



**National and Kapodistrian University of Athens**

**Department of Pharmacy**

**School of Health Sciences**

**STRUCTURAL AND FUNCTIONAL  
CHARACTERIZATION OF  
TRANSMEMBRANE TRANSPORTERS,  
USING COMPUTATIONAL APPROACHES.**

**Ph.D. Thesis**

**Iliana Zantza**

**Pharmacist**

Athens 2022

### **Advisory Committee**

Emmanuel Mikros, Professor, University of Athens, (Supervisor)

George Diallinas, Professor, University of Athens

Panagiotis Marakos, Professor, University of Athens

### **Examining Committee**

Emmanuel Mikros, Professor, University of Athens, (Supervisor)

George Diallinas, Professor, University of Athens

Panagiotis Marakos, Professor, University of Athens

Vittorio Limongelli, Professor, Università della Svizzera italiana

Stathis Frilligos, Professor, University of Ioannina

Antonios Kolokouris, Professor, University of Athens

Vasileios Myrianthopoulos, As. Professor, University of Athens

This research has been financed by the Stavros Niarchos Foundation (SNF)



## Περίληψη

Οι μεταφορείς είναι διαμεμβρανικές πρωτεΐνες που μεσολαβούν στην επιλεκτική μεταφορά διαλυμένων ουσιών μέσω των βιολογικών κυτταρικών μεμβρανών. Η οικογένεια Nucleobase Cation Symporter 1 (NCS1) που συναντάται σε προκαρυωτικούς οργανισμούς, μύκητες και ορισμένα φυτά, είναι ειδική για τις νουκλεοβάσεις. Η οικογένεια NCS1 απουσιάζει στα θηλαστικά και έτσι γίνεται ένας πολλά υποσχόμενος στόχος για αντιμικροβιακά φάρμακα. Επιπροσθέτως, αρκετοί ανθρώπινοι μεταφορείς (π.χ. μεταφορείς νευροδιαβιβαστών σεροτονίνης και ντοπαμίνης) σημαντικοί από βιολογική σκοπιά, έχουν την ίδια αναδίπλωση με τους μεταφορείς NCS1, ανοίγοντας το δρόμο για νέες γνώσεις σχετικά με την ανάπτυξη αντικαταθλιπτικών φαρμάκων και κατανόησης του μηχανισμού δράσης τους. Μια οικογένεια ειδική για νουκλεοβάσεις, διακριτή φυλογενετικά και δομικά, γνωστή ως Nucleobase Ascorbate Transporter (NAT) ή οικογένεια NCS2, υπάρχει σε βακτήρια, μύκητες, φυτά και μετάζωα, αλλά και σε ανώτερα πρωτεύοντα, συμπεριλαμβανομένων των ανθρώπων, όπου ορισμένα ομόλογα είναι ειδικά για L-ασκορβικό οξύ αντί για νουκλεοβάσεις. Ως εκ τούτου, ο δομικός και λειτουργικός χαρακτηρισμός των συστημάτων NCS1 και NAT/NCS2 αναδεικνύεται ως ένα σημαντικό πρώτο βήμα για μια επιτυχημένη διαδικασία σχεδιασμού φαρμάκων. Αν και υπάρχουν πολλά γενετικά, βιοχημικά και ορισμένα δομικά δεδομένα, λιγότερα είναι γνωστά για τον μηχανισμό μεταφοράς και την ειδικότητα τους. Αυτό είναι συνέπεια της δυσκολίας κρυστάλλωσης ή εκτέλεσης cryo-EM σε πολυτοπικά διαμεμβρανικά συστήματα με πολλαπλές διαμορφώσεις και των εξαιρετικά εκτενών υπολογιστικών προσεγγίσεων που απαιτούνται για τη θεωρητική περιγραφή τους. Ωστόσο, λίγα μέλη αυτών των οικογενειών έχουν κρυσταλλωθεί, υποδεικνύοντας έναν μηχανισμό εναλλασσόμενης πρόσβασης, σύμφωνα με τον οποίο ο μεταφορέας υφίσταται μεγάλες διαμορφωτικές αλλαγές για να αναγνωρίσει, να δεσμεύσει και να απελευθερώσει το υπόστρωμα κατά μήκος της μεμβράνης. Είναι αξιοσημείωτο ότι όλοι οι μεταφορείς NCS1 και NAT/NCS2 λειτουργούν ως συμμεταφορείς  $H^+$  ή  $Na^+$ .

Στην περίπτωση των NAT, 14 διαμεμβρανικά τμήματα (TMS) αναδιπλώνονται σε μια ανεστραμμένη επανάληψη 7 ελίκων (7+7). Οι κρυσταλλικές δομές έχουν ληφθεί από δύο μέλη των NAT, τον μεταφορέα ουρακίλης UraA του *E. coli* και τον μεταφορέα ουρικού οξέος-ξανθίνης του *A. nidulans*, και αποτελούνται από έναν τομέα πυρήνα και έναν τομέα διμερισμού και λειτουργούν ως διμερή. Ο μηχανισμός λειτουργίας τους θεωρείται πως είναι ο λεγόμενος «μηχανισμός ανελκυστήρα». Οι μεταφορείς NCS1, αποτελούνται από 12 TMS

αναδιπλωμένα σε μια ανεστραμμένη επανάληψη 5 ελίκων (5+5), ακολουθούμενα από δύο TMS που δεν έχουν άμεσο ρόλο στη μεταφορά του υποστρώματος. Αυτή η αναδίπλωση είναι χαρακτηριστική για όλους τους μεταφορείς της υπεροικογένειας APC, με κύριους κρυσταλλωμένους εκπροσώπους τον μεταφορέα υδαντοΐνης Mhp1, τον μεταφορέα λευκίνης LeuT, τον μεταφορέα ντοπαμίνης DAT και τον μεταφορέα σεροτονίνης SERT. Όλοι οι NCS1 μεταφορείς, λειτουργούν με τον μηχανισμό «rocking-bundle».

Η παρούσα μελέτη αποτελείται από τέσσερα μέρη. Στο πρώτο μέρος αποκαλύπτουμε τον τρόπο δέσμευσης, μετατόπισης και απελευθέρωσης της ουρακίλης και του H<sup>+</sup> από τον εξωκυττάριο χώρο στο κυτταρόπλασμα του μεταφορέα FurE (NCS1), χρησιμοποιώντας νέες προσομοιώσεις Μεταδυναμικής και ορθολογικά σχεδιασμένες μεταλλάξεις. Συγκεκριμένα, τα διαγράμματα της Επιφανείας Ελεύθερης Ενέργειας που προκύπτουν από τη Μεταδυναμική (FES), παρέχουν τη σχετική σειρά μεταφοράς της ουρακίλης και του πρωτονίου, επιτρέποντας επίσης την επιλογή ενδιάμεσων διαμορφώσεων, που σχετίζονται με τον κύκλο μεταφοράς. Η μεταδυναμική με χρήση περιοριστικής χοάνης (Funnel-Metadynamics), επιτρέπει δειγματοληψία αφενός των μονοπατιών που ακολουθεί το υπόστρωμα προκειμένου να συνδεθεί στη θέση πρόσδεσης και αφετέρου των τρόπων σύνδεσης της ουρακίλης και του πρωτονίου, δημιουργώντας ένα ολιστικό μοντέλο των γεγονότων που διέπουν τη μεταφορά υποστρώματος και που υποστηρίζονται από μελέτες μεταλλάξεων. Η εργασία μας όχι μόνο επεκτείνει τις υπάρχουσες γνώσεις σχετικά με τους μεταφορείς με τεταρτοταγή δομή τύπου FurE (LeuT, DAT, SERT, NET), αλλά επιβεβαιώνει για πρώτη φορά για ένα συμμεταφορέα πρωτονίων, τον λεγόμενο «rocking-bundle» μηχανισμό.

Στο δεύτερο μέρος, παρουσιάζουμε τα 3 μοντέλα του μεταφορέα UapA που κατασκευάστηκαν με χρήση Στοχευμένων Μοριακών Δυναμικών (Targeted Molecular Dynamics) και Μοντέλων Ομολογίας (Homology Modeling), σε διαφορετικές διαμορφώσεις, συνθέτοντας μαζί με την κρυσταλλική δομή τη διαμορφωτική αλλαγή που ακολουθείται κατά τον μηχανισμό του ανεγκυστήρα. Επιπλέον, εφαρμόζουμε Μεταδυναμική και αναδεικνύουμε σημαντικές πληροφορίες για τις διαμορφώσεις Inward-Open και Occluded, ενώ αποκαλύπτουμε την ασύγχρονη κίνηση των δύο πρωτομερών του διμερούς. Τέλος, περιγράφουμε τη συμβολή βασικών αμινοξέων στον μηχανισμό του ανεγκυστήρα και στη μεταφορά του υποστρώματος.

Στο τρίτο μέρος, διερευνούμε τον ρόλο του συντηρημένου μοτίβου NAT του UapA χρησιμοποιώντας μελέτες Επαγόμενης Πρόσδεσης (Induced Fit Docking), μεθόδους υπολογισμού Ελεύθερης Ενέργειας (Free Energy Methods) και ανάλυση μεταλλάξεων. Επιπλέον, προσφέρουμε νέα ευρήματα σχετικά με τον βασικό ρόλο του F528, ενός καταλοίπου

εκτός της θέσης δέσμευσης του υποστρώματος, το οποίο φαίνεται να ρυθμίζει την εξειδίκευση του UapA.

Στο τέταρτο μέρος, εξετάζουμε τις σχέσεις δομής-λειτουργίας των βακτηριακών μεταφορέων SmLL9, XanQ και ApcXanQ της οικογένειας NAT, χρησιμοποιώντας Μοριακές Δυναμικές (Molecular Dynamics), Μελέτες Επαγώμενης Πρόσδεσης (Induced Fit Docking) και μεταλλαξιγένεσης (site-directed mutagenesis), δίνοντας σημαντικά αποτελέσματα όσον αφορά την εξειδίκευση των μεταφορέων αυτών για τα υποστρώματα. Υπογραμμίζεται επίσης ο απροσδόκητα σημαντικός ρόλος των αμινοξέων εκτός της θέσης δέσμευσης στην εξειδίκευση, παρεμβαίνοντας στον μηχανισμό ολίσθησης του μεταφορέα.

## Abstract

Transporters are transmembrane proteins mediating the selective transport of solutes across the biological cell membranes. The Nucleobase Cation Symporter 1 (NCS1) family encountered in prokaryotes, fungi, and some plants is specific for nucleobases. The NCS1 family is absent in mammals and thus becomes a promising target for antimicrobial drugs. In addition, several biomedically important human transporters (e.g., serotonin and dopamine neurotransmitter transporters) have the same fold as NCS1 transporters, paving the way to new knowledge concerning the development of antidepressant drugs and understand their mechanism of action. A phylogenetically and structurally distinct nucleobase-specific family, known as Nucleobase Ascorbate Transporter (NAT) or NCS2 family is ubiquitously present in bacteria, fungi, plants and metazoa, and interestingly and in higher primates, including humans, some homologues are specific for L-ascorbic acid rather than nucleobases. Therefore, the structural and functional characterization of the NCS1 and NAT/NCS2 systems emerges as an important first step to a successful drug design process. Although there is plenty of genetic, biochemical and some structural data, less is known about their transport mechanism, and their selectivity. This is a consequence of the difficulty in crystallizing or performing cryo-EM of polytopic transmembrane systems in multiple conformations and the demanding computational approaches needed to theoretically describe them. However, few members of these families have been crystallized indicating an alternate-access mechanism, according to which the transporter undergoes large conformational rearrangements to recognize, bind and release the substrate across the membrane. Noticeably, NCS1 and NAT/NCS2 all transporters function as H<sup>+</sup> or Na<sup>+</sup> symporters.

In the case of NATs, 14 transmembrane segments (TMSs) are folded in a 7-helix inverted repeat (7+7). Crystal structures have been obtained from two NAT members, the UraA uracil transporter of *E. coli* and the UapA uric acid-xanthine transporter of *A. nidulans*, which form a core and a dimerization domain, and function as dimers. They are considered to operate under the so-called “elevator-mechanism”. NCS1 transporters, consist of 12 TMSs folded in a 5-helix inverted repeat (5+5), followed by two TMS that do not have a direct role in transport activity. This fold is characteristic of all APC superfamily transporters, with main crystallized representatives the Mhp1 hydantoin transporter, the LeuT leucine transporter, the DAT dopamine transporter, and the SERT serotonin transporter. All NCS1 operate under the “rocking-bundle” mechanism.

The present study consists of four parts. In the first part we reveal the binding mode, translocation, and release pathway of uracil/H<sup>+</sup> from the extracellular space to the cytoplasm of the NCS1 FurE transporter, using novel Metadynamics simulations and rationally designed mutational analysis. In particular, Metadynamics Free Energy Surface maps provide the relative order of internalization of the uracil and the proton, permitting also the selection of intermediate conformational states related to the transport cycle. Funnel Metadynamics allow the sampling of the binding pathway and modes of both uracil and proton, generating a holistic model of the transport events supported by mutation studies. Our work not only extends the existing knowledge on FurE same-fold transporters, namely LeuT, DAT, SERT, NET, but also confirms for the first time for a proton symporter, the so-called rocking-bundle mechanism. In the second part, we present the 3 models of UapA constructed using Targeted Molecular Dynamics and Homology Modeling, in different conformations, synthesizing along with the crystal structure the elevator-mechanism path. Furthermore, we apply Metadynamics and indicate important information about the Inward-Open and Occluded conformations, while revealing the asynchronous movement of the two protomers of the dimer. Finally, we describe the contribution of key residues in the elevator mechanism and substrate transport. In the third part, we investigate the role of the NAT signature motif of UapA using Induced Fit Docking, Free Energy Methods and mutational analysis. Additionally, we offer novel findings regarding the key role of F528, a residue outside the substrate binding site, which seems to regulate UapA specificity. In the fourth part, we address the structure-function relationships of the bacterial NAT transporters SmLL9, XanQ and AncXanQ, using Molecular Dynamics, Induced Fit Docking and site-directed mutagenesis, conducting important results regarding specificity. The unexpected role of residues outside the binding site to specificity, by interfering to the transporter's sliding mechanism is also underlined.

## Acknowledgements

First, I would like to thank my supervisor Professor Emmanuel Mikros, for his trust, guidance, support and encouragement throughout my PhD. I deeply admired his genuine curiosity about science and his ability to transfer it to me. I believe we had a creative collaboration which offered me the opportunity to grow as a scientist.

To continue I would like to thank my advisory committee, for guiding me as my thesis developed. In detail, I would like to thank Professor George Diallinas for the great collaboration all these years and Professor Panagiotis Marakos, for believing in me and guiding me since my student years in the Pharmacy school.

My sincere appreciations also go to Professor Stathis Frilligos, Professor Antonios Kolokouris, Assistant Professor Vasileios Myriantopoulos, for serving as my Examining Committee.

I want to express my biggest gratitude to Professor Vittorio Limongelli who kindly hosted me in his lab in Lugano. I am forever grateful to him for making me part of the team, for teaching me everything about Metadynamics and for assisting me throughout my PhD on the FurE and UapA projects. His scientific knowledge, can-do and problem-solving attitude inspired me and motivated me fundamentally. Special thanks to Dr. Stefano Raniolo for our excellent collaboration, his precious help, guidance, support, and encouragement. I would also like to thank Dr. Simone Aureli and Dr. Paolo Confliti for our collaboration, for always being there to help me in my projects and for supporting me.

I am also grateful to Professor Stathis Frillingos, Professor George Diallinas and their teams for our great collaboration over these years and for their guidance. Their passion for transporters motivated me a lot. Many thanks to Dr. Georgia Papadaki and Dr. Anezia Kourkoulou for our collaboration and for the enthusiasm and support they brought to our projects. It has been great working together.

I would also like to thank Dr. George Lambrinidis, for our collaboration, for his constant willingness to help me in my scientific projects, and for his encouragement. I also want to thank him for introducing me to Computational Chemistry when I was taking my first steps on the field.

In addition, I would like to thank the “Stavros Niarchos Foundation” for awarding me a scholarship and providing me the financial support to complete my Ph.D.

I am more than thankful to my fellow labmates in the Mikros' lab for creating a cheerful environment that enabled me to work with joy. Most of all I thank them all for their support and their valuable friendship during the bad and good days. You were the best partners in crime.

Finally, I cannot thank enough my family and friends for being always there for me, supporting and encouraging me. Words cannot explain my gratitude to my mother, Sofia, my father, Dimitris and my brother Akis for their unconditional love that kept me going through really demanding times. Thank you for believing in me so strongly that you kept me pushing into achieving what I considered impossible. I must admit that, until this moment that I am writing these lines, I have not met yet a person as intelligent as my little brother Akis, making me admire him so much. But even more than that, I admire him for being exactly the brother I needed through the years, and for that I thank him from the bottom of my heart.

# CONTENTS

Περίληψη.....	3
Abstract.....	5
Acknowledgements.....	7
Introduction .....	15
1.1 Membrane transport systems .....	15
1.2 Secondary Active Transporters categories .....	16
1.3 The Major Facilitator Superfamily (MFS).....	16
1.4 The Amino Acid-Polyamine-organo Cation (APC) Superfamily.....	17
1.4.1 The Neurotransmitter/Sodium Symporter (NSS) family.....	18
1.4.2 The bacterial amino acid transporter LeuT.....	18
1.4.3 The dopamine transporter (DAT).....	19
1.4.4 The serotonin transporter (SERT).....	20
1.5 The NAT (NCS2) family .....	22
1.5.1 The fungal uric acid-xanthine transporter UapA .....	22
1.5.2 The bacterial uracil transporter UraA.....	23
1.6 Mechanisms of transporter function .....	25
1.6.1 The Rocker-switch mechanism.....	25
1.6.2 The Elevator mechanism.....	26
1.6.3 The rocking-bundle mechanism.....	28
1.7 Approaches to the study of structure-function relationships in transporters.....	29
1.8 Gating mechanism and specificity in transporters.....	30
1.9 Residues important for specificity and transportation in UapA transporter.....	31
Aim of the thesis .....	32
Methods.....	33
2.1 FurE	
2.1.1 FurE protein model construction.....	33
2.1.2 Construction of the H <sub>3</sub> O <sup>+</sup> cation.....	33
2.1.3 FurE System Setup.....	34
2.1.4 Molecular Dynamics (MD).....	34
2.1.5 Metadynamics Simulations.....	35
2.2 XanQ and SmLL9 homolgy models' construction.....	41
2.2.1 Protein preparation .....	41



2.2.2 Induced-fit docking.....	42
2.2.3 Molecular Dynamics simulations.....	42
2.3 AncXanQ.....	43
2.3.1 Protein model construction.....	43
2.3.2 Protein preparation.....	43
2.3.3 Ligand preparation.....	43
2.3.4 Induced-fit docking.....	44
2.3.5 Molecular dynamics (MD).....	44
2.4 UapA transporter and mutants (SYSE, TFAE, SYSQ) .....	45
2.4.1 UapA-TFAE, UapA-SYSE and UapA-SYSQ Models Construction.....	45
2.4.2 Protein Preparation.....	45
2.4.3 Ligand preparation.....	46
2.4.4 Induced Fit Docking.....	46
2.4.5 Molecular Dynamics Simulations.....	46
2.4.6 Metadynamics Simulations.....	47
Results.....	48
3.1 Uracil/H <sup>+</sup> symport by the FurE transporter challenges the rocking-bundle mechanism of transport in APC transporters.....	48
3.1.1 Foundation of the study.....	48
3.1.2 The FurE 3D structure.....	50
3.1.3 Mutational analysis confirms the crucial role of specific residues in FurE transport function.....	56
3.1.3 Molecular Dynamics calculations.....	59
3.1.3.1 The binding mode of hydronium.....	59
3.1.3.2 The binding mode of uracil.....	61
3.1.3.3 The conformational transition of FurE from OO to IO.....	63
3.1.3.4 OO-to-Occ path.....	63
3.1.3.5 Occ-to-IO path.....	68
3.1.3.6 The internalization pathway of H <sub>3</sub> O <sup>+</sup> cation.....	69
3.1.3.7 The internalization pathway of uracil.....	71
3.1.4 Conclusions.....	72
3.2 NCS2/NAT family transporters (Structural and Functional characterization of UapA transporter.....	77
3.2.1 Foundation of the study.....	77

3.2.2 Construction of the three UapA models.....	78
3.2.3 The UapA crystal structure.....	78
3.2.4 The Inward-Occluded Model structure.....	80
3.2.5 The Outward-Open model structure .....	82
3.2.6 Comparison of the binding sites of the four models.....	87
3.2.7 Metadynamics simulations for the IOp-to-IOcc transition step.....	88
3.2.7.1 The Free Energy Surface of the IOp-to-IOcc transition step.....	88
3.2.7.2 The Inward Occluded Structure through Metadynamics .....	91
3.2.8 The NAT signature motif and its structural and functional role.....	93
3.2.9 Conclusion.....	93
3.3 Context-dependent cryptic roles of specific residues in substrate selectivity of the UapA purine transporter .....	94
3.3.1 Foundation of the study .....	94
3.3.2 Context-dependence role of the NAT signature motif in determining UapA specificity.....	95
3.3.3 Computational analysis gives insights on the specificity modifications of the various NAT signature motif versions.....	97
3.3.4 F528 dynamically interacts with the substrate binding site.....	100
3.3.5 Conclusions.....	101
3.4 NCS2/NAT family transporters (XanQ, SmLL9, AncXanQ): Structure, function, specificity, mechanism of transportation and crucial amino acids as key factors.....	102
3.4.1 The case of SmLL9 transporter.....	102
3.4.1.1 Foundation of the study .....	102
3.4.1.2 SmLL9 purine nucleobase transporter of <i>S. meliloti</i> has enlarged specificity profile.....	103
3.4.1.3 Homology modeling of SmLL9 and comparison with XanQ.....	103
3.4.1.5 Key binding-site residues are conserved between SmLL9 and XanQ preserving their functional role.....	114
3.4.1.6 Mutation of S93N converts SmLL9 to a xanthine-selective transporter.....	117
3.4.1.7 The important role of S93 in binding uric acid in SmLL9 transporter as revealed by Induced Fit Docking calculations (IFD).....	118
3.4.1.8 Conclusion.....	121
3.4.2 The case of AncXanQ transporter.....	122
3.4.2.1 Foundation of the study.....	123

3.4.2.2 Structural model of AncXanQ transporter .....	124
3.4.2.3 Transport assay analysis of AncXanQ .....	128
3.4.2.3 Binding mode of xanthine and guanine on AncXanQ.....	130
3.4.2.4 Combination of 5 mutations outside the binding site restricts specificity of AncXanQ to recognition of xanthine.....	133
3.4.2.5 The role of Ser/G377 in relevance to the xanthine specificity of XanQ as derived from experimental procedures.....	133
3.4.2.6 Structural alterations in the XanQ S377G mutant.....	134
3.4.2.7 Conclusion.....	140
References.....	143
Appendix: Publications and Pre-prints .....	159

## Abbreviations

5FU: 5-fluorouracil

APC: Amino acid-Polyamine-organocation superfamily

CV: collective variable

DPPC: Dipalmitoylphosphatidylcholine

FEP: Free Energy Surface

FM: Funnel Metadynamics

H<sub>3</sub>O<sup>+</sup>: hydronium

HX: Hypoxanthine

MD: Molecular Dynamics

MFS: Major Facilitator Superfamily

NAT: Nucleobase Ascorbate Transporter family NCBI: National Center for Biotechnology Information NCS1: Nucleobase Cation Symporter-1 family

NCS2: Nucleobase Cation Symporter-2 family

ND: Not Determined

ns: nanoseconds

NSS: Neurotransmitter Sodium Symporter family

PC: Phosphatidylcholine

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PE: Phosphatidylethanolamine

PI: Phosphatidylinositol

PCV: Path Collective Variable

RMSD: Root Mean Square Deviation

SulP: Sulfate Permease family

U: Uracil

UA: Uric acid

wt: Wild Type

X: Xanthine

# 1 Introduction

## 1.1 Membrane transport systems

The survival of cells depends on nutrients, ions and also other small non-polar or polar molecules located in the extracellular environment. Some lipophilic molecules can enter the intracellular space by simple membrane diffusion. Most other solutes and ions are dependent on structurally and mechanistically diverse class of transmembrane proteins in order to enter or exit the cell. Transmembrane proteins mediate the exchange of a large variety of molecules but also play an important role in other physiological and pathophysiological processes such as signal transduction and are associated with many diseases including cystic fibrosis, cancer, autism, epilepsy, and neurodegeneration<sup>1-5</sup>. Drug sensitivity or drug resistance (efflux proteins) directly involves many transporters. The above summarize the important role they fulfil in cells.

There are two types of transmembrane transport proteins, transporters, and channels/pores. Channels/pores do not have binding sites and can be open to both environments at the same time (pores) or exist in an open or a closed state (channels)<sup>6,7</sup>. The regulation of the opening and closing of channels is made by gating domains or gates in response to signals like ligand binding or membrane voltage changes<sup>6-8</sup>. Transporters do not have a continuous pore, but they have a gating mechanisms to control substrate translocation across the membrane differentiating from the channels<sup>6,7</sup>. They bind their substrates at one side of the membrane and undergo a series of conformational changes that bring the substrate-binding site on the other side of the membrane, where the substrate is released. The channels' selectivity is regulated by filters that restrict the size of the hole as they do not possess a binding site. On the other hand, transporters are characterized by a main substrate binding site, a rate of transport catalysis and a specificity profile. A transport system can be distinguished into passive or active based on the consumption of energy. When transport is passive, the molecules are transported towards the direction of their electrochemical gradient. Instead, active transport requires energy in order to move molecules from an area of lower to higher concentration. Both transporters (facilitated diffusion) and channels (passive diffusion) can facilitate passive transport. However, only transporters can execute active transport. Energy can be provided either directly (primary active transport), as for example from ATP-hydrolysis or from the electrochemical potential difference (secondary active transport). Most of the primary active transporters are also

ATPases proteins that provide the energy needed for transport by hydrolyzing ATP. There are four types of primary active transporters: P-type ATPases also referred as pumps (Na<sup>+</sup>-K<sup>+</sup> ATPase), F-ATPases (mitochondrial ATP synthase), V-ATPases and ABC transporters e.g. CFTR. On the other hand, in secondary active transport the movement of a driving ion down its electrochemical gradient is coupled to the movement of another molecule or ion against its concentration gradient. Taking advantage of the electrochemical gradients made by primary active transport, it is considered a form of active transport. Secondary active transporters can be categorized as symporters or antiporters if the transport of the substrates is in the same direction or if the substrates move in opposite directions, respectively.

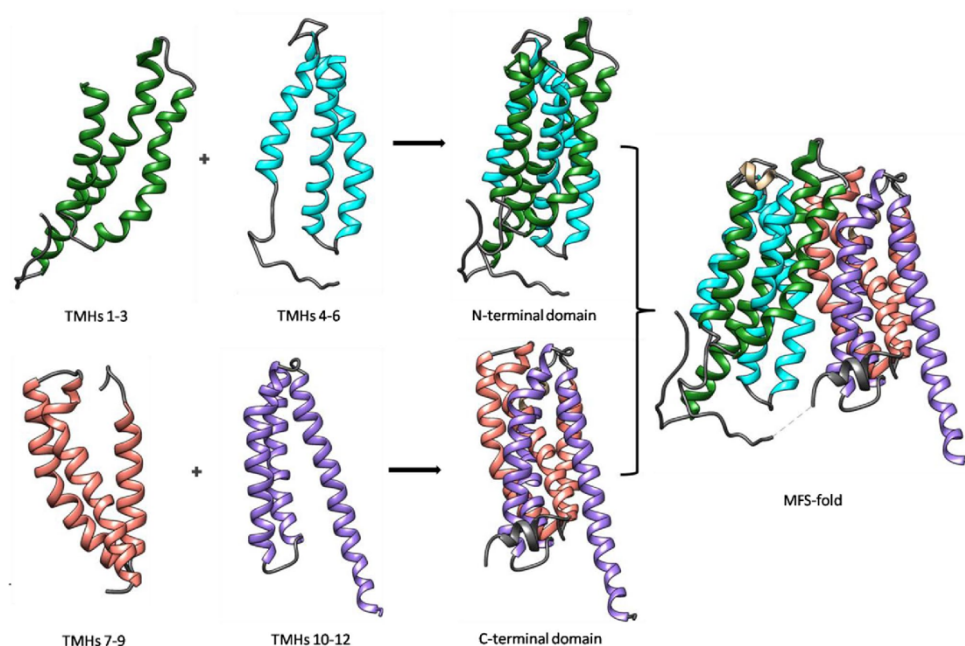
## 1.2 Secondary Active Transporters categories

Transport proteins are classified in the Transporter Classification Database (TCDB) based on their function and phylogenetic origin<sup>9-11</sup>. This classification organized them into classes, subclasses, families, subfamilies and transport systems, while several families have been organized into sequence divergent superfamilies. The three major superfamilies are the Major Facilitator Superfamily (MFS) and the Amino acid-Polyamine-organo Cation superfamily (APC), both of which predominantly include secondary active transporters, and the ATP-binding cassette superfamily (ABC), consisting of primary active transporters.

## 1.3 The Major Facilitator Superfamily (MFS)

The transporters within the MFS are symporters, antiporters or uniporters. The transfer of the molecules is performed by taking advantage of the electrochemical gradient of the co-transported ion or the concentration of the ligand to initiate the transport cycle. The substrates they carry are of great physiological and clinical importance, including sugars, drugs, neurotransmitters, amino acids, vitamins, organic and inorganic ions and many other small compounds<sup>12,13</sup>. The MFS protein structure (**Figure 1.1**) consists of 12 transmembrane segments (TMSs) arranged in two symmetrically related bundles of six TMSs each, connected by a loop. Within each bundle there is a pseudo two-fold axis of symmetry of three TMSs. The first crystalized representatives of this fold was the lactose-H<sup>+</sup> symporter LacY<sup>14</sup> and the GlpT<sup>15</sup>. The LacY mechanism of symport has been studied extensively and the key residues involved in the translocation pathway are well-characterized. Other MFS structures known include the following bacterial ones: the multidrug transporter EmrD<sup>16</sup>, glycerol-3-phosphate-

phosphate antiporter GlpT<sup>17</sup>, the xylose/proton symporter XylE<sup>18</sup> etc. and some eukaryotic ones: the *Piriformospora indica* phosphate transporter PiPT<sup>19</sup>, the plant nitrate transporter NRT1.1<sup>20</sup> and the human glucose transporter GLUT1<sup>21</sup>. The MFS transporters function as monomers although there is evidence of oligomerization in some plant members<sup>22</sup>, like NRT1.1 which can homodimerize according to the crystal structure<sup>20</sup>.



**Figure 1.1:** Secondary active transporters' shared folds. Representation of the MFS fold. The figure is adapted from <sup>23</sup>.

#### 1.4 The Amino Acid-Polyamine-organo Cation (APC) Superfamily

The second largest superfamily of secondary active transporters, the APC superfamily, consists of numerous families of transporters specific for nucleobases, amino acids, metal ions, purines, pyrimidines, peptides, vitamins etc and function as solute/cation symporters or solute/solute antiporters<sup>24-26</sup>. High-resolution crystal structures of APC proteins have been published, revealing their structure. Most of them contain 10 TMS organized in a 5-helix inverted repeat topology (5+5) known as the LeuT fold<sup>27</sup>, that form two structurally distinct bundles referred as the scaffold and core domain<sup>28</sup>. However, some display a 7+7 TMS fold. Among the 5+5 members, most contain two extra TMSs at their N- or C-terminal part. In some cases, there are four extra C-terminal TMSs or in other cases, there is one extra TMS before every inverted repeat unit. This structural diversity implies that multiple hairpin and domain duplication events occurred during the evolution of this family<sup>25</sup>. The substrate binding site is located near

the center of the transporter between TMS1 and TMS6 which are a symmetrical pair. They share a common mechanism of transport and gating reflected in the alternating access model. The APC superfamily includes also the Nucleobase/Cation Symporter family 1 (NCS1), the Nucleobase/Ascorbate Transporter (NAT) or Nucleobase Cation Symporter-2 (NCS2) family (7+7-fold), the Solute/Sodium Symporter (SSS) family and the Neurotransmitter/Sodium Symporter (NSS) family, the Sulfate Permease (SulP) family (7+7 fold), and finally the Anion Exchanger (AE) family.

#### **1.4.1 The Neurotransmitter/Sodium Symporter (NSS) family**

The Neurotransmitter/Sodium Symporter (NSS) family members catalyze the transport of a wide variety of substrates, including biogenic amines (serotonin, dopamine, norepinephrine), amino acids (GABA, glycine, proline, taurine) and osmolytes (betaine, creatine), using sodium and chloride electrochemical gradients<sup>29,30</sup>. Importantly, this family includes the dopamine transporter (DAT), the norepinephrine transporter (NET) and the serotonin transporter (SERT), which are very important pharmacological targets as their malfunction leads to multiple disorders including depression, Parkinson's disease, orthostatic intolerance, epilepsy, attention deficit hyperactivity disorder (ADHD) or obsessive-compulsive disorder (OCD)<sup>31-36</sup>. Consequently, they are targets for many antidepressant drugs and also for cocaine and amphetamine. This family also includes the prototype amino acid transporter LeuT.

#### **1.4.2 The bacterial amino acid transporter LeuT**

LeuT, the bacterial amino acid symporter from *Aquifex aerolicus*, is the most well-studied member of the NSS family. LeuT is in fact the role model of the APC superfamily establishing a common fold among its members. LeuT has been crystalized in all three different conformations of these type of transporters (outward-open, outward-occlude, inward) while several biochemical data are available, providing important details on the mechanism of substrate binding and transport. It is a non-polar amino acid/Na<sup>+</sup> symporter and even though it was named after leucine, it is specific for several aliphatic/aromatic amino acids including glycine, alanine, methionine and tyrosine and can also bind tryptophan but without transporting it<sup>37</sup>.

LeuT consists of 12  $\alpha$ -helical transmembrane segments (TMS1-TMS12) connected by loops and helices on the intra- and extracellular part of the cell membrane. The 5-helix inverted repeat



(TMSs 1-10) is organized relating TMSs 1-5 and TMSs 6-10 by an antiparallel pseudo two-fold axis located in the plane of the membrane. TMS 1 and TMS 6 are folded in an antiparallel orientation. They are the most conserved among the human transporter homologues and they share the common characteristic that they are not continuous helices implementing a tilt in the middle. TMSs 3 and 8 are also antiparallel and along with TMSs 1 and 6 comprise the substrate and sodium binding sites. TMSs 2 and 7 uphold TMSs 1 and 6 respectively, while TMSs 4 and 5 form a V-shape structure as do TMSs 9 and 10, supporting TMSs 3 and 8. TMS11 and TMS12, serve in the dimerization of the transporter. Two Na<sup>+</sup> binding sites have been identified and located near TMS1 and TMS6, having key roles in stabilizing the core of the transporter and the leucine binding. LeuT functions under the rocking-bundle alternating-access model mechanism.

### 1.4.3 The dopamine transporter (DAT)

Dopamine transporter, one of the most pharmacologically important transporters for human, has not been crystalized yet. Nevertheless, X-ray structure of the homologous *Drosophila melanogaster* dopamine transporter (dDAT) (50% sequence identity with its mammalian hDAT) has been determined. dDAT was crystalized in an inhibitor-bound, outward-open conformation<sup>38</sup>. Thus, dDAT is a powerful tool to study NSS pharmacology and transport mechanisms. As dDAT is the eukaryotic relative to the LeuT<sup>28</sup>, they share common structural features. Analytically, dDAT consists of 12 transmembrane segments of which TMSs 1-5 and 6-10 are characterized by antiparallel symmetry. TMS1 and TMS6 include the ligand and ion binding sites in their mid-helix breaks. A bonding network connects all three ions and the inhibitor. TMS3 also contributes through a bend, to the hydrophobic pocket where the ligand is bound.

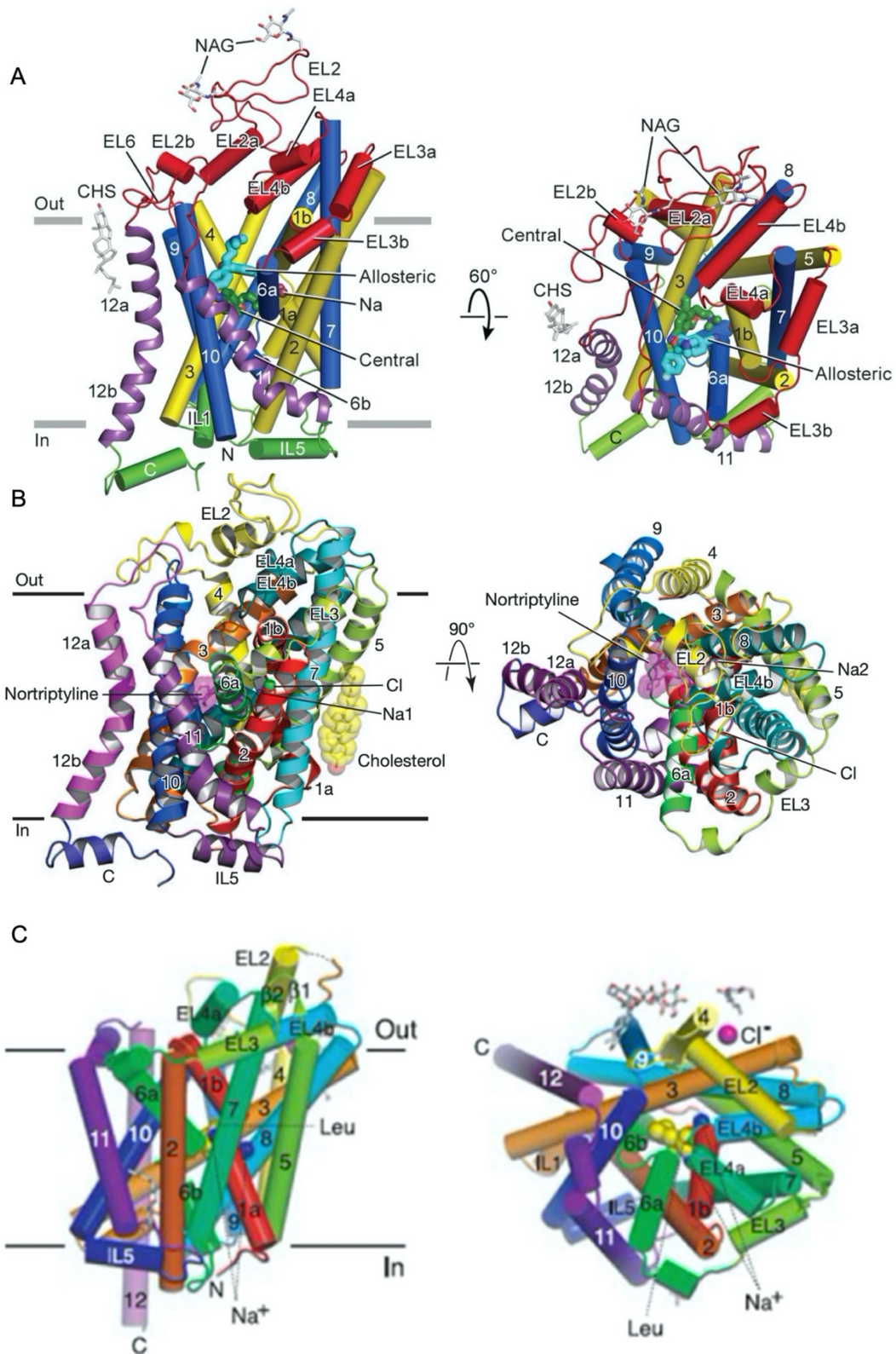
Interestingly, cholesterol plays an important role in the transport mechanism, as it modulates the movement of TMS1a during the neurotransmitter transport. Additionally, important features for the transport cycle are revealed from the the crystal structures, such a kink at a proline residue in TMS12 that causes a tilt in the center of the helix increasing the distance of the second half of the TMS12 from the inner helices. This feature explains why the crystal of dDAT is a monomer, unlike other NSS transporters, that form oligomers<sup>39,40</sup> in the interface of the last TMS. Moreover, the extracellular loop 2 (EL2) seems to be important for the proper transporter localization in the plasma membrane, as it consists of many N-linked glycosylation sites and a disulfide bond. Although the core of dDAT resembles that of LeuT,

they differ in a latch-like C-terminal helix of the first, that interacts extensively with the lower cytoplasmic part of the transporter, but also in a kink in the middle of TMS12 and in the presence of a cholesterol molecule between TMS1a, 5 and 7.

Several studies on the human homologue hDAT reveal that in NSS proteins, the N-terminal region of the transporter participates in the efflux process of the substrates. Still, the mechanism of the DAT reverse transport remains unclear, due to lack of structural information<sup>41-43</sup>. As the amino acid sequence of the N-terminus is not conserved among its homologues or any protein with known fold, ab initio structure prediction tools and extensive atomistic MD simulations were combined revealing two distinct structural elements: a beta sheet motif and a  $\alpha$ -helical segment. Phosphorylation and interactions with specific membrane lipids, the internal loops (ILs) of the transporter or the N-terminal region, proved to be critical for the efflux procedure<sup>41-43</sup>.

#### 1.4.4 The serotonin transporter (SERT)

The X-ray structure of the human serotonin transporter was determined in an inhibitor-bound, outward-open conformation<sup>44</sup>, characterizing the mechanism of antidepressant action and paving the path for potential drug design. SERT comprises of 12 transmembrane segments with TMSs 1-5 and 6-10 related by a pseudo-2-fold axis, similarly to the LeuT and DAT (**Figure 1.2**). As in dDAT, the intracellular part of the transporter is capped by IL1, IL5 and the C-terminal region, with the latter having a pronounced kink halfway across the membrane and a cholesterol hemi-succinate (CHS) molecule bound near TMS12a. Although transporter is a monomer in detergent, the oligomerization of the SERT was experimentally suggested<sup>45</sup>. In contrast with dDAT, the TMS9 of SERT is shifted towards TMS12 and TMS11 extends further into the putative membrane environment, providing a larger cavity for allosteric ligands. Moreover, interactions between cholesterol and the lipid molecules seems to support the conformation of TMS12. The C-terminal region of SERT resembles dDAT with a similar hinge and helix region<sup>44</sup>.



**Figure 1.2:** Structural representation of 5+5 fold transporters in side (left panel) and top view (right panel). (A) SERT<sup>44</sup>, (B) DAT<sup>38</sup>, (C) LeuT transporter<sup>28</sup>.

## 1.5 The NAT (NCS2) family

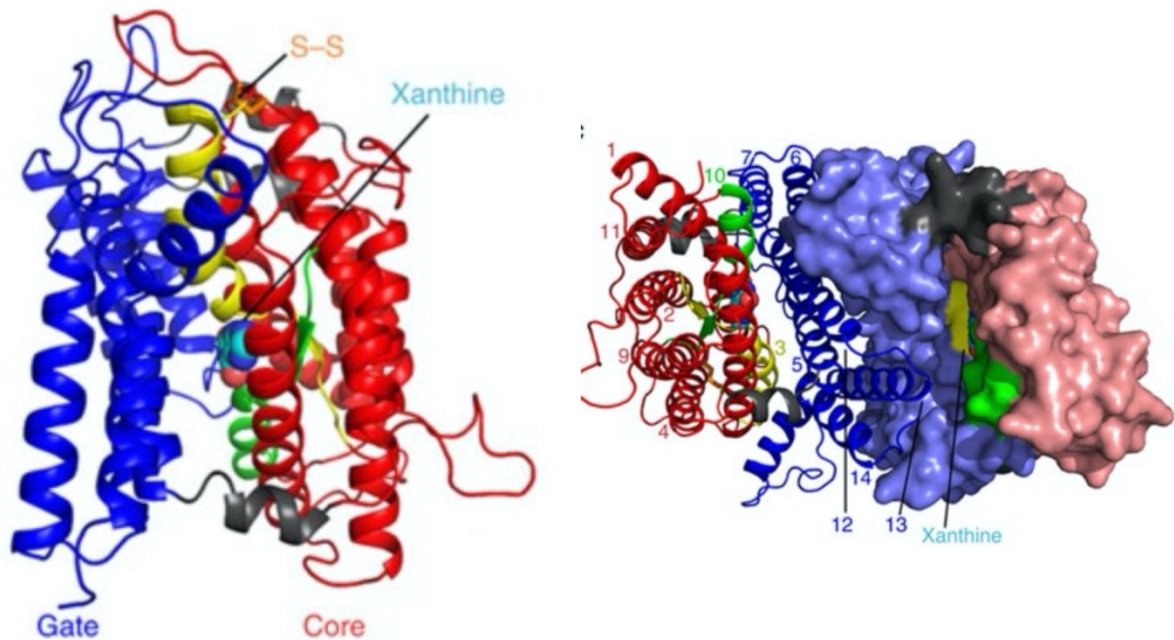
The NAT/NCS2 family includes bacterial, fungal and plant H<sup>+</sup> or Na<sup>+</sup> symporters, specific for purines, pyrimidines and related molecules. Remarkably, the mammalian L-ascorbate/Na<sup>+</sup> transporters SVCT1 and SVCT2 also belong in the NAT/NCS2 family. The first members characterized and crystalized were the uric acid-xanthine transporters of *Aspergillus nidulans* UapA and UapC and the uracil transporters of *Escherichia coli* UraA and PyrP<sup>46,47</sup>. Structures of eight structural homologues have been also resolved. From the SLC4 family, the borate transporter Bor1 from *Arabidopsis thaliana*, the Bor1p from *Saccharomyces mikatae*, the Band3 human anion exchanger 1 and the human acid-base/Na<sup>+</sup> transporter NBCe1<sup>48-51</sup>. The members of the SLC26 family are the fumarate/H<sup>+</sup> symporter SLC26Dg from the bacterium *Deinococcus geothermalis*, the bicarbonate transporter BicA from cyanobacteria *Synechocystis* sp., the Slc26a9 protein from *Mus musculus* and the human SLC26A9 protein<sup>52-55</sup>.

The crystal structures of NAT/NCS2 transporters revealed the tertiary structure of this family which contains 14 transmembrane segments organized in a 7+7 inverted repeat fold<sup>56-58</sup>.

### 1.5.1 The fungal uric acid-xanthine transporter UapA

UapA from *A. nidulans* is a xanthine-uric acid/H<sup>+</sup> symporter<sup>6</sup> (**Figure 1.3**). Recently, the crystal structure of UapA was determined<sup>58</sup>, including the mutation G411V and an N-terminal truncation ( $\Delta$ 1-11). Wild-type UapA, which was rather unstable, has not been successfully crystalized yet. This crystalized mutant version can bind, but cannot transport, xanthine. The overall structure of UapA contains 14 TMSs organized into a 7+7 fold and subdivided in a core (TMSs 1-4 and 8-11) and a gate domain (TMSs 5-7 and 12-14). The crystal structure of UapA also revealed that it functions as a homodimer. UapA compared to a bacterial homologue, UraA, has longer loops between the TMSs. TMSs 3 and 10 are  $\beta$ -strands located in the center of the transporter and contain the binding site approximately halfway across the membrane. The dimer is formed with the participation of TMSs 5, 12, 13 and 14 of the gate domain with TMS13 fitting into a cleft formed by the other protomer. Genetic, cellular and biochemical studies reveal a critical role of the dimerization in UapA trafficking and turnover<sup>59</sup>. The topology of important residues affecting specificity was highlighted e.g. binding site F406, Q408, F155 belonging to the binding site, Thr526, F528, V463 and A469 located along the putative substrate translocation pathway, R481 facing the substrate binding site of the opposite protomer, or Q113, A441 in hinge regions connecting the core and gate domains. The dimeric

crystal structure along with Molecular Dynamic simulations succeeded in explaining the important role of R481 which is an inter-dimeric contribution to substrate release in the cytoplasm. Molecular dynamics suggested a translocation pathway for xanthine internalization. This movement of the substrate is achieved due to the contribution of R481 from the opposite protomer.



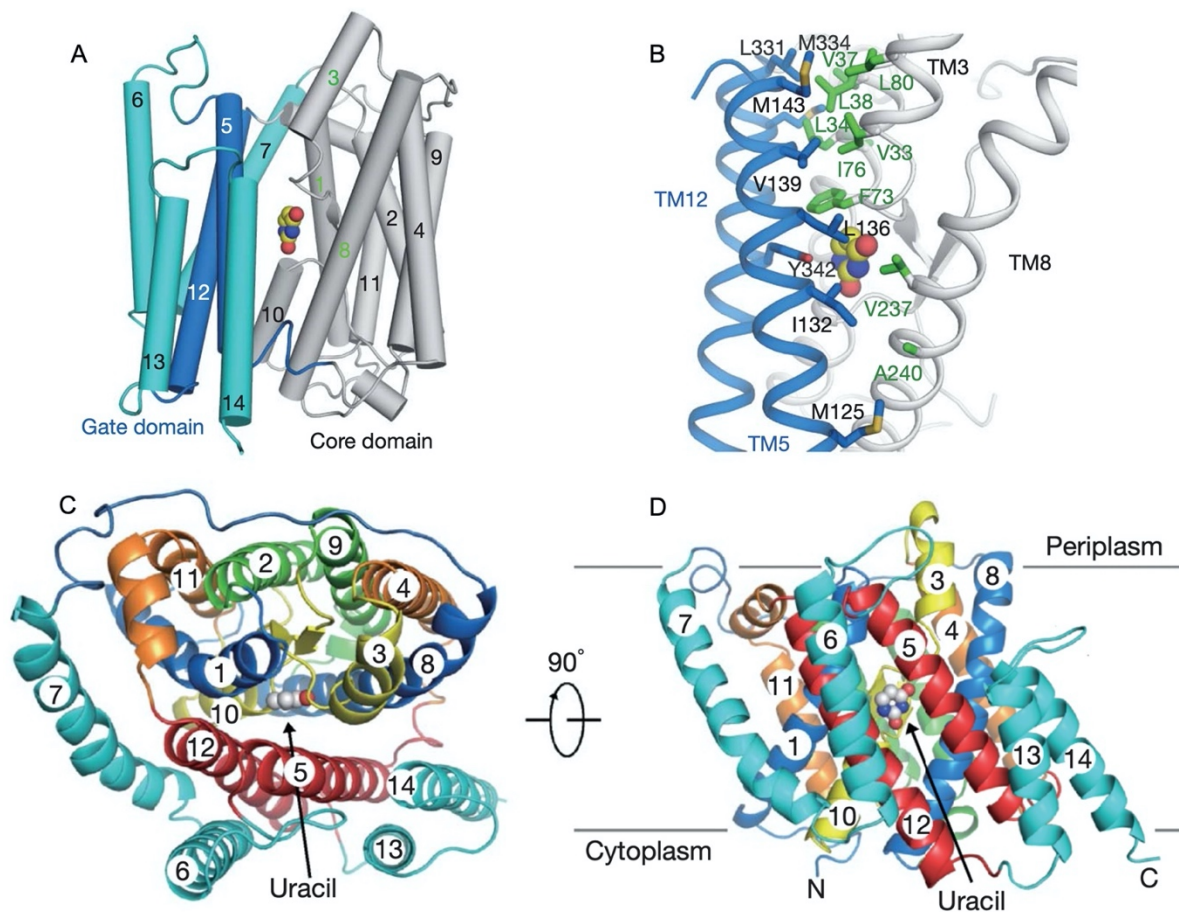
**Figure 1.3:** Ribbon representation of the UapA monomer. (On the left) The gate domain is shown in blue while most of the core domain is shown in red. The amphipathic helices that link the core and gate domains are shown in grey. Xanthine is shown in cyan as a space-filling model and the disulphide bond is shown in orange sticks. The N-terminal (residues 12–65) and C-terminal (residues 546–574) are missing for clarity reasons. (On the right) UapA dimer from a cytoplasmic point of view. One monomer is shown in ribbons, with the helices numbered, and the other is shown in surface representation with the same colouring. The figure is adapted from<sup>58</sup>.

### 1.5.2 The bacterial uracil transporter UraA

The uracil/H<sup>+</sup> symporter UraA from *E. coli*, has been crystalized<sup>56,57</sup> (**Figure 1.4**) and appears to function as a dimer. Systematic *in vitro* and *in vivo* characterizations support that UraA is in equilibrium between the monomer and dimer form and that dimerization is required for transport activity<sup>57</sup>. It consists of 14 transmembrane segments 12  $\alpha$ -helical and 2 antiparallel  $\beta$ -strands in TMS3 and TMS10 in the middle structure. The 14 TMSs are arranged into two structural repeats, TMSs 1-7 and TMSs 8-14, standing antiparallel along an axis parallel to the membrane. Furthermore, the TMSs can be divided in two distinct domains: the core domain consisting of TMSs 1-4 and 8-11, while the gate domain consisting of TMSs 5-7 and 12-14.



The TMSs participating in the dimerization are the same as in UapA. The substrate binding site is located between TMSs 1, 3, 8 and 9. Two water molecules have also been crystallized in the vicinity of the binding site, participating in a hydrogen bond network that stabilizes the substrate in the correct binding mode<sup>57</sup>. The residues involved in this hydrogen bond network are E241 and H245 (E356 and D360 in UapA). This result potentially provides a path for H<sup>+</sup> translocation. Molecular dynamics stressed the role of these residues in H<sup>+</sup> translocation revealing that protonation of E241 resulted in disruption of the hydrogen bond network that keeps uracil bound, while protonation of H245 further stabilizes it, without interacting directly but by stabilizing the orientation of E241.



**Figure 1.4:** Structural representation of UraA transporter. (A) Core (grey) and gate (cyan and blue) domains of UraA. (B) Antiparallel b-strands in TMS3 and TMS10 provide the organizing centre for the core domain. the residues that mediate the interdomain interactions are shown as green (core) and blue (gate) sticks. (C) Structure of FurE from a top view. (D) UraA from a side view. This figure is adapted from<sup>56</sup>.

## 1.6 Mechanisms of transporter function

Despite diversity in structure and mechanistic details, active membrane transporters all operate via the alternating access model<sup>60-62</sup>, during which substrate accessibility is switched between the two sides of the membrane. The transporter interconverts between two major functional states, the inward-facing (IF) and outward-facing (OF) states, through transition pathways visiting multiple intermediate states involving at least one occluded state which prevents free diffusion of the substrate along the gradient. Besides the major transition from the outward to the inward-facing conformation, transport cycle involves local conformational changes that regulate the closing and opening of extracellular and intracellular gating elements. The latter implement an additional control in transport catalysis by preventing the substrate leakage to the wrong direction and by contributing to the specificity determination. Gating element closure may anticipate the alteration from the outward to the inward-facing conformation<sup>63</sup>. Thus, 5 transporter states have been documented; outward-open, outward-occluded (i.e. outer gate closed), fully-occluded (outer and inner gates closed), inward-occluded (inner gate closed), inward-open. Based on structures and biochemical data, three major transport mechanisms have been described, the gated rocker-switch, the rocking bundle, and the elevator mechanism.

### 1.6.1 The Rocker-switch mechanism

The rocker-switch mechanism describes two structurally symmetric domains form a characteristic V-shaped architecture, move with respect to a binding site located at the interface, approximately halfway across the membrane<sup>64,65</sup>. During a transport cycle, the binding site is exposed on one side of the membrane, the substrate binds and then, the protein opens to the other side of the membrane with a rocker-switch like movement around the region of the binding site and the substrate is released<sup>61,66,67</sup>. In the simplest version of this mechanism the substrate binds symmetrically (e.g. SWEETs transporters<sup>68-70</sup>). In this case, the occluded conformation is mainly symmetric forming an O-like shape. On the other case, where substrate binding is asymmetric to varying degrees (e.g. MFS transporters)<sup>11,65</sup>, in most occluded MFS structures, cavities still have a clear orientation<sup>61</sup>. In these asymmetric cases, the states are typically referred to as outward- occluded or inward-occluded conformations. Furthermore, when substrate binds one or both the domains that occlude it, are prompted to local movement of their gating helices achieving transport<sup>71,72</sup>. Finally, it has been proposed for some MFS transporters that the transition between the two states is controlled by the breakage and re-

formation of salt bridges<sup>65,73,74</sup>, induced by the substrate binding which leads to direct or indirect movements of helices involved in the salt bridge. Two structurally similar domains, most often made by six transmembrane domains (TMS) each (i.e., transporters with 6+6 fold), move to provide access or release of substrates (S) from either side of the membrane<sup>11,61</sup>. The substrate binding site is made by residues located in both major domains.

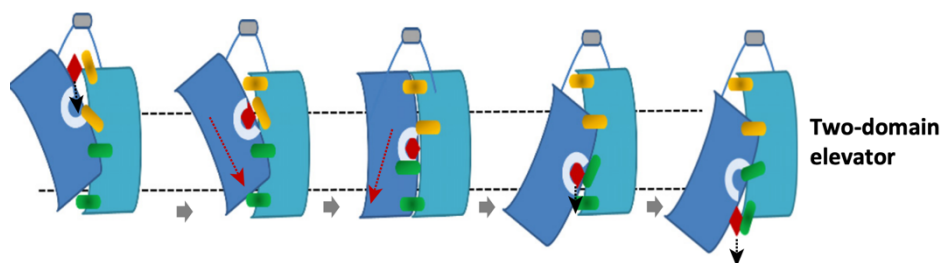
### 1.6.2 The Elevator mechanism

The elevator-type transporters are made up of two distinct domains<sup>61</sup>. One is the gate/scaffold domain which is anchored in the plasma-membrane via specific interactions with lipids and the core/transport domain which moves through the membrane remaining structurally unaltered (**Figure 1.5**). The binding site is included in the mobile core domain; thus the substrates are translocated from one side of the membrane to the other as the core domain undergoes a large rigid-body sliding against the rigid gate domain<sup>75,75</sup>. Despite structural differences, all elevator-type transporters (symporters and antiporters) share these common characteristics<sup>76</sup>. The gate domain is the dimerization domain, as transporters with this mechanism are usually dimers or trimers<sup>61,76</sup>. Homodimerization, is reported to be critical for subcellular trafficking, transport activity, and specificity<sup>76-79</sup>. However, the only case that oligomerization has a role in substrate specificity is that of an elevator-type transporter, namely UapA<sup>58,76</sup>. Remarkably, the gate domain seems to provide most of the gating elements and is much shorter and consequently buried in the membrane compared to the core domain, leading to the formation of water-filled areas in the membrane, in at least one of their conformational states. In rocker-switch and rocking-bundle transporters, both domains are of comparable height. The substrate binding sites are typically located at the bottom of the core domain but, the substrate can access the binding site easily and a substrate-navigation through a buried protein cavity is not required, due to the gate domain being shorter. These factors facilitate substrate diffusion to the binding site. According to the transport mechanism both solutes are bound before the protein gates are fully closed and allow the translocation step. The elevator mechanism can be classified into three types, based on the gating mechanism: the fixed-barrier with one gate, the fixed barrier with two gates and the moving-barrier with two gates<sup>75</sup>. Classification is possible only if all the different conformational states are determined for the protein, so the mechanism of many elevator proteins has not been experimentally determined. The first elevator-type mechanism described was a homolog of glutamate transporters the sodium–aspartate symporter GltPh from *Pyrococcus horikoshii*<sup>80,81</sup>, and characterizes mostly



transporters with a 7+7-fold. In elevator-type transporters, residues critical for specificity are mostly located in the scaffold domain and the interface of the core-gate domains. Structural data suggests that, in the apo state, elevator-type transporters are stabilized in an outward-facing state with the substrate binding site exposed to the extracellular environment and that the sliding to the inward open state requires substrate/cation to be bound<sup>82</sup>. However, due to lack of structural evidence for intermediate states for a single elevator-type transporter, until recently, formal support of this mechanism remained inconclusive.

In the case of DASS dicarboxylate/ $\text{Na}^+$  transporters<sup>83</sup>, the comparison of the outward- and inward-facing structures of four different states, that the sliding is initiated by the substrate and cation binding. Most DASS proteins are  $\text{Na}^+$ -coupled co-transporters, but some are exchangers of succinate for other dicarboxylates. The structures determined, and molecular dynamics simulations (MD) of the transition from outward- to inward-facing revealed, that if the carboxylate ion is absent, the substrate binding site has a positive net charge, which prevents the sliding of the elevator. This positive net charge derives by the binding of  $\text{Na}^+$  cations in the symporters, or by basic residues of the protein itself in antiporters. But when the carboxylate ion binds and neutralizes the binding site, the transporter changes between outward and inward conformations. It seems that binding of physiological substrates “unlocks” the elevator sliding and by exposing the binding site to the other side of the membrane the substrate is released in the cytoplasm<sup>83,84</sup>. In agreement the DASS transporters, it has also been suggested that in the case of NapA, an  $\text{Na}^+/\text{H}^+$  elevator-type antiporter,  $\text{Na}^+$  binding has to neutralize the charge of an aspartate which is strictly conserved in the ion-binding site. This way transition from outward to inward states is achieved as the aspartate rotates inwards between these two states, consistent with MD simulations<sup>85</sup>. Still, however, no report has addressed the molecular basis of substrate specificity in elevator-type transporters<sup>82</sup>.

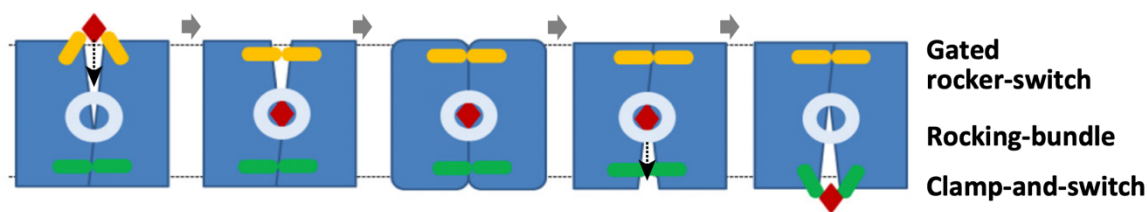


**Figure 1.5:** Cartoon representation of the elevator mechanism. As red diamonds are represented the solute/substrate, as yellow capsules the outward-gating elements, as green capsules the inward-gating elements<sup>76</sup>.

### 1.6.3 The rocking-bundle mechanism

In the rocking-bundle mechanism (**Figure 1.6**), substrate binds between two structurally distinct domains, and is directly coupled with the movement of outer and inner gates around the centrally located substrate binding site<sup>86</sup>. The amino acid/sodium symporter LeuT, an NSS-family homolog from *Aquifex aeolicus*, is the prototype of the rocking-bundle alternating-access mechanism. The LeuT fold is common, in many different transporter families that share the same architecture<sup>28,87-90</sup>. Three LeuT structures in different conformations: the outward-open, the inward-open<sup>91</sup> and the intermediate occluded structure highlighted essential aspects of the transport mechanism. In the LeuT fold, two distinct bundles are formed, namely the scaffold domain that consists of TMS3-TMS4 and TMS8-TMS9 and the core domain that consists of TMS1-TMS2 and TMS6-TMS7<sup>28</sup>. The scaffold and core domains are linked by the helices TMS5 and TMS10. Initially, the transporter is open to the extracellular, with its binding site accessible to the ligand. Proper binding of the substrate leads to the transition from the outward-open to the occluded state, when TMSs 1b and 6a move towards TMS3 and TMS10 and when TMS2 and TMS7 bend. The extracellular loop 4 (EL4), closes the outer cavity and is tightly packed with the core and the scaffold domain. These changes seem to precede the rocking movements of the core bundle that leads to the inward-occluded. TMS1a enables the cytosolic release of the substrate after a significant outward movement, leading to the inward-open conformation of the transporter. The gating on the extracellular side in LeuT when conformation changes from outward to occluded, there is an evolutionarily conserved salt bridge which is considered very important. LeuT and other same fold transporters like Mhp1, and vSGLT underwent molecular dynamics simulations, raising the idea of that the intracellular release of sodium and substrate facilitate conformational rearrangement towards the inward-open state. Crystal structures of LeuT-fold transporter BetP in all major conformations, indicate that changes in sodium coordination at the Na2 site during the transport cycle regulate substrate binding and dissociation. Remarkably, LeuT fold transporters diverge in terms of which of the two structural domains moves the most and by how much. The substrate-dependent local gating alterations also vary. For example, in the Mhp1 (hydantoin–sodium symporter), sodium binds only at the Na2 site but not at the Na1 site. Apparently, substrate binding induces gating movements of the extracellular TMSs different from those described for LeuT, involving TMS 10, which closes over the substrate binding pocket. Moreover, in Mhp1 the scaffold, also called hash domain, moves relatively to the mostly rigid core/bundle domain. A major difference between LeuT and Mhp1 is that the

substrate release to the intracellular is facilitated by the movements of TMS5, rather than TMS1a. Transporters with LeuT-fold vary significantly in the types of substrates and co-transported solutes. Additionally, many of the transporters belonging to NSS superfamily, like DAT, possess a secondary substrate-binding site, which serves in the substrate release in the cell, implying a regulatory role<sup>92</sup>. It has been also found that in LeuT, both substrates and noncompetitive inhibitors bind at the same extracellular site<sup>93</sup>. The contribution of gates in substrate specificity has been shown in rocking-bundle transporters<sup>82,94-96</sup>.



**Figure 1.6:** Cartoon representation of the rocking-bundle/rocker-switch mechanism. As red diamonds are represented the solute/substrate, as yellow capsules the outward-gating elements, as green capsules the inward-gating elements. The figure is adapted from<sup>76</sup>.

## 1.7 Approaches to the study of structure-function relationships in transporters

X-ray crystallography defines the structures of proteins with high resolution but being technically demanding<sup>97</sup> very few transmembrane protein structures have been resolved in atomic level (0.5-1.5Å)<sup>98</sup>. An exponential increase of the high-resolution structures arose due to the advances in the biochemistry of transmembrane proteins in conjunction with evolution of Cryo-electron microscopy (Cryo-EM)<sup>99</sup>, while new techniques<sup>100</sup> have been also employed for the analysis of the structure of transmembrane proteins, such as the single-domain antibodies (nanobodies)<sup>101</sup> and crystallography in a lipid environment with nanodiscs<sup>102</sup>. To compensate to some extent for the small number of resolved structures in secondary active transporters, we take advantage of their common structural characteristics to cluster homologous transporters according to structural features from crystallographic data<sup>103</sup>. This is rather important for transporters for which the production of sufficient quantities of building blocks of recombinant proteins is technically difficult. Chemical libraries are scanned in silico and based on a structural model, the transporters are identified as known or new ones<sup>104</sup>. The crystal structures usually correspond to one and rarely to more than one configuration corresponding to different steps of the transport cycle. However, by using homologous transporters to construct homology models and by combining this with state-of-the-art

computational techniques and the increasing computational power, then the whole simulation of the transport cycle could be achieved<sup>98,105–107</sup>. In these simulations small molecules can be included as well as the lipid membrane and reveal their role in the conformational rearrangement<sup>108–110</sup>. The most recent and highly successful, regarding the current data, method is the one that uses Artificial Intelligence and Machine Learning approaches to predict the 3D structure of molecules. This has been recently shown to be able to predict also other conformations of the protein<sup>111</sup>.

In this endeavor and especially for the execution of complex genetic approaches and rigorous biochemical techniques, the use of standard microorganisms is a significant advantage in the experimental methodology<sup>63</sup>. In order to functionally characterize a transporter, a combination of genetic, biochemical approaches and tests for uptake of radiolabeled substrates into fungi, bacterial or other cells are used. In the studies of the structure-function relationships of transmembrane proteins is mainly used systematic site-directed mutagenesis, but also cysteine scan mutations<sup>112</sup> (Cys-scanning mutagenesis), as a more efficient approach. Using Cys-scanning mutagenesis, it is possible to identify positions in the sequence important for function<sup>113</sup> and to indicate the structural organization of transmembrane segments<sup>114</sup> or the conformational change mechanism during substrate binding<sup>115</sup>. The prediction of the operation / specialization of a carrier based on the primary (or secondary) structure is not an easy task. It is not necessary that high-homogeneity related transporters recognize and / or transport the same substrates, whereas phylogenetically distant transporters have been shown to be able to display overlapping specialization profiles.

Studies on lactose permease (LacY) have shown that transporters with characteristic similarity in the binding site amino acid sequence, show unexpectedly significant differences in the substrate specificity<sup>116</sup>. Since the prediction of the function and specialization is not always reliable, knowledge of structure-function relationships is obtained mainly from mutagenesis studies and biochemical analyzes in prototypical microorganisms, but also from crystallographic structures or in silico structural analyzes (e.g. homology modeling, molecular dynamics).

## **1.8 Gating mechanism and specificity in transporters**

It is known that each transporter is substrate-specific and the transport catalysis occurs only when the appropriate ligand is bound in the binding site. The exact mechanism describing the transporter specificity or the reason why ligands with similar structure can bind, but are not

transported, or are not recognized even if they share high similarity with the physiological substrate is quite unclear.

An emerging theory is that if a substrate/ligand inhibits one of the alternative conformations of the transporter, transport cannot be executed. Specifically, in the case of LeuT, a crystal structure was obtained with Trp, a competitive inhibitor of LeuT that binds with high affinity but is not transported, bound in the outward open conformation<sup>117</sup>. In the presence of all amino acids transported by LeuT the crystals obtained were in an occluded conformation, but only Trp resulted in outward open conformation<sup>117</sup>. Moreover, a secondary binding site of lower affinity was suggested as in the open conformation a second Trp molecule was bound ~4 Å above the first Trp, closer to the extracellular vestibule<sup>117</sup>.

Additionally, the existence of multiple binding sites was suggested for the human glucose facilitator (GLUT1)<sup>118,119</sup>. In detail, computational studies propose that glucose might bind in multiple binding sites with different affinities, sliding in GLUT1 cavity until it is released in the cytoplasm or the periplasm, using a channel-like mechanism. Until now, there are not any experimental data supporting this mechanism.

The crystal structure of Mhp1 complexed with a bulky hydantoin analogue (NMH) was captured in the open outward conformation<sup>120</sup>, similarly to LeuT-Trp case. Comparing different conformations of Mhp1, rigid body movements occur after substrate binding and TMS10 closes towards TMS1, shielding the binding cavity and leading from the outward open to the occluded conformation. Compared with other analogues that could be transported and were captured bound in occluded conformations, NMH was a competitive inhibitor that was predicted to clash with a Leu residue of TMS10 inhibiting closure. NMH was transported after the substitution of the specific Leu with an A1. Also, in the case of a double mutant of LacY, replacement of two Gly residues with Trp at the termini of TMS2 and TMS8 led to abolishment of transport, but retained substrate binding<sup>121</sup>. A crystal structure of this mutant revealed an almost occluded structure, which is partially open to the periplasmic side and tightly closed to the cytoplasmic side<sup>122</sup>. The Trp substitutions seem to inhibit closure and subsequently the conformational change from one structure to the other<sup>122</sup>.

The closure of the extracellular gate and occlusion of the binding site from both sites is proposed as a necessary step in order to transport a molecule successfully by the computational and structural studies mentioned above, as well as on other transporters, e.g. the dopamine transporter from *Drosophila melanogaster* (DAT)<sup>123</sup> or the human serotonin transporter (SERT)<sup>124</sup>. The question arising is whether are able to recognize their physiological substrates among all other compounds.

## 1.9 Residues important for specificity and transportation in UapA transporter

Mutations rationally designed or randomly selected acquired via genetic screens in UapA modified transport kinetics (i.e.,  $K_m$  and apparent  $V_{max}$ ) and, interestingly, led to specificity changes<sup>1,8,125–128</sup>. Surprisingly, specificity mutations are mostly located in the scaffold domain and not in the core domain which also contains the binding site. In particular, these mutations are located at the interface overlooking the trajectory of the sliding elevator. Most of the specificity mutations, enable UapA to transport a range of purines, pyrimidines, and a plethora of nucleobase analogs enlarging specificity, albeit with very low binding affinities. Specificity mutations did not affect the high binding affinity for physiological substrate, although they increased moderately the apparent transport rate of UapA. In addition, combinations of specificity with substrate binding mutations in the core domain proved functionally additive, resulting in novel specificities and transport kinetics<sup>1,8,125,128</sup>. Due to these observations it has been proposed that specificity mutations affect residues acting as selectivity filters or gating elements along the substrate translocation trajectory<sup>8,76,125–127</sup>. Still the mechanistic details determining specificity remain elusive as there is very little structural data for NAT transporters<sup>82</sup>.

### Aim of the thesis

The mechanism under which nucleobase transporters and in general, transmembrane transporters operate can be investigated in the cases that the tertiary structure is determined in multiple conformations. However, to obtain a crystal structure is very difficult for transmembrane systems, especially because the system is constantly in conformational change making it difficult to understand the functionalities of the system. Crystal structures have been resolved only for UraA, UapA, Mhp1.

In this study, the first goal was to shed light on the structural alterations during the binding and translocation of uracil-substrate and the symporting proton of *Aspergillus nidulans*' nucleobase transporter FurE, as well as to elucidate the role and the order of the internalization of both uracil and the proton. This was addressed using state-of-the art computational techniques.

In addition, we aimed to introduce the outward-open structure of the dimeric UapA transporter and reveal important transport details of the elevator mechanism, using Molecular Dynamics and Metadynamics approaches to reproduce the movement from the Occluded to the

Inward-Open state. We also aimed to characterize the role of important residues, indicated by mutational analysis, involving their contribution to the conformational change of the transporter.

Moreover, we tried to gain insights into the specificity regulation of these transporters by investigating computationally the ligand binding and rationalizing mutational data for UapA from Prof. Diallinas' lab and for the homologous transporters: XanQ, AncXanQ, SmLL9 from Prof. Frillingos' lab. These transporters haven't been resolved and belong to fungi and bacteria.

## 2 Methods

### 2.1 FurE

#### 2.1.1 FurE protein model construction

The model of the FurE transporter was constructed based on homology modeling using Prime 2018-4 (Schrödinger, LLC, New York, NY, 2018) on Maestro platform (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018). As a template structure was used Mhp1 in the three crystal structure conformations provided in <https://www.rcsb.org/> namely, OO (2JLN), Occ (4D1B), IO (2X79). These crystal structures share a 35% similarity with FurE. The sequence alignment was formulated according to previous work<sup>94</sup> using HHpred and is presented in **Figure 3.1.2**. Further, in the case of 2X79 Mhp1 IO crystal structure a part of TMS9 was represented as coil. In order to fix TMS9 we started with the OO FurE structure and using Targeted Molecular Dynamics (plumed-v2 software),<sup>129</sup> a constant force of 500000 kj/(mol\*nm<sup>2</sup>) was applied on the Ca atoms of the helices resulting in the occluded conformation of the FurE and then in the inward conformation. The constant force value was gradually turned to zero and each model system was subjected to stabilization.

#### 2.1.2 Construction of the H<sub>3</sub>O<sup>+</sup> cation

The H<sub>3</sub>O<sup>+</sup> topology was provided from Bryce group<sup>130</sup> while H<sub>3</sub>O<sup>+</sup> was prepared using Antechamber<sup>131</sup> and the general Amber force field.<sup>132</sup>

### 2.1.3 FurE System Setup

The protein-ligand complex construction was initiated using the CHARMM-GUI<sup>133</sup> platform. Each model was inserted into a heterogeneous fully hydrated  $120 \text{ \AA} \times 120 \text{ \AA} \times 120 \text{ \AA}$  bilayer, consisting of YOPC, POPI lipids and ergosterol at a ratio of 40:40:20 accordingly. The membrane embedded system was solvated with TIP3P explicit water molecules. The solution contained neutralizing counter ions and 150 mM  $\text{Na}^+$  and 150 mM  $\text{Cl}^-$ . In the case that  $\text{H}_3\text{O}^+$  was present, a water molecule was replaced with a  $\text{H}_3\text{O}^+$ , and the system was neutralized by having one  $\text{Na}^+$  counter ion less. The assembled simulation system consisted of  $\sim 160,000$  atoms.

### 2.1.4 Molecular Dynamics (MD)

All simulations were conducted using GROMACS software, version 2019.2<sup>134</sup>. CHARMM36m<sup>135</sup> force field was chosen for protein and lipids,  $\text{H}_3\text{O}^+$  was provided from Bryce group<sup>130</sup> while the ligand and  $\text{H}_3\text{O}^+$  were prepared using Antechamber<sup>131</sup> and the general Amber force field.<sup>132</sup> The protein orientation into the membrane was calculated using the OPM database of the PPM server<sup>136</sup>. All model systems were minimized and equilibrated to obtain stable structures. Minimization was carried out for 5,000 steps with a step size of 0.001 kJ/mol applying a steepest descent followed by a conjugate gradient algorithm, and the system was equilibrated for 20ns by gradually heating and releasing the restraints to expedite stabilization. Finally, the system proceeded to further simulations free of restraints at a constant temperature of 300K using Nose-Hoover thermostat,<sup>137</sup> the pressure was kept constant at 1 bar using Parrinello-Rahman semi-isotropic pressure coupling<sup>138</sup> and compressibility at  $4.5\text{e-}5 \text{ bar}^{-1}$ . The van der Waals and electrostatic interactions were smoothly switched off at 1.2 nm, while long-range electrostatic interactions were calculated using the particle mesh Ewald method.<sup>139</sup> All bonds were constrained using the LINCS algorithm,<sup>140</sup> allowing a time-step of 2.0 fs. The trajectories were further examined for structural stability by RMSD calculation of protein Ca (up to 1.2  $\text{\AA}$ ) and by visual inspection with VMD platform<sup>141</sup> thus ensuring that the thermalization did not cause any structural distortion.



### 2.1.5 Metadynamics Simulations

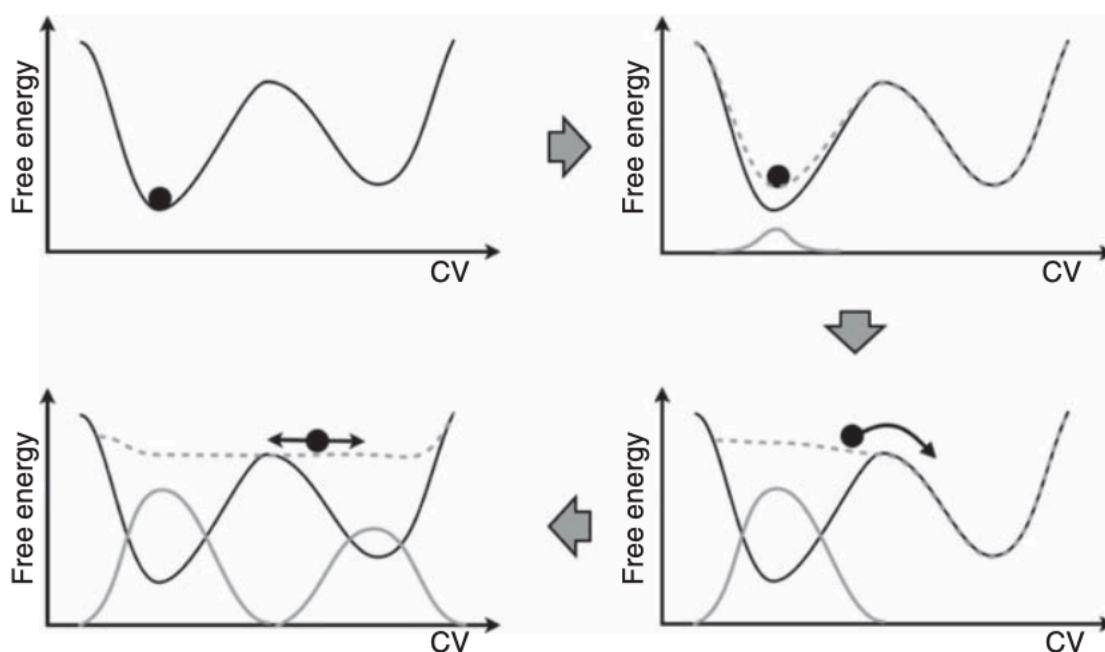
Molecular dynamics (MD) and the Monte Carlo simulation method have had a very deep influence on diverse fields, especially to pharmacology. Yet, these simulation methods suffer from limitations such as limited time scale explored by the present-day computer technology and sampling algorithms which result in reducing the scope of their applications. MD is used for sampling the canonical distribution having as important aspect the ergodic hypothesis. According to this, if a system is simulated long enough all the states relating to the canonical ensemble will be explored, where each one has its own correct statistical weight. Unfortunately, this hypothesis cannot be proven for most of the systems and, even when verified, the length of a simulation necessary in calculating ensemble averages is often unreachable for numerical calculations. This led to the development of several enhanced sampling algorithms that were designed to alleviate this difficulty in the past years. Specifically, such cases where the free energy surface (FES) has several local minima separated by large barriers are the conformational changes in solution, protein folding, first-order phase transitions, and chemical reactions. In such circumstances it is almost impossible for a simulation started in one minimum to be able to move spontaneously to the next minimum and it can happen only under very favorable circumstances. Metadynamics<sup>142-145</sup> stand as a solution to these problems allowing the system to escape from local minima of the FES and explore the 3D space while at the same time permit a quantitative determination of the FES as a result of the integrated process. In Metadynamics, the FES is systematically explored by the metadynamics bias potential, which is adaptively constructed as a sum of Gaussian functions centered on the explored points in the collective variables (CVs) space along the system trajectory in a history depended manner. CVs are degrees of freedom defined by the user and have a preassigned width ( $\sigma$ ), which needs to be fixed for each of the CVs ( $\sigma_\alpha$ ), and height ( $w_G$ ) and are deposited every  $\tau_G$  time units as the simulation proceeds. The bias potential at time  $t$  is:

$$V(s, t) = \sum_{i=1}^{t/\tau_G} w_G \exp \left[ - \sum_{\alpha=1}^{N_{CV}} \frac{(s_\alpha - s_\alpha(i\tau_G))^2}{2\sigma_\alpha^2} \right]$$

This potential forces the system away from the kinetic traps in the potential energy surface and out into the unexplored parts of the energy landscape. The free energy can be reconstructed from a metadynamics calculation because the final bias is given by:  $V(s^\rightarrow) = -F(s^\rightarrow)$ .

The  $w_G$  and  $\tau_G$  parameters are not independent, and if they are chosen within meaningful ranges, what really matters is their rate  $\omega = w_G/\tau_G$ , also known as the deposition rate. Equivalently, it is possible to state that the bias potential  $V(s)$  evolves with time according to the following equation of motion. In the limit of small  $\tau_G$ , the expression for the history-dependent bias is:

$$\dot{V}(s, t) = \omega \exp \left[ - \sum_{\alpha=1}^{N_{CV}} \frac{(s_{\alpha} - s_{\alpha}(t))^2}{2\sigma_{\alpha}^2} \right]$$



**Figure 2:** A sketch of the process of metadynamics. First the system evolves according to a normal dynamics, then a Gaussian potential is deposited (solid gray line). This lifts the system and modifies the free-energy landscape (dashed gray line) in which the dynamics evolves. After a while the sum of Gaussian potentials fills up the first metastable state and the system moves into the second metastable basin. After this the second metastable basin is filled, at this point, the system evolves in a flat landscape. The summation of the deposited bias (solid gray profile) provides a first rough negative estimate of the free-energy profile. This figure is adapted from<sup>146</sup>

In this study we used Path Collective Variable Metadynamics (PCV) and Funnel Metadynamics (FM).

Path Collective Variable Metadynamics (PCV) is a dimensionality reduction approach, which can study complex chemical and biological processes involving a large number of degrees of freedom in one or two collective variables. Two functions are introduced to describe the position of a point in the configurational space relative to a preassigned guessed path: the

progress along the path ( $R(s)$ ) and the distance from it ( $R(z)$ ) mathematically described as follows:

$$s = \frac{\sum_{i=1}^N i \exp(-\lambda R[X - X_i])}{\sum_{i=1}^N \exp(-\lambda R[X - X_i])}, \quad z = -\frac{1}{\lambda} \ln \left[ \sum_{i=1}^N \exp(-\lambda R[X - X_i]) \right].$$

where  $X$  represents the atomic coordinates at the current simulation time-step, while  $X_i$  represents those of the  $i$ -th snapshot. The function  $R$  represents usually RMSD, which measures the distance between configuration states. The  $\lambda$  parameter smooths the variation of the  $s$  variable.

Funnel-Metadynamics (FM) accelerates the sampling of ligand-protein binding (LPB) process allowing the identification of the ligand binding mode to its binding site and the accurate estimate of the absolute protein-ligand binding free energy. In FM, the binding free energy surface (BFES) is thoroughly explored by the metadynamics bias potential, while as  $cv$  is usually considered the distance between the ligand and the assumed binding pocket. During the simulation, a funnel-shape restraint potential is applied to the protein which is a combination of a cone restraint, including the ligand binding site, and a cylindric part directed toward the solvent. If the ligand is within the funnel area, no repulsive bias is applied on the system and the simulation proceeds as standard metadynamics. As the ligand reaches the edge of the funnel, a repulsive bias is applied to the system preventing it from visiting areas out of the funnel. Subsequently, the sampling at the binding site is not affected by the external bias, while in the bulk water the repulsive potential decreases the space to explore to a cylindric region. This way it is possible to observe multiple binding/unbinding events leading to a fast convergence of the results. When FM simulation is converged, the ligand binding mode corresponds to the lowest free-energy minimum in the BFES, while the absolute protein-ligand binding free energy  $\Delta G^0$  is computed from the free-energy difference between the bound and unbound state, depending solely on the free-energy value at the two states and is independent from the path that connects these two states. Mathematically, the absolute protein-ligand binding free energy  $\Delta G^0$  is computed as  $\Delta G^0 = -k_B T \ln(C^0 K_b)$ , where  $K_b$  is the equilibrium binding constant calculated as follows:

$$K_b = \pi R_{\text{cyl}}^2 \int_{\text{site}} dx e^{-\frac{1}{k_b T} [W(x) - W_{\text{ref}}]}$$

where  $\pi R_{\text{cyl}}^2$  is the surface of the cylinder in the unbound state, while the potential  $W(x)$  and its value in the unbound state ( $W_{\text{ref}}$ ), are derived from the potential of mean force obtained through FM calculations.

FM has many pros as a computational method for LPB prediction. FM aim at reproducing the physical binding pathway of the ligand toward the protein target, allowing the sampling of the binding process without knowing a priori the binding mode and site of the ligand, thus, binding area can be approximately hypothesized, unlike end-point methods (Linear Interaction Energy method<sup>147</sup>, MM-PBSA/MM-GMSA<sup>148,149</sup>) and alchemical transformation techniques (Free Energy Perturbation<sup>150</sup>) that demand knowing in advance the ligand binding mode. It also considers the full flexibility of the ligand, the protein and the water molecules, hence being able to include the dynamics of waters during the process and to consider the positional, conformational, orientational entropy of both protein and ligand. FM has affordable computational cost for a rare event such as the binding process, compared to standard MD calculations. Finally, FM is self-diagnostic, connotating that the free energy calculation converges only to the exact value, which is a critical advantage of FM, and metadynamics in general. The method has proven to be successful in reproducing binding processes in ligand/protein and ligand/ DNA systems, as it effectively predicts crystallographic binding modes and experimental binding free energies.

For the metadynamics simulations the plumed-v2 software was used.<sup>129</sup>

### ***Funnel-Metadynamics for the H<sub>3</sub>O<sup>+</sup> cation.***

a) The FurE transporter used was in outward-open (OO) conformation. Since no data are available concerning the binding site of the H<sub>3</sub>O<sup>+</sup> cation, a wide area around the equivalent one of the Na<sup>+</sup> cation in Mhp1 was circumvented by the funnel cone. The cone's starting point was T332 Ca, while the cylinder had a direction towards the extracellular waters. The funnel had a switching point between the cone and cylinder region at 4.0 nm, the amplitude of the cone was 0.27 rad, the radius of the cylinder section for the unbound region at 0.1 nm, the minimum and maximum value sampled as projection of the ligand's center of mass (COM) along the funnel axis was at 0.25 and 4.6 nm respectively, as long as, the lowest and highest value for fps.lp

used to construct the funnel-shape restraint potential was at 0.00 and 4.8 nm respectively. The value for the spring constant of the funnel-shape restraint potential was set to 5000  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$ . As collective variable (CV) was selected the distance between the Ca atom of T332 and the center of mass of  $\text{H}_3\text{O}^+$  cation. The width of the Gaussian functions was set to 0.05 nm, the height to 2  $\text{kJ}/\text{mol}$  and the deposition stride to 500 simulation steps. The rescaling factor of the Gaussian function's height was 20 as we performed well-tempered metadynamics.

b) The study of binding/unbinding process of the  $\text{H}_3\text{O}^+$  cation in the cytoplasmic solvent was initiated by using well-tempered metadynamics with the FM method on FurE in the Occ state of FurE. Uracil was included in the system, placed at the binding site, whereas an upper wall of 20000  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$  enforced the COM in distances lower than 0.9 nm from the Cg of N341. The constructed funnel included all the possible routes that could lead the  $\text{H}_3\text{O}^+$  to the exit to the cytoplasm. The cone region started at Ca of S56. The direction of the funnel axis was cytoplasm-oriented passing through D348 Cb atom. The switching point between the cone and the cylinder region was at 3.6 nm, the amplitude of the cone section was set to 0.5 rad and the radius of the cylinder for the unbound region to 0.1 nm. The minimum and maximum value sampled as projection of the ligand's COM along the funnel axis was set to 0.29 and 4.2 nm respectively, the lowest and highest value for  $\text{fps.lp}$  used to construct the funnel-shape restraint potential was set to 0.2 and 4.3 nm respectively, while, the value for the spring constant of the funnel-shape restraint potential to 7000  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$ . As CV was selected the distance between the Ca of E51 and the center of mass of  $\text{H}_3\text{O}^+$  cation. The width of the Gaussian functions was calculated at 0.01 nm, the height was arranged at 2  $\text{kJ}/\text{mol}$  with a rescaling factor of the Gaussian function at 20 and the deposition stride was set to 500 simulation steps.

### ***Funnel-Metadynamics for the substrate (Uracil)***

a) FM were performed aiming to highlight the binding mode of uracil in the binding site and the binding mechanism as it approaches the binding pocket from the extracellular. FurE was in the occluded state and the  $\text{H}_3\text{O}^+$  cation was included in the system, as in crystallographic results of other transporters, in particular Mhp1, ion and substrate co-exist in the Occ state. In detail, the funnel used, oriented from Ca atom of V323 deep in the binding area, with direction of the funnel axis to the extracellular solute. The switching point between the cone and cylinder region was set to 2.7 nm, the amplitude of the cone section to 0.37 rad, the radius of the cylinder section for the unbound region to 0.1 nm, the minimum and maximum value sampled as projection of the ligand's COM along the funnel axis to 0.2 and 3.3 nm respectively, the lowest

and highest value for  $\text{fps.lp}$  used to construct the funnel-shape restraint potential to 0.05 and 3.6 nm respectively. The value for the spring constant of the funnel-shape restraint potential was 30000  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$ . As CV was selected the distance between the Ca of N341 and the center of mass of uracil. The width of the Gaussian functions was 0.01 nm, the height was arranged at 2  $\text{kJ}/\text{mol}$  and the deposition stride at 500 simulation steps. The rescaling factor of the Gaussian function's height was 20.

b) The uracil internalization process was implemented using again, well-tempered metadynamics with the FM method, on FurE transporter in IO conformation containing uracil and not  $\text{H}_3\text{O}^+$ , as the latter was already proved from PCV simulations that leaves first the transporter in order to allow uracil to exit too (see Main Text). The funnel was constructed as to include all the possible exiting pathways from the binding pocket to the TMS5 outer gate. The cone restraint started at backbone C atom of F53, while the direction of the funnel axis was cytoplasm-oriented passing through S342 O atom. The switching point between the cone and the cylinder region was set to 3.4 nm, the amplitude of the cone section to 0.49 rad and the radius of the cylinder for the unbound region to 0.1 nm. The minimum and maximum value sampled as projection of the ligand's COM along the funnel axis was set to 0.21 and 4.1 nm respectively, the lowest and highest value for  $\text{fps.lp}$  used to construct the funnel-shape restraint potential was set to 0.1 and 4.2 nm respectively, while, the value for the spring constant of the funnel-shape restraint potential was set to 30000  $\text{kJ}/(\text{mol}\cdot\text{nm}^2)$ . As CV was chosen the distance between the backbone of A50 and the center of mass of uracil. The width of the Gaussian functions was calculated at 0.01 nm, the height was arranged at 2  $\text{kJ}/\text{mol}$  with a rescaling factor of the Gaussian function at 20 and the deposition stride was set to 500 simulation steps.

### ***Metadynamics Simulations with Path Collective Variable (PCV):***

a) OO-to-Occ path: In this case we used the  $\text{C}\alpha$  atoms of the residues belonging to FurE helices involved in hash and bundle motif. This choice was found to be appropriate because the calculated FESs were well reproducible. The initial path was obtained through a carefully chosen set of frames with equally distant RMSDs, derived from a steered MD simulation where the OO FurE was biased to Occ conformation using a stable force on  $\text{C}\alpha$  atoms of helices. 6 frames were used to construct the path in total, while the average distance between adjacent frames was 0.13 nm. The RMSD matrix was constructed and plotted, confirming that the frames were appropriate for the calculation. The  $\lambda$  value calculated for s was equal to 200  $\text{nm}^2$ . The width of the Gaussian functions for hills deposition was 0.035  $\text{nm}^2$  based on the

structure fluctuation in unbiased MD, the height was arranged at 0.5 kJ/mol and the deposition stride at 500 simulation steps. An upper wall of 500000 kJ/(mol\*nm<sup>2</sup>) was set to constrain the distance from the path at a value lower than 0.06, based in unbiased MD simulations of more than 200 ns where the cv's fluctuation did not reach values higher than 0.03. If uracil is part of the system, it is constrained in the previously calculated position in the binding site with a distance restraint of 20000 kJ/(mol\*nm<sup>2</sup>) at 0.7 nm between the center of mass of the substrate and Cd atom of Q134. The same constraint was applied on the distance of H<sub>3</sub>O<sup>+</sup> cation from Cd atom of E51 at 0.45 nm.

b) The same rationale and method were used in the Occ-to-IO case. Here, the  $\lambda$  value for s was equal to 110 nm<sup>-2</sup>, the width of the Gaussian functions for hills deposition was 0.037 nm<sup>2</sup>, the upper wall of 500000 kJ/(mol\*nm<sup>2</sup>) was set to constrain the z at a value lower than 0.1.

## **2.2 XanQ and SmLL9 homolgy models' construction**

Homology model building was performed using Prime 2018-4 (Schrödinger, LLC, New York, NY, 2018)<sup>151,152</sup>. The alignments were done using HHpred as presented in **Figure 3.2.1.1**.

### **2.2.1 Protein preparation**

The protein was prepared for the IFD calculations using the Protein Preparation Protocol (Schrödinger Suite 2018, Protein Preparation Wizard) implemented in Schrödinger suite and accessible within the Maestro program (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018). Specifically, the hydrogen atoms were added. The orientation of hydroxyl groups of Asn, Gln, and the protonation state of His were optimized to maximize hydrogen bonding. Finally, the ligand – protein complex was refined with a restrained minimization performed by Impref utility, which is based on the Impact molecular mechanics engine (Impact, Schrödinger, LLC, New York, NY, 2016) and the OPLS2005 force field, setting a max rmsd of 0.30. Ligand preparation for docking was performed with LigPrep (LigPrep, version 2018-4, Schrödinger, LLC, New York, NY, 2018) application which consists of a series of steps that perform conversions, apply corrections to the structure, generate ionization states and tautomers, and optimize the geometries.

### 2.2.2 Induced-fit docking

The IFD protocol (Schrödinger Suite 2018-4 IFD protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2018), is intended to circumvent the inflexible binding site and accounts for the side chain or backbone movements, or both, upon ligand binding. In the first stage of the IFD protocol, softened-potential docking step, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018-4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues having at least one atom within 8 Å of an atom in any of the 20 ligand poses. The Prime refinement starts with a conformational search and minimization of the side chains of the selected residues and after convergence to a low-energy solution, an additional minimization of all selected residues (side chain and backbone) is performed with the truncated-Newton algorithm using the OPLS parameter set and a surface Generalized Born implicit solvent model. The obtained complexes are ranked according to Prime calculated energy (molecular mechanics and solvation), and those within 50 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. In the final round, the ligands used in the first docking step are redocked into each of the receptor structures retained from the refinement step. The final ranking of the complexes is done by a composite score which accounts for the receptor – ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

### 2.2.3 Molecular Dynamics simulations

The XanQ and SmLL9 models were minimized and equilibrated further to obtain the stable structures. Each model was inserted into a heterogeneous fully hydrated bilayer  $120 \text{ \AA} \times 120 \text{ \AA} \times 120 \text{ \AA}$ , consisting of DPPC lipids using CHARMM-GUI and CHARMM36m<sup>135</sup> force fields for lipids as well as protein. The protein orientation into the membrane was calculated using the PPM server<sup>136</sup>. The structures were then solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. The assembled simulation system consisted of ~119,000 atoms. The biomolecular systems were simulated using GROMACS<sup>134</sup> software. Minimization was carried out for 2,000 steps with a step size of 0.001 kJ/mol applying a steepest descent and then, a conjugate gradient algorithm, and system was equilibrated for 15 ns while gradually heating and releasing the restraints to



expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 300K for 40 ns using Nose-Hoover thermostat<sup>137</sup> and Parrinello-Rahman<sup>138</sup> semi-isotropic pressure coupling and compressibility at 4.5e-5 bar-1. The van der Waals interactions were smoothly switched off at 1.2 nm by switching function as electrostatics, while long-range electrostatic interactions were calculated using particle mesh Ewald method.

## **2.3 AncXanQ**

### **2.3.1 Protein model construction**

A homology model of XanQ has already been constructed in an inward-open conformation based on the UapA structure 5I6C (see paragraph 2.2.1 of Materials and Methods). The XanQ model was used as a template to generate AncXanQ model by Homology Modeling using Prime 2018–4 (Schrödinger, LLC, New York, NY, 2018) on Maestro platform (Maestro, version 2018–4, Schrödinger, LLC, New York, NY, 2018). The two transporters, AncXanQ and XanQ, share 76% sequence identity and their alignment is depicted in **Figure 3.2.2.1** as derived from HHpred.

### **2.3.2 Protein preparation**

The modeled transporter was prepared using the Protein Preparation Protocol implemented in Schrödinger suite (Schrödinger Suite 2018, Protein Preparation Wizard) which is accessible within the Maestro suite (Maestro, version 2018–4, Schrödinger, LLC, New York, NY, 2018). The preparation procedure starts by adding hydrogen atoms. Then, follows the optimization of the orientation of hydroxyl groups of Gln and Asn and of the protonation state of His guided by the maximization of hydrogen bonding. The final step was energy minimization of the transporter, using the OPLS3 force field.

### **2.3.3 Ligand preparation**

Ligand preparation was performed with LigPrep application (LigPrep, version 2018–4, Schrödinger, LLC, New York, NY, 2018), which consists of a series of steps that perform conversions, apply corrections to the structure, generate ionization states and tautomers, and optimize the geometries. The force field used was OPLS3<sup>153</sup>.

### 2.3.4 Induced-fit docking

For Induced-Fit-Docking (IFD), Schrödinger Suite protocol was used (Schrödinger Suite 2018–4 Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2018), taking into account both the side chain or backbone movements, upon ligand binding. In the first softened-potential docking step of the protocol, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018–4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues within 5 Å of any out of the 20 ligand poses. The Prime refinement starts by performing conformational search and by minimizing the side chains of the selected residues. After convergence to a low-energy result, an additional minimization of all selected residues (side chain and backbone) is performed with the Truncated-Newton algorithm using the OPLS3 parameter set<sup>153</sup> and a surface Generalized Born implicit solvent model. The ranking of the obtained complexes is implemented according to Prime calculated energy (molecular mechanics and solvation), and the complexes within 30 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. Finally, the ligands used in the first docking step are redocked into each of the receptor structures retained from the Prime refinement step. The final ranking of the complexes is performed by a composite score which takes into account the receptor – ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

### 2.3.5 Molecular dynamics (MD)

In order to construct the protein–ligand complex we used the CHARMM-GUI platform. Each model was inserted into a heterogeneous fully hydrated bilayer 120 Å × 120 Å × 120 Å, consisting of DPPC lipids. The membrane embedded system was solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. CHARMM36m force field was used for protein and lipids, while the ligand was prepared using Antechamber and the general Amber force field<sup>132</sup>. The protein orientation in the membrane was calculated using the PPM server<sup>136</sup>. The assembled simulation system consisted of ~130,000 atoms. The systems were simulated using GROMACS software<sup>134</sup>. The models were energy-minimized and equilibrated to obtain stable structures. Minimization was carried out for 2000 steps with

a step size of 0.001 kJ/mol applying a steepest descent followed by a conjugate gradient algorithm, and the system was equilibrated for 20 ns by gradually heating and releasing the restraints to expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 300 K for 100 ns, using Nose-Hoover thermostat<sup>137</sup> and Parrinello-Rahman semi-isotropic pressure coupling<sup>138</sup> and compressibility at  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. The van der Waals and electrostatic interactions were smoothly switched off at 1.2 nm by switching function, while long-range electrostatic interactions were calculated using the particle mesh Ewald method<sup>154</sup>. The calculations performed were of 500 ns in time length. Each calculation was produced three times using different seed numbers.

## **2.4 UapA transporter and mutants (SYSE, TFAE, SYSQ)**

### **2.4.1 UapA-TFAE, UapA-SYSE and UapA-SYSQ Models Construction.**

Models of UapA-TFAE, UapA-SYSE and UapA-SYSQ were generated by mutating the specific residues on Maestro platform (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018), on the basis of the structure of UapA crystallized in the inward-open conformation (PDB ID: 5i6c). In order to construct the model of UapA outward conformation Band3 transporter (4yzt) was used as template. Band3 is an anion exchanger of transmembrane helices, crystallized in the outward conformation. The model was constructed using Prime Homology in Prime 2018-4 (Schrödinger, LLC, New York, NY, 2018), by superimposing helices one by one and aligning all the important residues.

### **2.4.2 Protein Preparation**

The protein was prepared using the Protein Preparation Protocol implemented in Schrödinger suite (Schrödinger Suite 2018, Protein Preparation Wizard) and accessible within the Maestro program (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018). Specifically, at first hydrogen atoms were added followed by an optimization of the orientation of hydroxyl groups of Asn, Gln, and of the protonation state of His to maximize hydrogen bonding. The final step was that of minimization of the protein, using the OLS3 force field<sup>153</sup>.

### **2.4.3 Ligand preparation**

Ligand preparation for docking was performed with LigPrep application (LigPrep, version 2018-4, Schrödinger, LLC, New York, NY, 2018), which consists of a series of steps that perform conversions, apply corrections to the structure, generate ionization states and tautomers, and optimize the geometries. The force field chosen was OPLS3.

### **2.4.4 Induced Fit Docking**

Induced Fit Docking Schrödinger Suite protocol was used (Schrödinger r Suite 2018-4 Induced Fit Docking protocol; Glide, Schro“dinger, LLC, New York, NY, 2016; Prime, Schro“dinger, LLC, New York, NY, 2018), taking into account the side chain or backbone movements, or both, upon ligand binding, thus circumventing an inflexible binding site. In the first softened-potential docking step, of the protocol, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018-4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues within 5 Angstrom of any out of the 20 ligand poses. The Prime refinement starts by performing conformational search and by minimizing the side chains of the selected residues. After convergence to a low-energy result, an additional minimization of all selected residues (side chain and backbone) is performed with the TruncatedNewton algorithm using the OPLS3 parameter set and a surface Generalized Born implicit solvent model. The ranking of the obtained complexes is implemented according to Prime calculated energy (molecular mechanics and solvation), and the complexes within 30 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. Finally, the ligands used in the first docking step are re-docked into each of the receptor structures retained from the Prime refinement step. The final ranking of the complexes is performed by a composite score which takes into account the receptor–ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

### **2.4.5 Molecular Dynamics Simulations**

For the construction of the protein–ligand complex CHARMM-GUI platform was used. Each model was inserted into a heterogeneous fully hydrated bilayer 120 Å 120 Å 120 Å,

consisting of DPPC lipids and ergosterol. The membrane embedded system was solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. CHARMM36m force field was used for protein and lipids, while the ligand was prepared using Antechamber and the general Amber force field. The protein orientation into the membrane was calculated using the PPM server<sup>136</sup>. The assembled simulation system consisted of ~130,000 atoms. The systems were simulated using GROMACS software. The models were minimized and equilibrated to obtain stable structures. Minimization was carried out for 2,000 steps with a step size of 0.001 kJ/mol applying a steepest descent followed by a conjugate gradient algorithm, and the system was equilibrated for 20 ns by gradually heating and releasing the restraints to expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 300 K for 100 ns, using Nose-Hoover thermostat<sup>58</sup> and Parrinello-Rahman semi-isotropic pressure coupling and compressibility at 4.5e-5 bar-1. The van der Waals and electrostatic interactions were smoothly switched off at 1.2 nm by switching function, while long-range electrostatic interactions were calculated using the particle mesh Ewald method.

#### 2.4.6 Metadynamics Simulations

The C $\alpha$  atoms of the residues belonging to UapA helices of both protomers was used to apply the bias potential, as the calculated FESs were well reproducible. The initial path was obtained through a carefully chosen group of frames with equally distant RMSDs, derived from a steered MD simulation where the IO UapA was biased to Occ conformation using a stable force on Ca atoms of helices and then to OO conformation. The average distance between the 5 adjacent frames chosen for the path was 0.1 nm. The RMSD matrix was constructed and plotted, confirming that the frames were appropriate for the calculation. The  $\lambda$  values calculated for each protomer were equal to 135, 136 nm<sup>2</sup>, respectively. The width of the Gaussian functions for hills deposition was 0.019 and 0.024 nm<sup>2</sup>, the height was arranged at 0.5 kJ/mol and the deposition stride at 1000 simulation steps. An upper wall of 500000 kJ/(mol\*nm<sup>2</sup>) was set to constrain the distance from the path at a value lower than 0.2, based on unbiased MD simulations of more than 200 ns where the cv's fluctuation did not reach values higher than this. Xanthine is constrained in the binding site with a distance restraint of 50000 kJ/(mol\*nm<sup>2</sup>) at 0.35 nm.

## 3 Results

### 3.1 Uracil/H<sup>+</sup> symport by the FurE transporter challenges the rocking-bundle mechanism of transport in APC transporters<sup>155</sup>

#### 3.1.1 Foundation of the study

Secondary active transporters are transmembrane proteins, fundamental for the transport of nutrients inside the cells. They recognize and translocate their substrates using the energy provided by the electrochemical gradient of the membrane. The transport also involves the symport or antiport of Na<sup>+</sup>/H<sup>+</sup> cations with other solutes. Though secondary active transporters could be functionally, structurally, or evolutionary distinct, they share common folds, associated to specific protein conformational changes linked with the transport cycle, according to structural studies. The typical model for the transport mechanism is identified as the “alternating-access model”. In detail, the transporter receives or releases the substrate by changing conformations from an *outward-open* (OO) state facing the extracellular environment to an *inward-open* (IO) state facing the cytoplasm.<sup>61,62,156–158</sup> Depending on the folding and specific conformational rearrangements of the transporter, three major mechanisms have been proposed, namely the rocker-switch, the rocking-bundle and the sliding-elevator.<sup>11,75,86,158,159</sup> Important structural and functional information about the rocking-bundle mechanism, which characterizes one of the largest transporter families, the so-called Amino Acid-Polyamine-Organocation (APC) Family (TC# 2.A.3), emerge from seminal studies on the bacterial leucine and alanine specific transporter LeuT.<sup>28,156,158</sup> LeuT adopts the 5+5 helical inverted repeat (5HIRT), shaped by transmembrane helices 1-10, whose structural elements and conformational changes determine substrate identification and transport. In total, LeuT and similar transporters acquire twelve  $\alpha$ -helical transmembrane segments (TMSs). Nonetheless, the role of TMS11 and TMS12 is not clarified yet. The rocking-bundle model assumes that substrate translocation follows the OO-to-IO conformational change and is facilitated by the relative movement between two motifs, the so-called *hash/scaffold* domain (TMS3, TMS4, TMS8, TMS9) and the *bundle/core* domain (TMS1, TMS2, TMS6, TMS7). TMS5 and TMS10 function as gates. It has been proposed that binding of the substrate in the OO conformation requires the simultaneous binding of a positive charge (Na<sup>+</sup> or H<sup>+</sup>) and induces the conformational change of the protein towards the IO conformation. This example of substrate translocation mechanism has been supported by studies on the eukaryotic dopamine (DAT)<sup>123</sup> and serotonin (SERT)<sup>160</sup> transporters (neurotransmitter/sodium symporter family-NSS), and a

number of prokaryotic transporters, including the MhsT transporter specific for hydrophobic L-amino acids (NSS family),<sup>161</sup> the sodium-galactose symporter vSGLT (sodium-solute symporter-SSS family),<sup>89</sup> the sodium-betaine symporter BetP<sup>88</sup> and CaiT<sup>162</sup> (betaine/choline/carnitine transporter-BCCT family), the AdiC<sup>163</sup> and ApcT<sup>90</sup> (amino acid/polyamine/organocation-APC family) and, finally, the bacterial benzyl-hydantoin/Na<sup>+</sup> symporter Mhp1<sup>87,120,164,165</sup> (nucleobase cation symporter 1 or NCS1 family). Topological variations have been found during the transition from the OO to the IO state, although all these proteins fold in the same way. Crystal structures of LeuT and MhsT, indicate that the bundle/core domain (TMS1, TMS2, TMS6, TMS7), undergoes substantial conformational changes during the OO/IO transition pivoting around the hash/scaffold motif. Moreover, there are two further rearrangements operating as opening/closing gates. Explicitly, in LeuT the TMS1b, TMS6a displacement, acts as an extracellular gate. Alongside, a kink of 45° of TMS1a followed by a limited unwinding of TMS5, serves as the intracellular gate. Contrastingly, the Mhp1 transporter transits from the OO to the IO state by pivoting the hash motif around the bundle domain. This also stimulates TMS10 to move towards TMS1b and TMS6a to enclose the substrate in the occluded conformation. Besides, TMS5 bending opens the inward facing cavity, differentiating from the TMS1a of LeuT.

Various fungal members of the NCS1 transporter family, related to the APC superfamily, have been extensively studied by Diallinas' group, revealing crucial information about substrate specificity, regulation of expression, subcellular trafficking and turnover, and transport kinetics.<sup>94-96,166-169</sup> Transporters of this family are H<sup>+</sup> symporters of high or low affinity selective for uracil, allantoin, cytosine, thiamine, uridine, or nicotinamide riboside and secondarily for xanthine and uric acid.<sup>166,167,170</sup> NCS1 transporters are present mostly in fungi and bacteria, hence illustrating the relative transport mechanism in atomic detail will set light on the complex mechanism of LeuT-type transporters, and may also provide valuable insight into discovering new antifungal or antibacterial drugs. In previous work, we used as a structural template the Mhp1 transporter, in order to construct the models of specific members of the NCS1 family. The results were assessed using functional data acquired by extensive mutational analyses. In these studies, we described the putative substrate binding sites in several NCS1 members of the fungus *Aspergillus*, namely the FcyB purine-cytosine transporter,<sup>168</sup> the FurA allantoin transporter,<sup>95</sup> the FurD uracil transporter,<sup>95</sup> and the FurE uracil-allantoin-uric acid transporter.<sup>95,96</sup> Recently, focusing on the role of the cytosolic N-and C-terminal segments of FurE, we showed the crucial role they play in regulating endocytic turnover, transport kinetics, and surprisingly in affecting substrate specificity.<sup>94,96</sup> Related to the specificity, using a

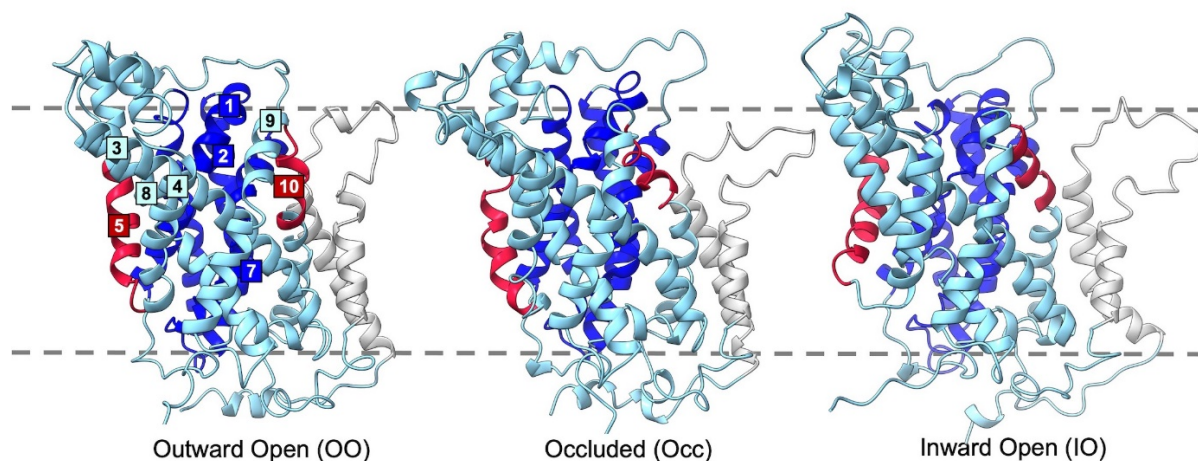
homology model we found that the N-terminus of FurE can control substrate specificity by interacting dynamically and in a pH-dependent manner with specific intracellular loops.<sup>94,96</sup> Remarkably, the N-terminal region appears to have the same role in the *Drosophila melanogaster* DAT (dDAT) and human SERT (hSERT), which are eukaryotic LeuT-type transporters.<sup>171</sup>

In this study, we aim to describe the functional conformational changes operated by FurE from the OO to the IO state at atomistic resolution. We explored the translocation pathway from the OO to the IO state through the occluded state (Occ) and illuminated the role and the order of the internalization of both uracil as substrate and H<sup>+</sup> cation which is considered in the form of hydronium H<sub>3</sub>O<sup>+</sup>. For that, we used innovative computational methods and additional mutational analyses, wisely designed to assess the *in silico* results. At first, we constructed the FurE homology models of the three conformational states based on the crystal structures of Mhp1, and performed metadynamics calculations,<sup>142</sup> a powerful enhanced-sampling technique able to reproduce slow-rate biological procedures at an reasonable computational expense. In particular, we implemented funnel-metadynamics (FM)<sup>143</sup> to determine the binding sites and translocation pathways of both H<sub>3</sub>O<sup>+</sup> and uracil substrate. FM is a variant of metadynamics deliberately developed to study ligand binding. Overall, we managed to characterize the large-scale conformational changes of the transporter from the OO to the IO state, including the transport of both the substrate and H<sub>3</sub>O<sup>+</sup>, thus providing an inclusive new portrait of the functional mechanism of FurE and in general, of NCS1 transporters.

### 3.1.2 The FurE 3D structure

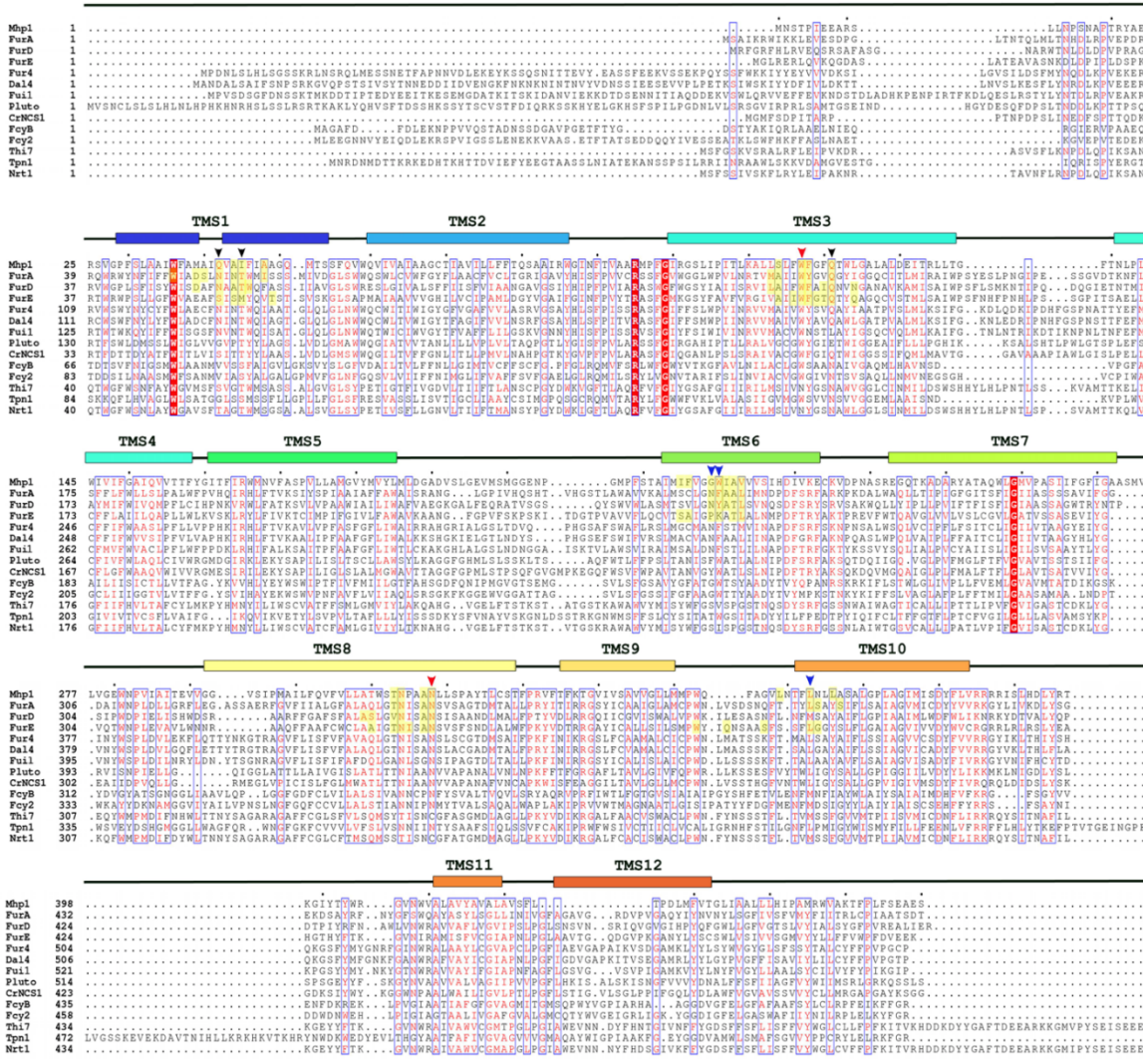
Using homology modeling and Targeted Molecular Dynamics calculations, we constructed the FurE structure the three main conformational states, Outward Open (OO), Occluded (Occ) and Inward Open (IO), using as template the Mhp1 crystal structures with PDB IDs 2JLN, 4D1B, and 2X79, respectively.<sup>87,120,165</sup> The sequence identity of FurE and Mhp1 is 23.9%, excluding the highly variable cytosolic terminal regions. No major difference is observed in the length of TMS helices and loops of FurE, compared to Mhp1. All structures were subjected to 50 ns of MD simulations leading to structure relaxation and equilibration. The three FurE structures obtained are depicted in **Figure 3.1.1**.





**Figure 3.3.1:** The three homology model structures of FurE (Outward Open, Occluded, Inward Open) based on corresponding Mhp1 template crystal structures shown in a side view (orientation parallel to the membrane lipids). The “bundle” helices are colored blue, the “hash” helices are colored cyan, the outer and inner gates are colored red and the TMS11, and TMS12 are grey. The yellow dashed lines represent the membrane plane.

Important conserved residues in the family are revealed by the sequence alignment of selected and with varying specificity NCS1 transporters, namely Mhp1 (D6R8X8; *M. liquefaciens*), FurA (Q5BFM0; *A. nidulans*), FurD (A6N844; *A. nidulans*), FurE (Q5ATG4; *A. nidulans*), Fur4 (P05316; *S. cerevisiae*), Dal4 (Q04895; *S. cerevisiae*), Fui1 (Q04895; *S. cerevisiae*), PLUTO (Q9LZD0; *A. thaliana*), CrNCS1 (A8J166; *Chlamydomonas reinhardtii*), FcyB (C8V329; *A. nidulans*), Fcy2 (P17064; *S. cerevisiae*), Thi7 (Q05998; *S. cerevisiae*), Tpn1 (P53099; *S. cerevisiae*) and Nrt1 (Q08485; *S. cerevisiae*) (**Figure 3.1.2**).

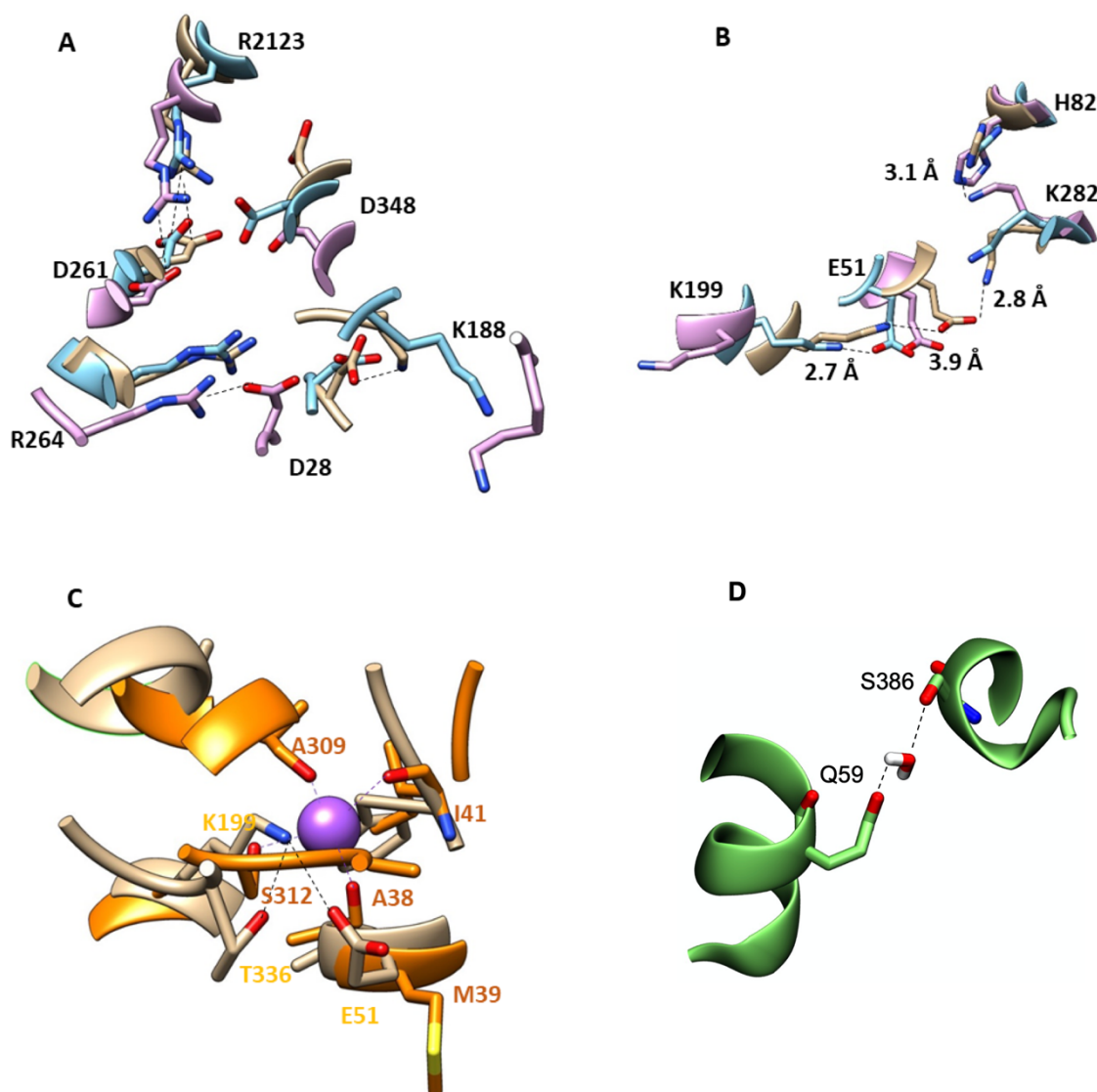


**Figure 3.1.2:** The sequence alignment of selected and with diverse specificity NCS1 transporters, namely Mhp1 (D6R8X8; *M. liquefaciens*), FurA (Q5BFM0; *A. nidulans*), FurD (A6N844; *A. nidulans*), FurE (Q5ATG4; *A. nidulans*), Fur4 (P05316; *S. cerevisiae*), Da14 (Q04895; *S. cerevisiae*), Fui1 (Q04895; *S. cerevisiae*), PLUTO (Q9LZD0; *A. thaliana*), CrNCS1 (A8J166; *Chlamydomonas reinhardtii*), FcyB (C8V329; *A. nidulans*), Fcy2 (P17064; *S. cerevisiae*), Thi7 (Q05998; *S. cerevisiae*), Tpn1 (P53099; *S. cerevisiae*) and Nrt1 (Q08485; *S. cerevisiae*), reveals important conserved residues in the family.

A tryptophan, semi conserved among NCS1 family members, but fully conserved in the Furs (W39 in FurE) is in TMS5, very close to the so-called inner gate TMS5. This tryptophan appears to be able to play the role of the gatekeeper for the exiting of the substrate from the transporter. In FurE position 48, a fully conserved tryptophan is also located in TMS1 but is rather buried in the protein cavity, rather lower from the presumed binding pocket. The

presence of an acidic residue in all Furs is quite characteristic and is absent in other NCS1 family members. In all the Furs, in position 54 and 56 (TMS1) there is an amino acid with an amide side chain, while FurE has a serine, being the only exception. The role of TMS1 is important as resides in the vicinity of the ion binding site, as seen in the Mhp1 crystal structure. A fully conserved arginine and glycine, respectively, exist in the loop between TMS2 and TMS3 and in residue positions 107 and 111 of FurE, far from the binding cavity and inside the intracellular environment. TMS3 in Mhp1 crystal structure, contains important substrate binding site residues, namely W117, Q121, corresponding to W130, Q134 respectively, in FurE. These are widely conserved in the family, signifying the preservation of the binding site. In TMS5, there are well conserved residues, like R193, L195, L202 (FurE numbering). K199 of TMS5 is conserved only in the Furs however, seems strategically located in a position that consists a bridge between the intracellular environment and the ion and substrate binding sites. TMS6 contains W220, important for substrate binding in Mhp1 and highly conserved in the family. Aromatic residues are in this position in the other transporters. However, in FurE, the corresponding residue is K252. K252, points towards the binding site and is important for FurE specificity, regulating the binding of at least one of the three FurE substrates. N318 in TMS8, is highly conserved in the family and is one of the main binding site residues of Mhp1. In FurE this is also an asparagine. In position 313 of Mhp1, all Furs except FurD including Mhp1 possess a threonine. In Mhp1-position 314 there is a conserved asparagine within the family. In the beginning of TMS10, there is a highly conserved phenylalanine, located in the upper part of the transporter which seems to regulate the outer gate of the transporter by interacting with the also aromatic substrates.

Inspecting the structures closely, it seems that the majority of charged residues are arginines and lysines located at the outer surface of FurE and facing the cytoplasm. Few exceptions are essential for the structure and function of the transporter. For example, R123 (TMS3) and D261 (TMS6) can form a salt bridge, mimicking the interaction observed in Mhp1 between K110 (TMS3) and D229 (TMS6). In Mhp1, the two residues interact at 3.87 Å in the IO conformer, whereas in the Occ and OO conformers they are at more than 5 Å. At variance with Mhp1, in FurE due to the bigger size of arginine compared to lysine this interaction is formed in all three conformers. The distance between the carboxylate of D261 and R123 guanidinium is 3.0 Å in OO, 2.85 Å in Occ, and 3.5 Å in IO (**Figure 3.1.3A**). In very close proximity, the D348 (TMS9) carboxylate, which was considered as neutral since it had a pKa equal to 8, points towards the D261-R123 interaction and thus does not seem to participate in the D261-R123 charge neutralization.



**Figure 3.1.3:** Details of interactions between residues in FurE models. **(A)** Interactions between D261-R123 (and vicinity with D348), R264-D28 and D28-K188 shown with black dashed lines and compared in the three FurE conformers (color code: OO in chaki, Occ in cyan, IO in pink). **(B)** Interaction between K199 and E51 in the different FurE conformers are shown in black dashed lines along with distances in Å (color code as in A). **(C)** In Outward Open (OO) conformation the interaction between K199-E51-T336 in FurE model (in chaki) is shown with black dashed lines. When compared with the corresponding cavity of Mhp1 (superimposed in orange), the FurE K199 amino group is located in the same position as the co-crystallized sodium (in magenta) in Mhp1 structure. Mhp1 sodium is coordinated by the S312 and T311 side chains along with A38, I41 and A309 backbone oxygens (magenta dashed lines). **(D)** Interaction between Q59 and S386 through a water molecule in FurE model MD simulation.

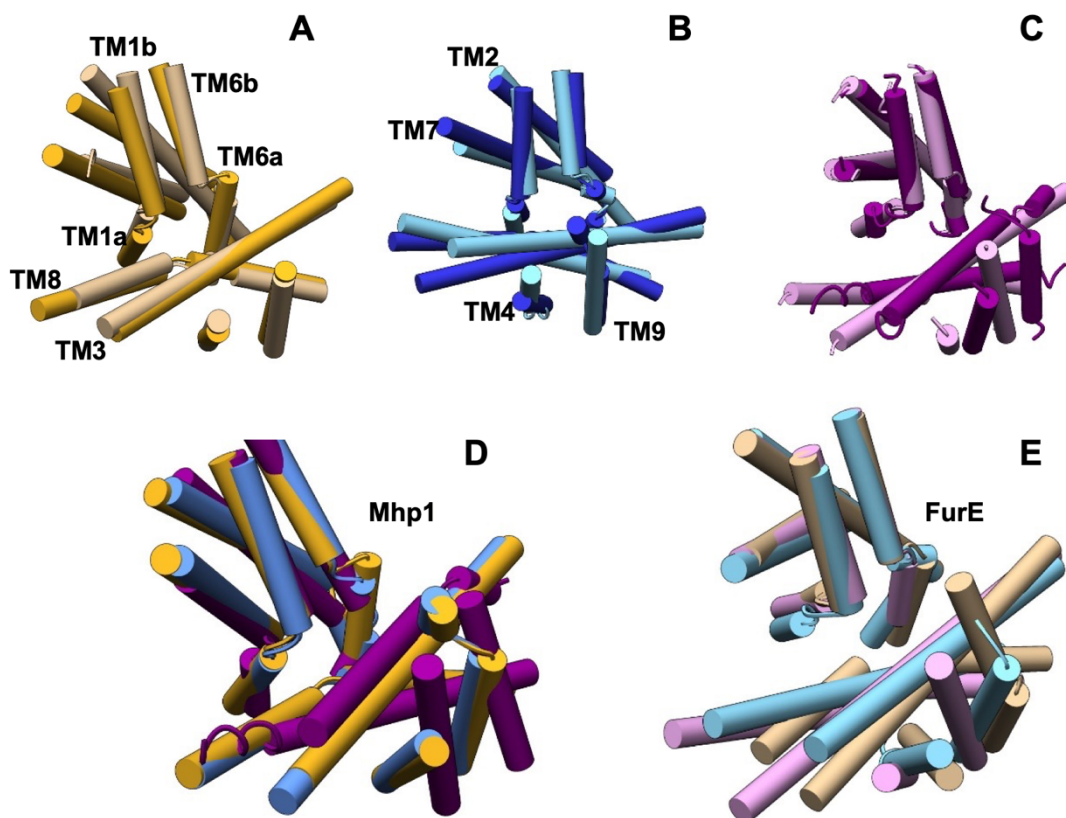
Another important interaction is the one formed between E51 at the edge of TMS1a and K199 positioned in the middle of TMS5. The side chains of both residues are quite flexible, and their interaction seems not stable, with a distance ranging between 2.7 Å in Occ and 3.9 Å in OO (**Figure 3.1.3B**). In IO, K199 side chain has been completely reoriented pointing to the lipid interface of the protein (**Figure 3.1.3B**). Intriguingly, in both OO and Occ states, the K199 side chain amino group is positioned in the position of the co-crystallized Na<sup>+</sup> cation in the Occ

conformation of Mhp1 and very close to that of the second Na<sup>+</sup> (Na2) existing in the corresponding structure of LeuT (**Figure 3.1.3C**). It is worth noting that an additional K252 (TMS6), which has been shown to affect substrate specificity,<sup>96</sup> might also form a second salt bridge with E51 in the OO conformation (**Figure 3.1.3B**). This residue corresponds to W220 in Mhp1, which is considered one of the principal residues participating in substrate recognition and translocation. In the other two conformers, no interaction between K252 and E51 is observed. Finally, the interaction between the D28 N-terminal and K188 (TMS5), reported by Papadaki *et al.*, exists only in the OO structure.<sup>94</sup> In Occ, D28 forms a stable salt bridge with R264. In IO the distance between the oxygens of D28 carboxylate and nitrogen of R264 guanidinium atoms is approximately 5 Å (**Figure 3.1.3A**).

In addition to the above-mentioned, “static” salt bridges, other interactions possibly involved in the function of the outer gate can be related to the OO-to-Occ conformational change. For instance, the Q59 side chain (TMS1b) could approach that of S386 (TMS10) (**Figure 3.1.3D**), as well as S43 (TMS1) and R193 (TMS5), at least in the OO structure, while in the other conformers R193 is oriented towards the surrounding lipids. A similar gating network involving TMS1, TMS5 or TMS10 have also been suggested in LeuT and Mph1.

Overall, in the structures of FurE generated, the relative orientation of the TMSs in OO, Occ and IO conformations is in accordance with that of the Mhp1 structures, however small discrepancies occur. Aligning the structures on the C $\alpha$  atoms of the “bundle” motif, we observe that FurE OO and Occ conformations are in quite good agreement with the Mhp1 templates, whereas in the IO conformer TMS4 and TMS9 are not optimally aligned with the corresponding template helices (**Figure 3.1.4**). Considering that the detected TMS shift is observed after only 50 ns of MD simulation relaxation, it implies a relative flexibility of TMS4 and TMS9.





**Figure 3.1.4.** Relative orientation of transmembrane helices of the “hash motif” (TMS3, 4, 8, 9) compared to the “bundle” (TMS1, 2, 6, 7) in both Mhp1 template crystal structures and the homology models of FurE as resulted after a 50 ns MD relaxation. All structures are aligned on the “bundle motif” and are shown as cylinders for better resolution. (A) Comparison between OO structures of Mhp1 (gold) and FurE (yellow). No major differences are observed. (B) Comparison between Occ structures of Mhp1 (blue) and FurE (cyan). Small difference in TMS3 is observed. (C) Comparison between IO structures of Mhp1 (magenta) and FurE (pink). Differences in the orientation of TMS4 and TMS9 are observed. (D) Superimposition of the three structures of Mhp1 OO (gold), Occ (blue), IO (magenta). (E) Superimposition of the three structures of FurE OO (khaki), Occ (cyan), IO (pink).

### 3.1.3 Mutational analysis confirms the crucial role of specific residues in FurE transport function

The FurE structural models highlight two salt bridges, namely E51-K199 and R123-D261, and a polar interaction between S384 or S386 and Q59 as significant in stabilizing the different states and are considered vital in the conformational transition. Some other residues associated with transient interactions related with the FurE assumed gate conformational changes and the putative substrate pathway are W39, T63, S64, R193, F196, R264, N347 and F385. In order to support the previously suggested substrate interactions during the translocation pathway as hypothesized from the described models defined by MDs, prof. Diallinas' lab performed a rational mutational analysis of the aforementioned residues in FurE,

using a standard directed mutagenesis protocol. The following Ala substitutions mutations were constructed: **W39A, E51A, Q59A, T63A, S64A, R123A, R193A, F196A, K199A, D261A, R264A, N347A, S384A, F385A and S386A**. Other residues predicted from the FurE structure as important for transport activity, such as the interaction of D28 with K188, and the critical role of K252 in substrate binding and specificity, have been previously supported by analogous Ala substitutions.<sup>94</sup> Genetic transformation in the *A. nidulans*  $\Delta 7$  strain, genetically lacking all major nucleobase-related transporters was used to introduce all the mutated versions of FurE. The mutant versions of FurE, among a wild-type FurE control, were expressed *via* the strong *gpdA* promoter, fused at their C-terminus in frame with the *gfp* gene, as previously described.<sup>94,96</sup> Several selected transformants for each mutation were purified and tested for growth in media containing purines as nitrogen sources or 5-fluorouracil (5-FU, a uracil toxic analogue). As expected, transformants expressing a specific mutation exhibited similar growth phenotypes.

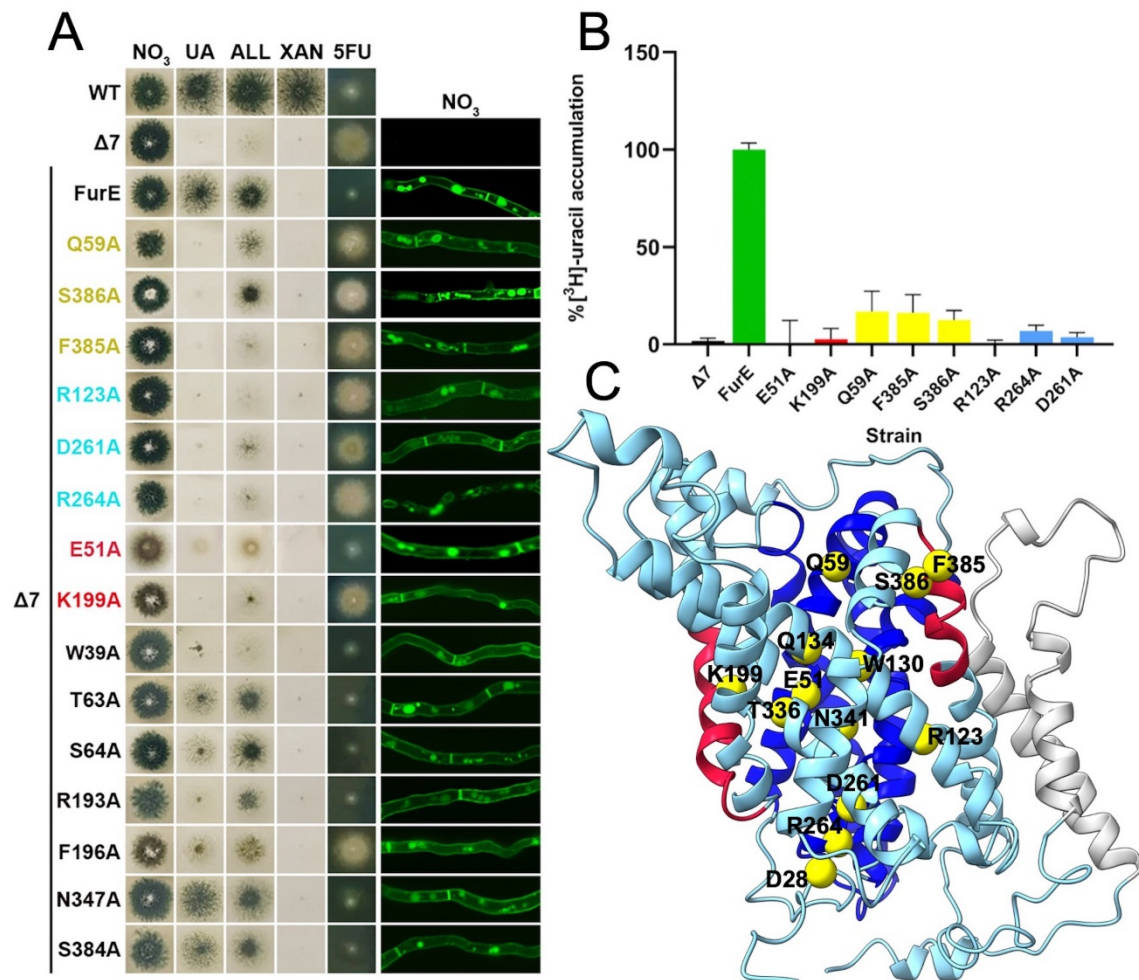
The upper left panel (**Figure 3.1.5**) summarizes the growth phenotypes tested for all engineered mutants and control strains. As expected, the positive control strain expressing wild-type FurE grows on allantoin and uric acid and is sensitive to 5-FU, while the negative control strain not expressing FurE is resistant to 5-FU and displays a N starvation growth phenotype. The residues predicted to form the two major salt bridges (R123-D261 and E51-K199) were subjected to Ala substitutions and scored as loss-of-function mutations as shown by abolishment or dramatic reduction of growth on allantoin or uric acid. They relatively increased resistance to 5-FU, mostly in the case of R123A and D261A. Similar great loss of FurE transport activity was obtained in R264A and F385A mutants, while Q59A and S386A FurE versions seemed to have lost their transport activity for uric acid or 5-FU, but conserved some capacity for allantoin transport. Hence, the important functional role of the interactions between E51-K199, R123-D261 and Q59-S385-S386 was confirmed by mutational analyses, as was also the importance of R264 as it interacts with N-terminal D28, according to the predictions, as proved by D28 mutation.<sup>94</sup> This analysis also showed the importance of W39, as its substitution led to loss transport of for uric acid and allantoin in FurE, although since sensitivity to 5-FU is retained, transport seemed to be conserved. These results further revealed that substitution of T63, S64, R193 or F196 with Ala has sensible negative effects on FurE apparent activity, as shown by the reduction of growth on uric acid and certain increase in 5-FU resistance (e.g. F196A). Finally, residues N347 and S384 were found non-important for FurE activity.

Epifluorescence microscopic analysis, performed by prof. Diallinas' team, shown in the right panel of **Figure 3.1.5A** confirmed that mutational disruption of the major interactions tested, like R123-D261, E51-K199 and Q59-F385-S386, did not affect the normal PM localization and stability of FurE. This fact further confirms that the related growth defects in specific mutants reflect defects in FurE transport activity *per se*, and do not affect protein folding or subcellular trafficking. This conclusion was further strengthened by direct transport assays performed by prof. Diallinas' team, which showed that FurE-mediated radiolabeled uracil transport was practically abolished in the respective mutants (**Figure 3.1.5B**). The R264A mutation was the only case that apparent loss-of-function proved emerge because of the abolishment of translocation to the PM, because of misfolding and ER-retention of this mutant form of FurE. This suggests that R264 participates in interactions that stabilize the expression level of FurE, in addition to its functional interaction with D28. Notice, however, that D28A substitution does not lead to FurE misfolding.<sup>94</sup>

In conclusion, FurE transport function was found to be affected by mutations of residues proposed by the homology models and MDs, to participate in specific dynamic interactions along the substrate pathway. In particular, all residues proposed to be involved in salt bridges associated with uracil/H<sup>+</sup> binding and symport, namely E51-K199 and R123-D261, all led to dramatic reduction of FurE transport activity. Mutations in residues like Q59A-F385-S386A which were suggested to be involved in polar interactions in the outer gate, resulted in reduction in transport capacity, but also in modifications in substrate specificity. Q59A and S386A mutants significantly lost uric acid, uracil and 5-FU transport, but retained apparent transport activity for allantoin, as confirmed by growth tests. Notably, the F385A mutation recorded for all substrates as a loss-of-function mutation, which could be due to its proximity to W130, also important in substrate specificity. Remarkably, mutations T63A and S64A, on residues that could possibly interact with S386 of the outer gate, also showed specific reduction



of uric acid transport, but retained a substantial capacity for allantoin or 5-FU accumulation, whereas mutations R193A and F196A led to reduction in uric acid transport.



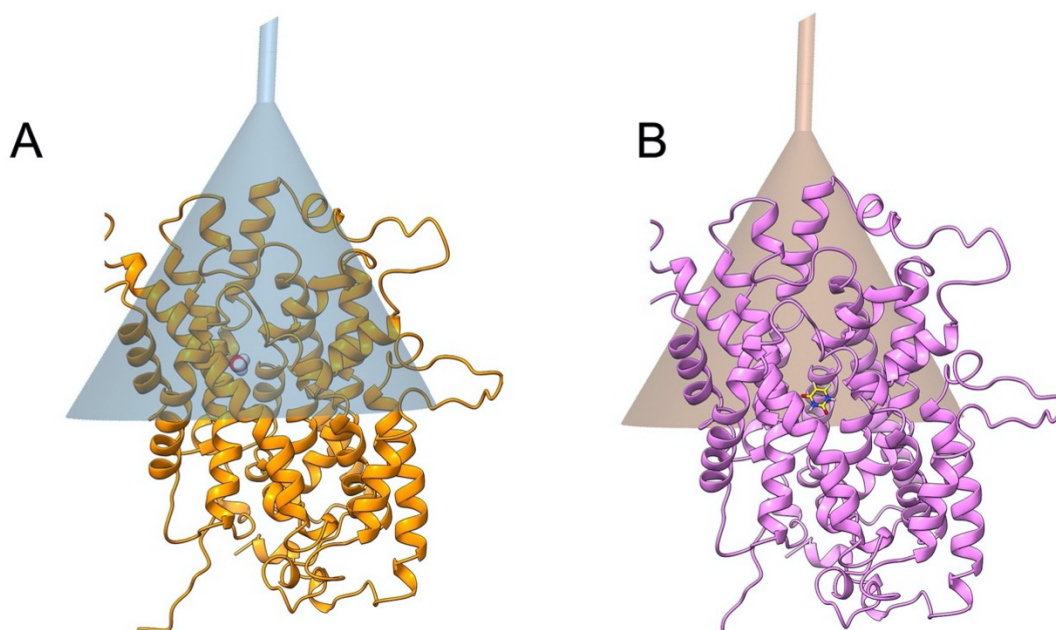
**Figure 3.1.5:** Functional analysis of FurE mutants. (A) Growth tests of isogenic strains expressing distinct FurE mutant versions in a  $\Delta 7$  genetic background (i.e. genetically lacking other nucleobase-related transporters), compared to a positive (FurE) and a negative ( $\Delta 7$ ) control strain (for strain details see materials and methods). NO<sub>3</sub>, UA, ALL, Xan denote MM supplemented with nitrate, uric acid, allantoin or xanthine as sole N source. 5FU is MM+NO<sub>3</sub> supplemented with 5-FU. WT denotes a standard *A. nidulans* wild-type strain expressing all major nucleobase transporters. *In vivo* epifluorescence microscopy of the same strains is shown in the right panel. All FurE mutants are functionally tagged with GFP. Notice that all FurE mutant versions, except R264A, exhibit normal (i.e. wt FurE-like) plasma membrane localization and vacuolar turnover. R264A is trapped in the perinuclear ER rings, typical of misfolded versions of FurE or other transporters (for details see Materials and methods) (B) Direct uptake assays of selected FurE mutants, using 0.2  $\mu$ M [<sup>3</sup>H]-radiolabeled uracil. The figure shows relative % initial uptake rates (1 min) of mutants, when wild-type FurE transport is taken as 100%, performed with 107 germinated conidiospores, as described by Kryptou and Diallinas, 2014. All functional assays shown are performed by prof Diallinas' team.

### 3.1.3 Molecular Dynamics calculations

#### 3.1.3.1 The binding mode of hydronium

All the characterized NCS1 transporters seem to function via proton symport except for Mhp1, which is a Na<sup>+</sup> driven NCS1 symporter. Still, proton interactions don't seem to be

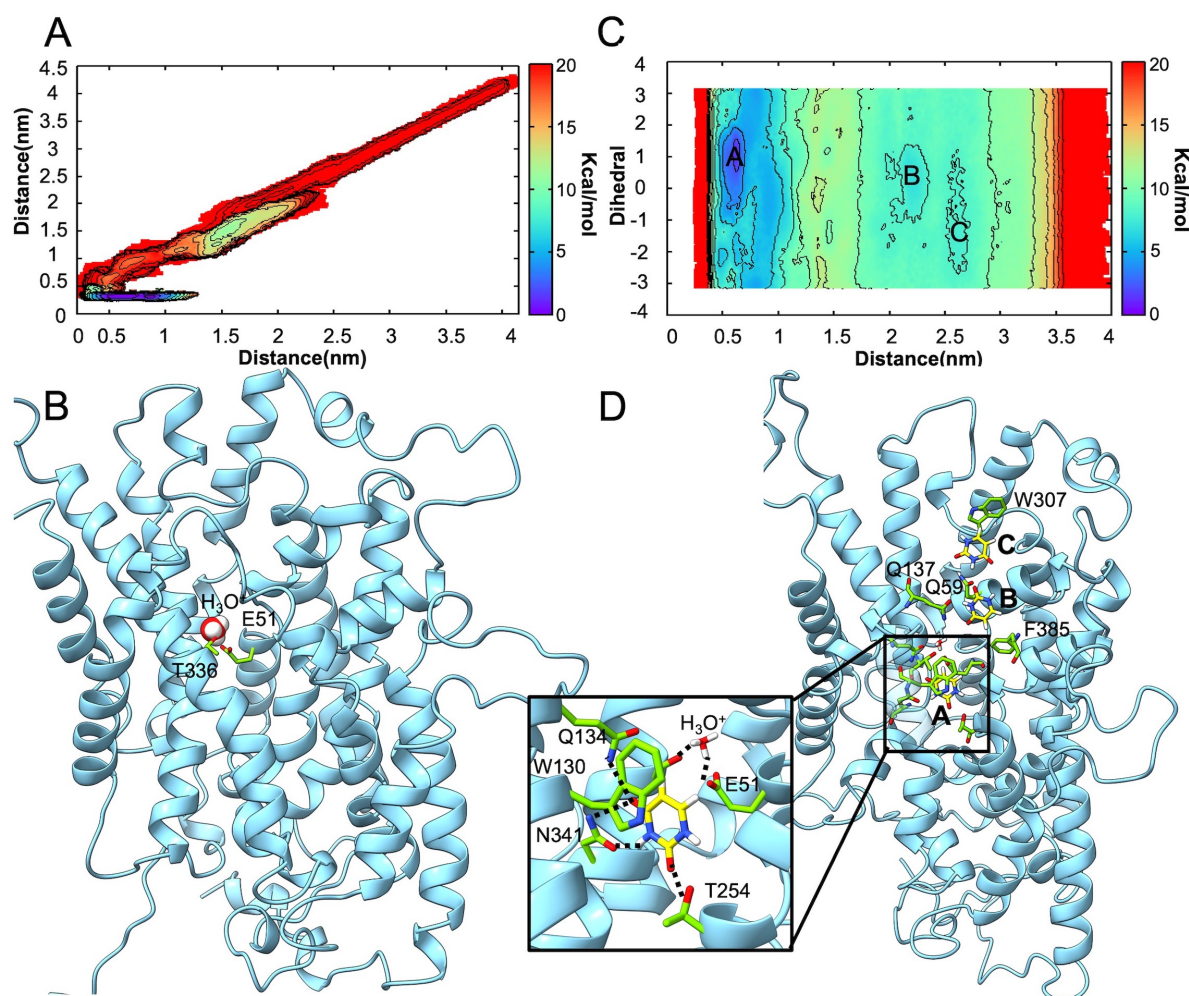
elucidated for none of them. In the case of the FurE purine/proton symporter, no experimental information is available regarding the binding site/mode of the cation. In Mhp1, the symported ion stabilizes the OO conformation for substrate binding and accumulation,<sup>164</sup> while in LeuT induces a conformational change.<sup>172</sup> Aiming to enlighten this critical part of substrate/ion co-transport and in order to characterize the proton interactions, we investigated the binding of a hydronium molecule ( $\text{H}_3\text{O}^+$ ) to FurE by utilizing Funnel-Metadynamics (FM), a cutting-edge method for molecular binding free-energy, introduced in our group, which is widely used to study ligand-protein systems.<sup>143</sup> Due to FM the process of binding of hydronium from its fully solvated state in the extracellular environment to the binding site in the OO state can be reproduced, using an adaptively constructed time-dependent bias potential accumulated as a sum of Gaussian functions. That potential accelerates the sampling along a definite degree of freedom called collective variable (CV), which in this case is the distance of  $\text{H}_3\text{O}^+$  from FurE. Throughout the FM simulations, the complete binding pathway was simulated, and all the potential binding sites energetically evaluated (**Figure 3.1.6A**).



**Figure 3.1.6:** Funnel-shaped restrain potential used for the four cases of FM simulations: (A) the funnel used for the investigation of the entrance of  $\text{H}_3\text{O}^+$  in FurE transporter. (B) The funnel investigating the entrance of uracil in FurE transporter.

The binding site of hydronium was identified in the Binding Free Energy Surface (BFES) as the lowest energy state (**Figure 3.1.7A**) and proved to be identical with the one presented for  $\text{Na}^+$  in Mhp1. This site is located at the interface of TMSs 1 and 8 of the rigid and moving part of FurE, forms interactions with E51 and T336 of TMS1b and TMS8, respectively (**Figure 3.1.7B**). The structural stability of the obtained of the FurE/ $\text{H}_3\text{O}^+$  complex

was further assessed by standard MD simulation of 150 ns. We note here, that the hydronium binding mode is very similar to that in both Mhp1 and LeuT and the  $\text{Na}^+$  binding pose in the co-crystallized structures. In addition, the residue in position 336 in FurE, which is a threonine, is conserved in these three transporters. Finally, increased flexibility displayed in the presence of  $\text{H}_3\text{O}^+$ , was observed in the case of residue K199.



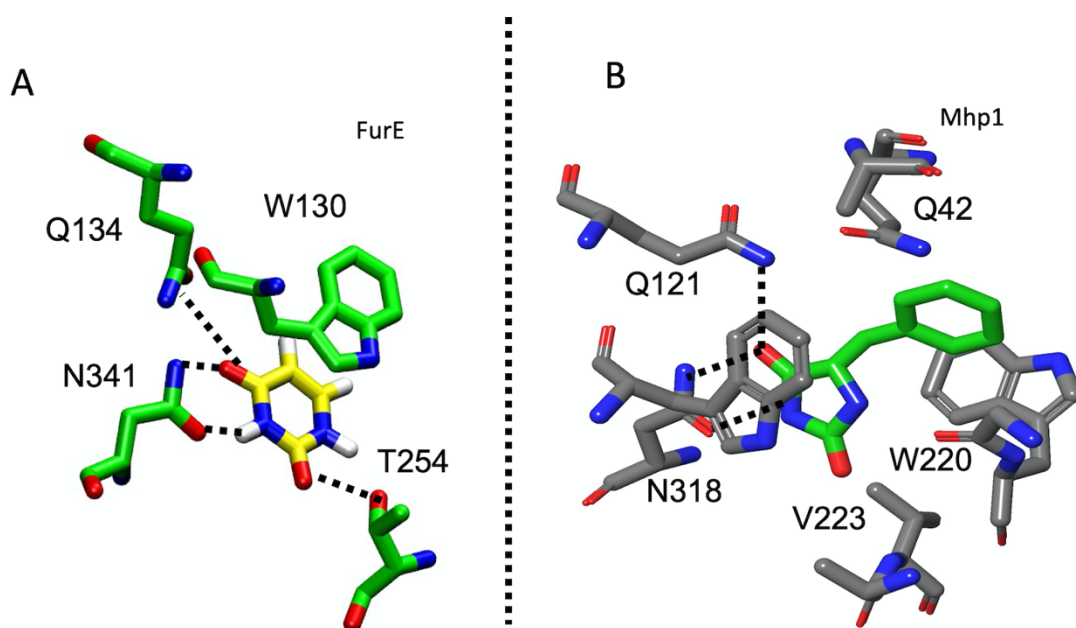
**Figure 3.1.7:** The binding site of  $\text{H}_3\text{O}^+$  and uracil accompanied with other key low energy states. (A) The BFES of  $\text{H}_3\text{O}^+$  binding process. Contour lines are drawn every 2 kcal/mol. (B) The binding mode of  $\text{H}_3\text{O}^+$  cation in FurE transporter as resulted from the global energy minimum in the FES. (C) The BFES of uracil binding process in FurE transporter. (D) The intermediate states (local minima in the BFES) of uracil entering FurE transporter and the binding mode in the binding site (inset) as resulted from the BFES in C.

### 3.1.3.2 The binding mode of uracil

The binding process of uracil ( $K_m = 1$  mM) to its putative binding site in FurE was also simulated, using FM. As suggested by mutagenesis data, this site was confined to TMS1, TMS3, TMS6 and TMS8, as is also the binding site in Mhp1 and other NCS1 transporters.<sup>95</sup> We simulated the binding process of uracil, as performed for  $\text{H}_3\text{O}^+$ , from its fully solvated state



to the binding site in the Occ state, using as CV, the distance of uracil from the binding site. We used the result of docking calculations as a starting structure for the uracil binding mode. This pose does not affect the result, as during FM the ligand explores all the possible binding poses at the binding site. Given that the hydronium is bound to the transporter in the Occ state,<sup>120,164</sup> it was included in this system, with the binding geometry identified by the FM calculation previously. A big cone-shaped perimetric constraint is used around the binding site during the FM simulation, to ensure the sampling of a wide area (**Figure 3.1.6B**). The selected model of uracil binding mode derived from the global minimum of the BFES (**Figure 3.1.7C**), (**Figure 3.1.7D**) and is outstandingly similar with that of the hydantoin analogue, (5S)-5-benzylimidazolidine-2,4-dione, in Mhp1 crystal structure (**Figure 3.1.8**).<sup>120</sup> In detail, H-bonds interactions were observed between T254 (TMS6) and uracil C2=O, N341 (TMS8) and uracil C4=O and N3, and  $\pi$ - $\pi$  stacking interactions between W130 (TMS3) and uracil. Two other energetic minima were discovered at higher energy values, which denote intermediate binding poses of the ligand across its pathway to the binding site. At these binding poses, uracil interacts at first with Q59 (TMS1b), via a bidentate interaction with C2=O and N3, and a  $\pi$ - $\pi$  stacking with W307 (L7 loop) (**Figure 3.1.7 C**). Next, moving lower in the FurE binding cavity, it interacts via a bidentate bond involving C4=O and N3 with Q137 (TMS3) (**Figure 3.1.7D**). Markedly, W130 of the binding pocket remained parallel to the membrane axis, keeping the binding site exposed for the substrate. Finally, both uracil and W130 interact with F385 (TMS10), through  $\pi$ - $\pi$  and T-shaped stacking interactions.



**Figure 3.1.8:** Comparison of the binding mode of (A) uracil in FurE with (B) (5S)-5-benzylimidazolidine-2,4-dione in Mhp1.

### 3.1.3.3 The conformational transition of FurE from OO to IO

Aiming to systematically portray the FurE large-scale conformational change, also considering the role of hydronium, uracil and the relative order of their transport, we employed path collective variables (PCVs), which is a dimensionality reduction methodology. PCV is able to study complex chemical and biological processes including a large number of degrees of freedom using one or two collective variables.<sup>144</sup> In this case, the above-mentioned conformational change can be studied by providing a set of frames describing the required movement. In these frames are included the positions of key atoms from the start to the end of the conformational transition, giving us the opportunity to track the transition stage during the simulation and accelerate its sampling using Metadynamics. The whole OO to IO transition of FurE was explored through two sets of simulations. The first described the OO-to-Occ transition and the second the Occ-to-IO one. For that end, four systems were created and investigated considering all possible ligand stoichiometry combinations:

- i) FurE - H<sub>3</sub>O<sup>+</sup> - uracil (*apo*);
- ii) FurE + H<sub>3</sub>O<sup>+</sup> - uracil;
- iii) FurE - H<sub>3</sub>O<sup>+</sup> + uracil;
- iv) FurE + H<sub>3</sub>O<sup>+</sup> + uracil.

In the case that hydronium and uracil are inside the system, they occupy the binding mode previously identified through FM. Considering the FES obtained at the end of the PCV calculations, we extracted the FurE structures representing the global minimum and clustered them. The central structure of the most cluster containing the most structures, was selected and subjected to a 100ns standard MD simulation to assess the stability. These structures are fully described in the following paragraphs. Moreover, the interactions between the most important residues during the simulated pathways and the simulation time were monitored and reported in **Figure 3.1.9A** and in **Table 3.1.1**, accordingly.

### 3.1.3.4 OO-to-Occ path:

i) *Apo state (FurE - H<sub>3</sub>O<sup>+</sup> - uracil)* - A single broad free energy minimum, compared to the other systems, is depicted between OO and Occ on the FES, (**Figure 3.1.9B**). This clearly shows that, FurE adopts a variety of conformations in the absence of ligands. Standard MD simulations on the *apo* form of the OO state, confirmed this flexibility (**Figure 3.1.10**). Notably, comparing with Mhp1, the relative orientation between “hash” and “bundle” motives

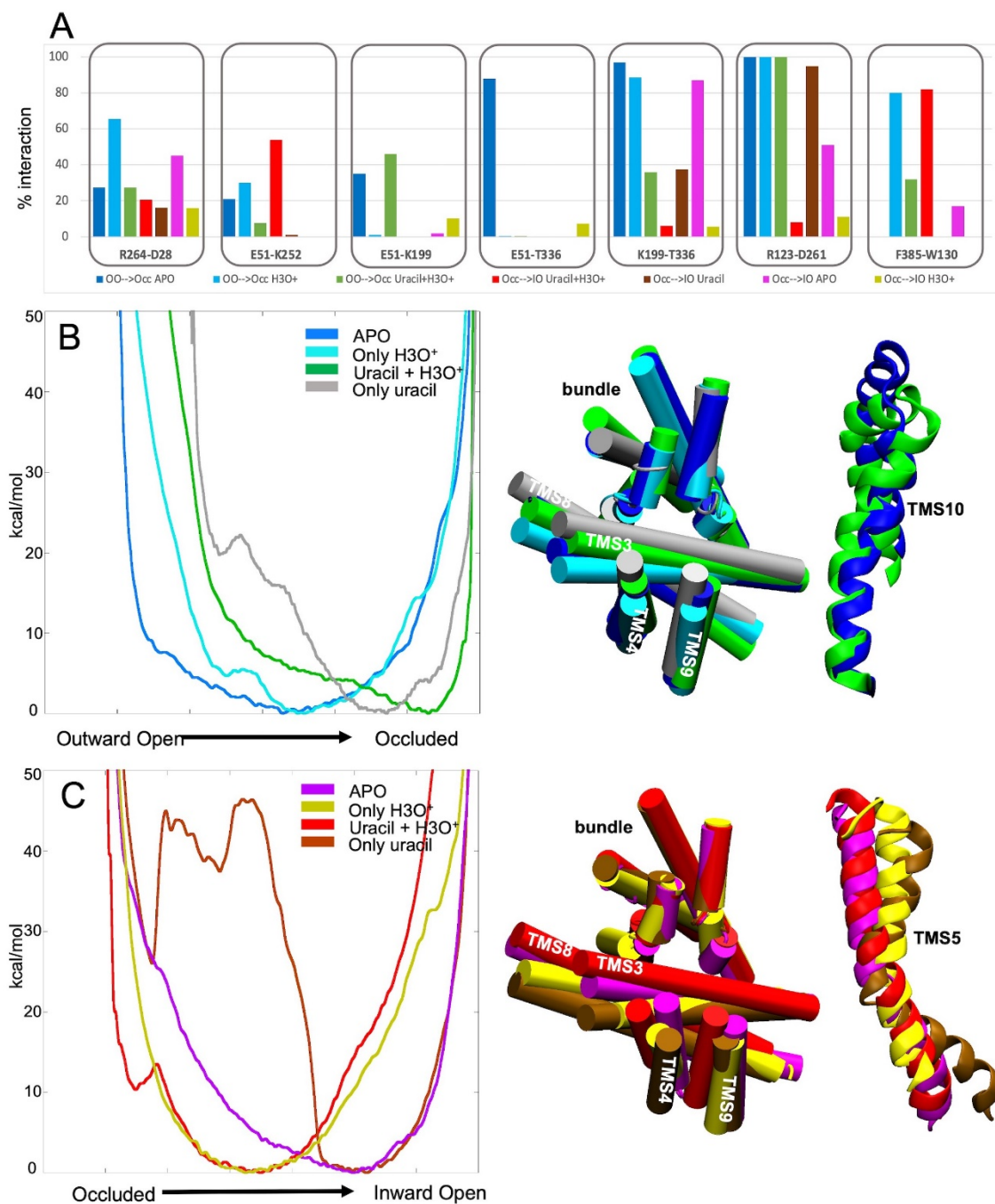
remained almost the same. Also, the primary part of TMS10 is located much closer to TMS1a with respect to the starting OO structure in the energy minimum. K199-T336, E51-T336 and R123-D261 are the most stable interactions in the *apo* state, while E51-K199 and R264-D28 interact at a minor degree (**Figure 3.1.9A**). Additionally, a water bridge is possibly formed between S386 and Q59 (**Figure 3.1.11**), while a water molecule assists the interaction between Q134 and T336 (**Figure 3.1.12A**).

*ii) Hydronium bound ( $FurE + H_3O^+ - uracil$ )* – In the case that the  $H_3O^+$  cation is bound to FurE, the FES is quite similar to the *apo* form, although the minimum is narrower, indicating a reduced flexibility of the transporter and specifically of TMS10 (**Figure 3.1.9B**). This result suggests the impact that the  $H_3O^+$  has on the free energy landscape. In detail,  $H_3O^+$  binding locks TMS10 in a position beneficial for the substrate binding. More precisely,  $H_3O^+$  forms a salt bridge with E51 and a H-bond with T336, and as a consequence the bond between K199 and E51, between E51 and T336 and the water bridge between Q134 and T336. Consequently, T336 interacts only with K199 and the H-bond between T336 and Q134 is lost, making Q134 available to interact with uracil (**Figure 3.1.12B**). In line with the above-mentioned experiments, we obtained strong evidence for a firm connection between  $H_3O^+$  and substrate binding and transport.

*iii) Uracil bound ( $FurE - H_3O^+ + uracil$ )* – A FES minimum located close to the Occ state is observed when only uracil is bound to FurE the system (**Figure 3.1.9B**). However, in this pose there is still a relative distance between the first part of TMS10 and TMS1a. The proximity of TMS10 to TMS1a literally describes the accurate Occ state, as indicated in the experimental Mhp1 structure (PDB ID 4D1B). Nevertheless, unbiased MD simulations performed on this system show that uracil is not stable in the binding pocket and leaves the binding pocket after 20ns. During this event, TMS10 fluctuated between the OO and Occ state (**Figure 3.1.10**).

*iv) Hydronium and uracil bound ( $FurE + H_3O^+ + uracil$ )* - In the case both hydronium and uracil are bound, the lowest energy minimum represents the Occ state (**Figure 3.1.9B**). Comparing FurE in this state and Mhp1 in the crystallized Occ state, negligible differences are observed. These were a tilt in TMS5 towards the IO conformation, as well as in TMS3 and TMS9. In particular, TMS10 is stabilized in the occluded position merely due to the contribution of the interaction between F385 and W130, leading consequently to the motion of TMS9 which is though not detected in the Mhp1 crystal structure. Furthermore, a slight tilt of TMS5 is evident connotating that FurE in this case has moved slightly closer to the IO state regarding the conformation, foreshadowing a low energy barrier between the occluded and an

Inward Occlude state. Additionally, the uracil binding stabilizes forms an H-bond with Q134, stabilizing it in a position competent of making an H-bond network with Q59 and water molecules. Moreover, S386 (TMS10) can interact with Q59 (TMS1) either via a water molecule or directly (**Figure 3.1.11**).



**Figure 3.1.9:** FurE conformational change and mechanism of transport. (A) The side chain interactions of important residues have been studied in all structures retrieved from each FES global minimum and are represented as percentage over the ensemble of the structures. (B) The FESs of the OO-to-Occ transition using diverse stoichiometry of ligands bound to the transporter (colour code: protein in the *apo* form in blue, complexed only with H<sub>3</sub>O<sup>+</sup> in cyan, complexed with both uracil and H<sub>3</sub>O<sup>+</sup> in green, complexed only with uracil in grey). Each

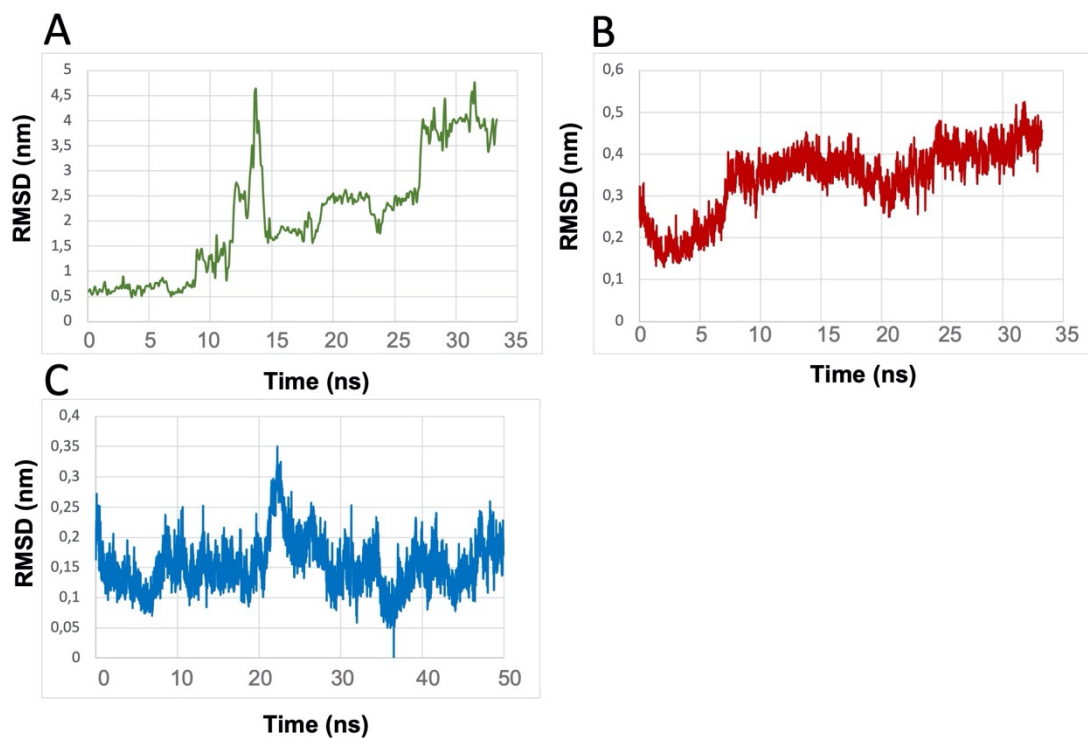
tick in the x axis represents one unit. (C) The FESs of the Occ-to-IO transition are presented using different stoichiometry of ligands bound to the transporter. The system containing both uracil and  $\text{H}_3\text{O}^+$  is represented in red, the system containing only uracil is represented in brown, the *apo* form is represented in magenta, the system containing only  $\text{H}_3\text{O}^+$  is represented in yellow.

**Table 3.3.1:** The simulation time of each case of the PCV Metadynamics simulations.

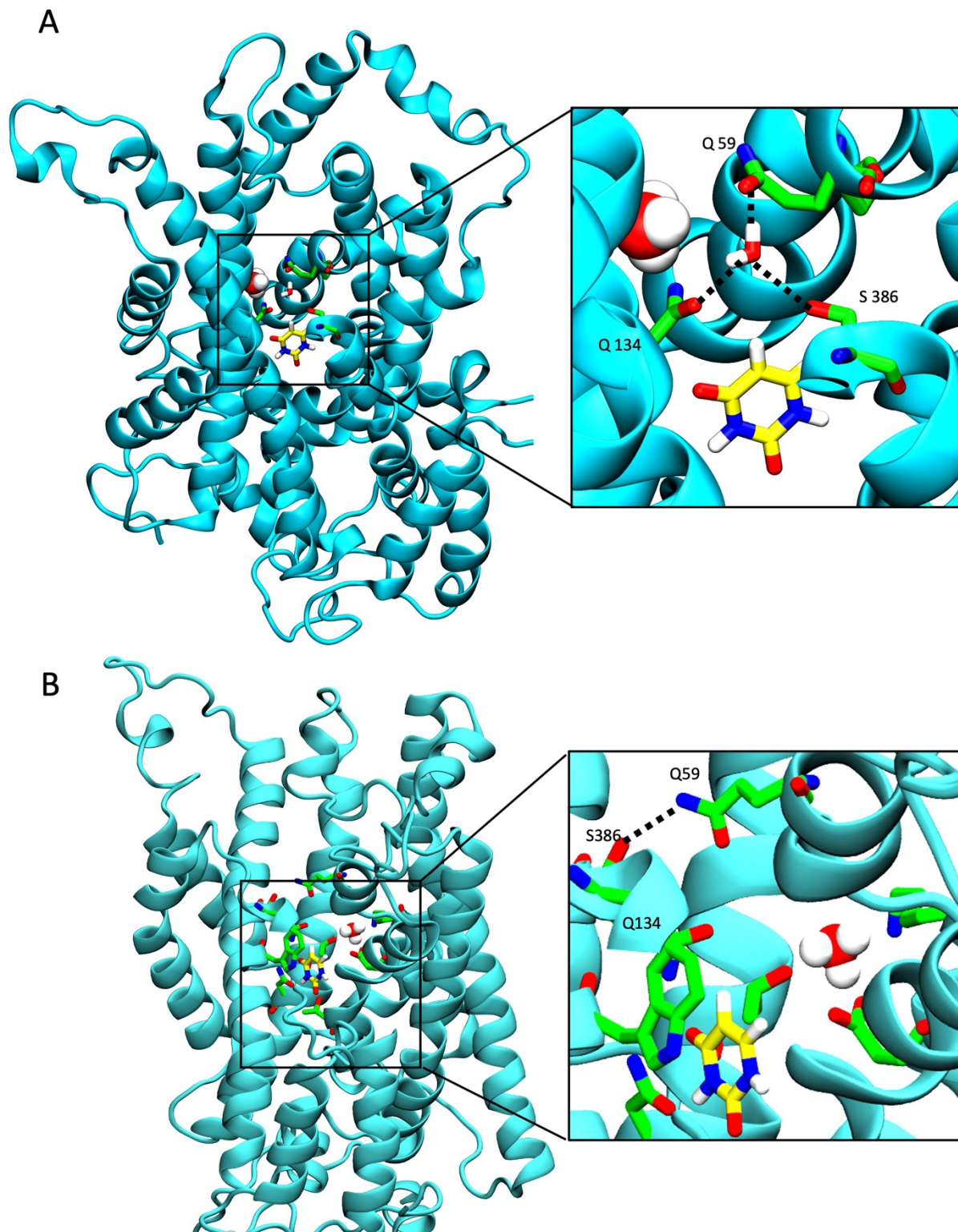
<b>Op-to-Occ</b>	<i>apo</i>	$\text{H}_3\text{O}^+$	$\text{H}_3\text{O}^+$ +uracil	uracil
Time(ns)	352	550	600	430
<b>Occ-to-In</b>	<i>apo</i>	$\text{H}_3\text{O}^+$	$\text{H}_3\text{O}^+$ +uracil	uracil
Time(ns)	330	390	314	298

Summarizing, given the results presented, we can provide exceptional structural insight into the OO-to-Occ transition of FurE. Specifically, it is undoubtedly shown that the FurE conformation is stabilized in the presence of hydronium and thus is prepared for the binding of uracil. Further, the binding of both hydronium and uracil is necessary to lock the FurE Occ conformation. the low RMSD values ( $\sim 1$  Å) computed for the backbone  $\text{C}\alpha$  atoms of the transporter in unbiased MD calculations indicate that the Occ state is very stable as. Furthermore, the disruption of bonds that stabilized TMS5 in OO state in a closed position and retained hash-bundle domain in a firm orientation, namely K199-E51, K199-T336, E51-T336, allow the FurE structure to move towards IO state.

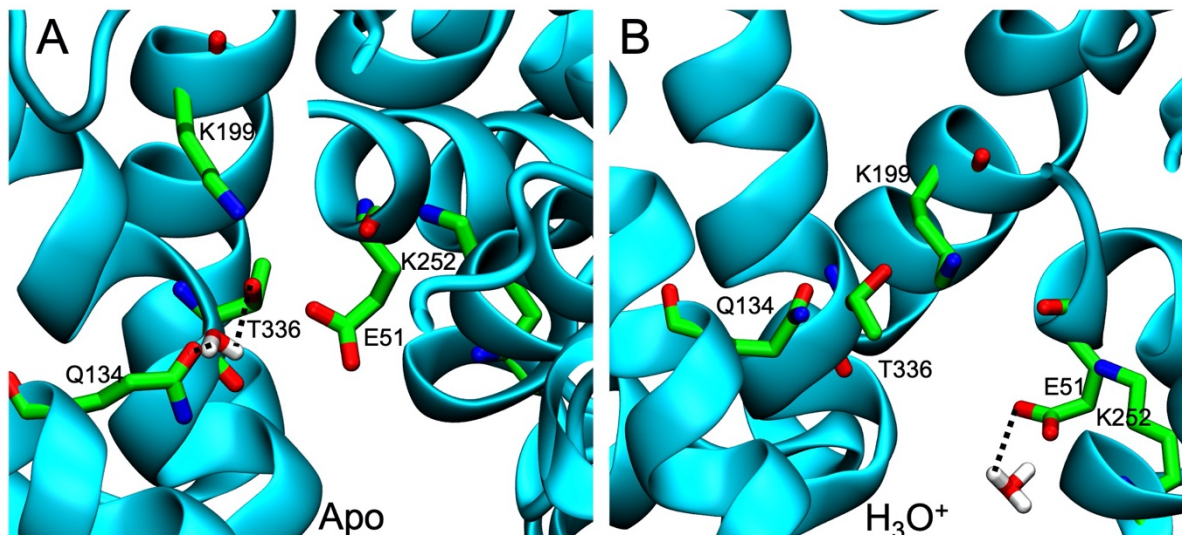




**Figure 3.1.10:** RMSD diagrams. (A) The RMSD fluctuation of uracil during an unbiased Molecular Dynamics simulation in the occluded FurE structure until it is out of the transporter, with uracil included and  $\text{H}_3\text{O}^+$  not included. (B) The RMSD fluctuation of TMS10 during an unbiased Molecular Dynamics simulation in the occluded FurE structure, with uracil included and  $\text{H}_3\text{O}^+$  not included. (C) The RMSD fluctuation of TMS10 during an unbiased Molecular Dynamics simulation in the outward *apo* FurE structure.



**Figure S3.1.11:** S386-Q59 interaction stabilize the Occ conformation of FurE by closing the TMS10 outer gate. **(A)** S386 interacts with Q59 through an H-bond network which involves a water molecule. **(B)** S386 forms a direct H-bond with Q59.



**Figure 3.1.12:** Different residue interactions in the absence and presence of  $H_3O^+$  cation in the OO FurE conformation. **(A)** When FurE is in APO Form, E51 interacts strongly with K199 and less often with K252. The orientation of T336 oxygen is towards TMS3 as T336 interacts with Q134 through H-bond network involving a water molecule. **(B)** When  $H_3O^+$  is present the electrostatic bond between K199 and E51 is less likely to happen, while is more likely for K252 to interact with E51. T336 side chain has also rotated by 180 degrees as it interacts with  $H_3O^+$  as displayed in Figure 1B. When  $H_3O^+$  is present T336 side chain oxygen is directed towards TMS5, thus the interaction with Q134 is broken and the latter is free to bind the substrate (uracil).

### 3.1.3.5 Occ-to-IO path

*i) Hydronium and uracil bound ( $FurE + H_3O^+ + uracil$ )* – In the case that both uracil and  $H_3O^+$  are bound to FurE, a low energy conformation that very much resembles the IO state, albeit not reaching it, is obtained by the transporter. We defined this structure as the Inward Occluded (IOcc) state. A tilt in TMS3, induces both TMS4 and TMS5 to obtain a quasi-open state. The disruption of the electrostatic interaction between D261 (TMS6) and R123 (TMS3) seem to lead to the motion of TMS3 (**Figure 3.1.9A**). Instead, R123 interacts with T254, an uracil binding site residue and with uracil.  $H_3O^+$  is in a position more buried compared to the one in Occ state, and approaches D28 of the N-terminal loop. Furthermore, the E51 side chain, following the cation, rotates resulting in a solid interaction with K252 (**Figure 3.1.9A**).

*ii) Uracil bound ( $FurE - H_3O^+ + uracil$ )* – In this state, a minimum close to the IO conformation appears in the FESs, which is narrow and has a high-energy barrier (**Figure 3.1.9C**). Based on this finding we suggest that  $H_3O^+$  unbinds first and leaves FurE and then the later acquires a conformation close to the IO, favorable for uracil release. Comparing FurE in this conformation with the inward Mhp1, we notice that TMS5 tilts more and induces the movement of TMS8 and finally the rearrangement of the “hash” motif and TMS3.

iii) *Apo state (FurE - H<sub>3</sub>O<sup>+</sup> - uracil)* - The FES of the *apo* state presents the lowest energy minimum close to IO, similar to the uracil bound state. Yet the minimum of the *apo* is wider, representative of a higher conformational freedom of the transporter in this state. Greater flexibility was noticed for TMS5, while TMS3 bents slightly if compared with the Occ state. As FurE is encountered in the *apo* state once both the ligands are released, such conformational freedom might be helpful to service the reverse transition of the transporter to the outward state. The FurE flexibility was confirmed by standard MD simulations carried out on the structure corresponding to the energy minimum.

iv) *Hydronium bound (FurE + H<sub>3</sub>O<sup>+</sup> - uracil)* - As only H<sub>3</sub>O<sup>+</sup> is bound to FurE, the structure corresponding to the energy minimum, is between Occ and IO state (**Figure 3.1.9C**). Here, TMS5 is very close to the position assumed in Occ. This suggests that in the absence of uracil the protein is not able to reach the IO state.

Our enhanced sampling protocol shows that as H<sub>3</sub>O<sup>+</sup> is still bound to the protein the conformation is stabilized in an intermediate state between Occ and IO, suggesting that the sequence of events first includes displacement and detachment of the H<sub>3</sub>O<sup>+</sup>, while the uracil is necessary to shift to the final IO state and then it can also be released. Besides, this transition from Occ to IO is related to TMS3 tilt and TMS8 shift which are also related to both H<sub>3</sub>O<sup>+</sup> and substrate interactions.

### 3.1.3.6 *The internalization pathway of H<sub>3</sub>O<sup>+</sup> cation*

According to our PCV calculations on the FurE Occ-to-IO transition, hydronium is the first to be released in the cytoplasm. Therefore, we used FM simulations to investigate the unbinding of H<sub>3</sub>O<sup>+</sup> from the FurE transporter (**Figure 3.1.13C**).



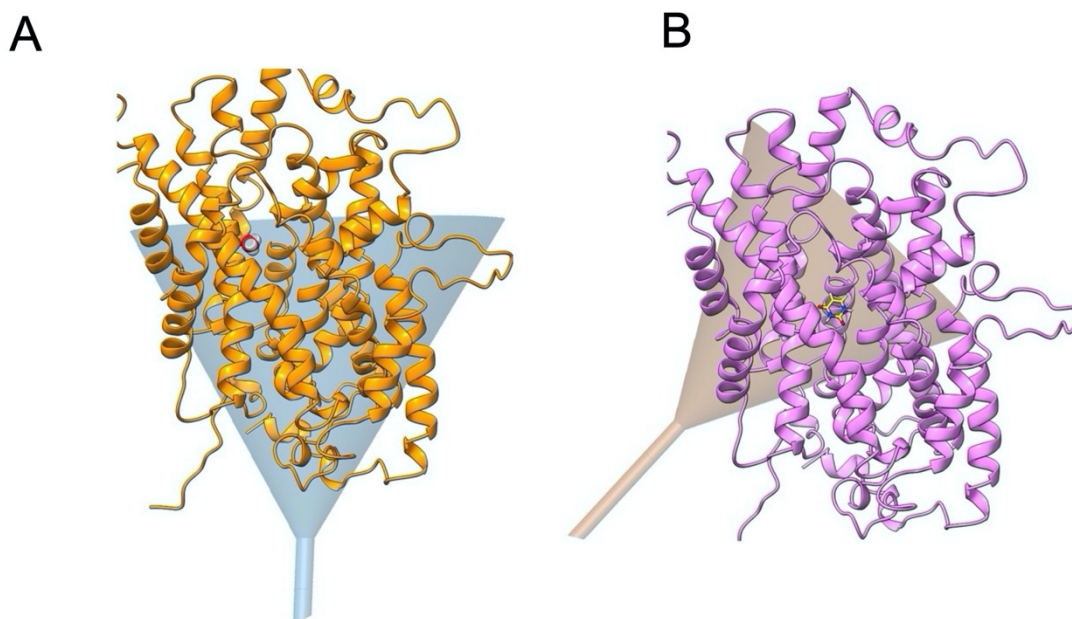
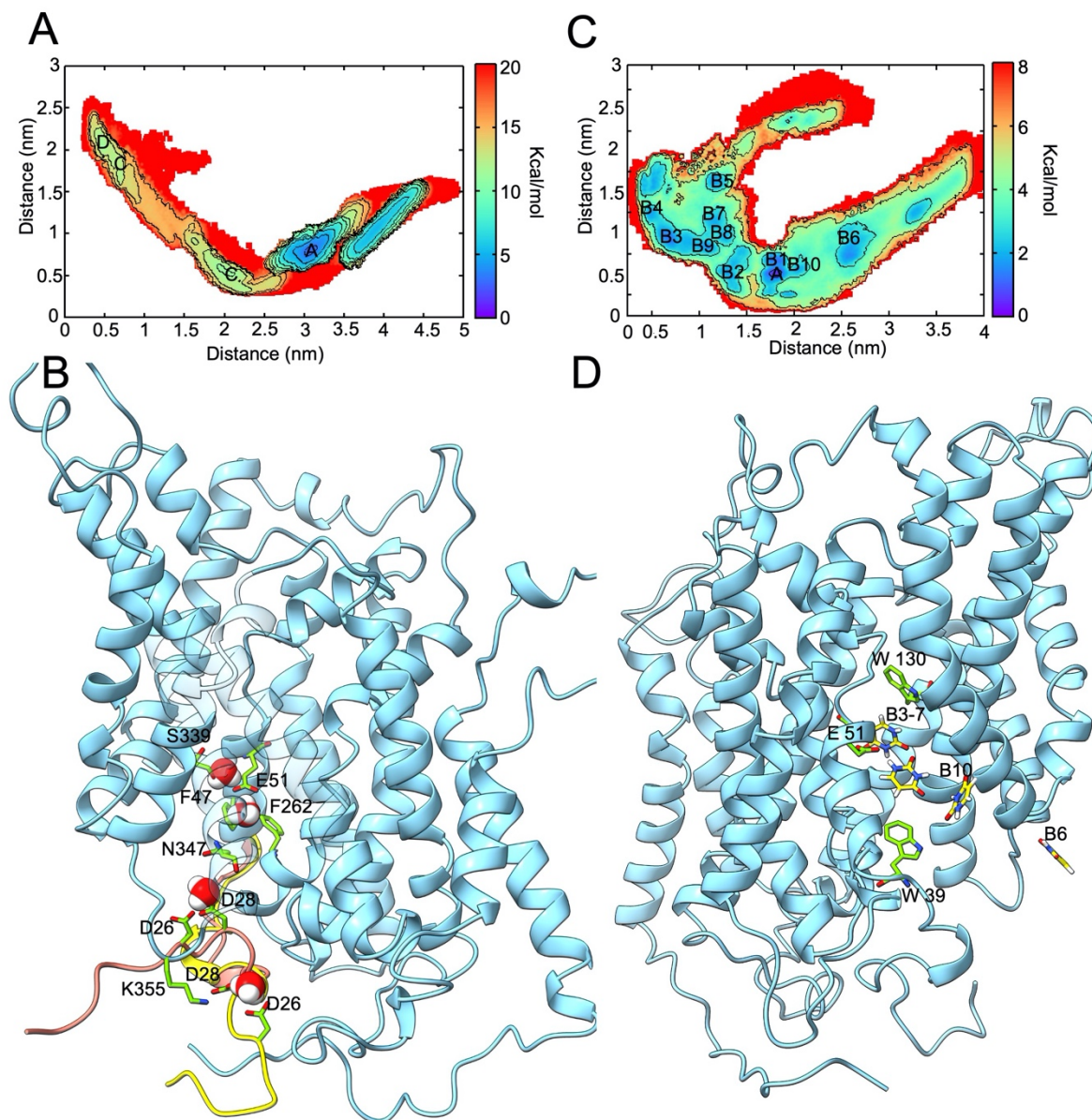


Figure 3.1.13: (A) The funnel investigating the internalization process of  $\text{H}_3\text{O}^+$  in FurE transporter. (B) The funnel investigating the internalization process of uracil in FurE transporter.

The results indicate that hydronium can move towards the intracellular part of FurE passing through different binding modes (**Figure 3.1.14A, 3.1.14B**). First, the interactions between  $\text{H}_3\text{O}^+$  and T336 are disrupted and a H-bond with S339 is created, while the salt bridge with E51 is preserved. This state corresponds to the minimum D in the FES presented in **Figure 3.1.14A**. Afterwards,  $\text{H}_3\text{O}^+$  binds in a cleft formed by F47, F262 and E51, corresponding to minimum C (**Figure 3.1.14A**). Next,  $\text{H}_3\text{O}^+$  loses the interaction with E51 and it binds to D28, D26 of the N-terminal LID<sup>94</sup> and N347, corresponding to minimum B of the FES (**Figure 3.1.14A**). Lastly,  $\text{H}_3\text{O}^+$  reaches the lowest energy minimum A, in which it is bound to D28 and D26 (**Figure 3.1.14B**), and then it is completely released in the cytoplasm. The movement of E51 along with the  $\text{H}_3\text{O}^+$  unbinding process prompts the disruption of the E51-K199 and D28-R264 interactions (**Figure 3.1.9A**). Generally, our simulations denote that the flexibility of the FurE N-terminal LID plays a major role in the hydronium release in the intracellular solute.



**Figure 3.1.14:** The unbinding process of  $\text{H}_3\text{O}^+$  and uracil characterizing the exiting route to the cytoplasm. (A) The BFES of  $\text{H}_3\text{O}^+$  internalization course. The separation between contours is 2 kcal/mol. (B) The binding sites of  $\text{H}_3\text{O}^+$  cation in FurE transporter down the internalization pathway as derived from the low energy states in the BFES in A. In orange is represented the LID when  $\text{H}_3\text{O}^+$  is bound in D26, D28 and N347, while in yellow when  $\text{H}_3\text{O}^+$  is released in the cytoplasm guided by D26, D28 and LID conformational change. (C) The BFES of uracil internalization pathway. The split between contours is 2 kcal/mol. (D) The intermediate states of uracil internalization pathway while exiting FurE transporter as derived from the BFES in C.

### 3.1.3.7 The internalization pathway of uracil

After hydronium is released, uracil can also be discharged in the intracellular space. FM calculations were used to investigate the unbinding of uracil from the FurE IO state. We considered all the possible exiting pathways from the binding pocket up to the inner gate

(TMS5) (**Figure 3.1.13D**). In **Figure 3.1.14C, 3.1.14D** we present the FES and the energetically related poses assumed by the ligand along the exit pathway from the transporter, correspondingly. Throughout uracil unbinding process three residues play a critical role, namely W130, E51 and W39 (**Figure 3.1.14D**). W130 obtains a vertical geometry compared to the z axis of the membrane, hence shielding the access to the extracellular environment, whereas E51 favors the translocation of the ligand towards the TMS5 inner gate by forming a H-bond with uracil. Finally, W39 forms  $\pi$ - $\pi$  and T- stacking interactions with uracil validating the significance of W39 as highlighted by mutagenesis analysis. It is significant to note that standard MD simulations carried out on the metastable states recognized along the uracil unbinding, showed an unhindered release of the ligand in the solvent, demonstrating that this process is barrier-less and semi-diffusive.

### 3.1.4 Conclusions

The molecular foundation of coupled nucleobase/proton symport is an essential mechanistic aspect of NCS1 transporters. Very few cases of proton symporters have been structurally described and the structural and functional aspects of the transport have not been explained yet, neither computationally nor experimentally. We focused on the FurE transporter, to address this subject by performing extended free-energy simulations, targeted on the substrate translocation coupled with proton symport, which was performed by introducing a  $\text{H}_3\text{O}^+$  cation. Meanwhile, we also investigated the large conformational change of the FurE structure that complements these events. In this study, we simulated the proton symport by introducing  $\text{H}_3\text{O}^+$  as a second separate substrate. Proton binding could be explored by implementing MD simulations, considering all different protonation combinations of Asp, Glu and His residues or by applying Quantum-mechanics/molecular-mechanics (QM/MM).<sup>173</sup> However, in the FurE case the possible input combinations for standard MDs should be up to 32 and this is extremely computationally demanding. The same is true for QM/MM.

To reveal in atomic resolution, the binding sites of both uracil and  $\text{H}_3\text{O}^+$  cation in FurE, we used a state-of-the-art binding free-energy technique, named Funnel-metadynamics. Unlike other simulative binding techniques, FM allows the sampling of the binding process without knowing *a priori* the binding mode of the ligand and offers a detailed classification of all possible binding modes. FM is not dependent on the initial position, because the full binding pathway is sampled, leading to an accurate estimate of the absolute ligand-protein binding free energy and a systematic retrieval of all possible binding modes. All the molecules in the system are explicitly simulated, including water molecules, thus yielding information about their

dynamics during the binding process. The recently developed code of FM also incorporates a roto-translational matrix to accurately contemplate the entropy of the system. FM has an affordable computational cost for studying rare events such as binding processes, if compared to standard MD calculations. Finally, it is self-diagnostic, so the free energy calculation converges only to the exact value and critical points, such as non-optimal CVs or bad choice of simulation parameters, can be identified the calculation will not converge.

Our results illustrate that  $\text{H}_3\text{O}^+$  interacts with three negatively charged residues, namely E51, D28 and D26. The initial  $\text{H}_3\text{O}^+$  binding location was found to be exactly at the same position where  $\text{Na}^+$  is co-crystallized in Mhp1.  $\text{H}_3\text{O}^+$  binding, according to the FES of the OO-to-Occ simulation, stabilized the rather flexible *apo* structure in an intermediate conformation between the initially constructed models of OO and Occ, contributing to the conformational change from OO to Occ. The FurE- $\text{H}_3\text{O}^+$  interaction seemed to initiate rearrangements in specific residues. It induced the availability of Q134, leading to a binding site geometry prone to accept uracil, without significant protein conformational change. This is in agreement with previously reported data for Mhp1. Indeed, fluorescence quenching experiments in Mhp1<sup>87</sup> show that a tenfold increased affinity for benzyl-hydantoin occurs as a result of the sodium presence. Additionally, in Mhp1 distance measurements between pairs of spin labels showed that  $\text{Na}^+$  binding does not shift Mhp1 conformational equilibrium and deviates from the rigid-body motion model of the “hash” motif.<sup>174</sup> This dynamic behavior is different in other 5+5 folded transporters like LeuT, dDAT, hDAT, SERT, where  $\text{Na}^+$  binding favors an occluded conformation. FRET experiments in LeuT<sup>175</sup> suggest that a cation-dependent conformational rearrangement can form and stabilize the substrate binding site. However, in the FurE case, the lowest energy conformation shifts towards the occluded structure, only if both substrates are bound. In that case, both TMS10 (outer gate) and TMS5 (inner gate) are closed. Furthermore, a relatively small change was observed in the “hash” helices. Specifically, W130 relatively moves and probably induces a slight bend in the last part of TMS3 assisted by a glycine in residue position 132. These are followed by a similar bend of the first part of TMS8. Moreover, TMS9, following the movement of TMS10, induces a small shift to TMS4 (**Figure 3.1.9**). We identified key interactions between the ligand and FurE for the OO-to-Occ transition, namely uracil-W130-F385 and W130-F388, as well as minor ones, like the Q59-S386 interaction. This agrees with Mhp1 data, according to which, a similar network of interactions is observed and elucidates how ligand binding stabilizes the TMS10 tilt in the occluded conformation.<sup>176</sup> Notably, the functional importance of residues E51, W130, Q134, K199, K252, T336 and F385, derived from our *in silico* results was supported by mutagenesis data presented here. The



substrate release to the extracellular environment, in the Occ FurE transporter, is mainly hindered by W130 (TMS3). The prevention of substrate release to the extracellular solute is reinforced by polar residues S386 (TMS10) and Q59 (TMS1b), interacting mainly through the formation of H-bonds involving intermediate water molecules. In agreement with our results, in the LeuT transporter, the substrate release from the occluded state to the extracellular solute is prevented by Y108 (corresponding to W130 in FurE) and F253 (TMS6), while the structure is stabilized by the interaction of D404 (TMS10) and R30 (corresponding to Q59 in FurE) and two water molecules.<sup>28</sup>

a conformational change leading to substrate translocation occurs, once both the substrate and the hydronium are bound in FurE. Especially, according to the computed FESs, FurE assumes an intermediate structure between Occ and IO, in which case,  $\text{H}_3\text{O}^+$  cation should be free to move towards the intracellular domain, whilst D28 loses the interaction with R264 to engage the moving hydronium. At the same time, the interaction between E51 and K199 is lost and the first interacts frequently with K252, releasing the TMS5 intracellular gate. Also, there is a destabilization of the substrate in the binding site, as the interaction between R123 and D261 breaks, and the substrate interacts with both R123 and T254. These events induce a relative motion of TMS3, and subsequently of TMS4 and the first part of TMS5, followed by a major shift of TMS8. In this conformation, we detect bending of the first part of the unleashed TMS5 in position 204 (proline), opening an inner gate. This state can be considered as the inward-occluded conformation in which both uracil and  $\text{H}_3\text{O}^+$  cation are still bound. In both LeuT<sup>92</sup> and DAT<sup>177</sup>, based on experimental studies, the inward-closed (occluded) state is stabilized by a  $\text{Na}^+$  ion bound to the Na2 site (Na2 ion). Nevertheless, in LeuT-type transporters two sites have been located for sodium binding, playing key roles both in stabilizing transporter conformations, and in substrate binding. In fact, in LeuT, the leucine carboxyl group binds directly to Na1 and has been proposed that  $\text{Na}^+$  and substrate internalization are coupled.<sup>28</sup> In the FurE case, two positively charged residues exist in positions close to Na1 and Na2 sites. Particularly, the K199 side chain group is in the Na2 site, the position of  $\text{Na}^+$  in Mhp1, while K252 side chain displays a relative flexibility and very often is found close to the LeuT Na1 site, which does not exist in Mhp1.

it is evident that only if  $\text{H}_3\text{O}^+$  is released, FurE can reach the IO state, as only in such a case the FES is shifted to IO. As a result,  $\text{H}_3\text{O}^+$  must be the first to leave and this event triggers the transporter transition to IO and allowing uracil to be finally released, as similarly proposed also for DAT and LeuT.<sup>92,178–180</sup> Computational studies on DAT,<sup>42</sup> showed that the inward open conformation is obtained after the spontaneous release of the  $\text{Na}^+$  ion from the Na2 site. The

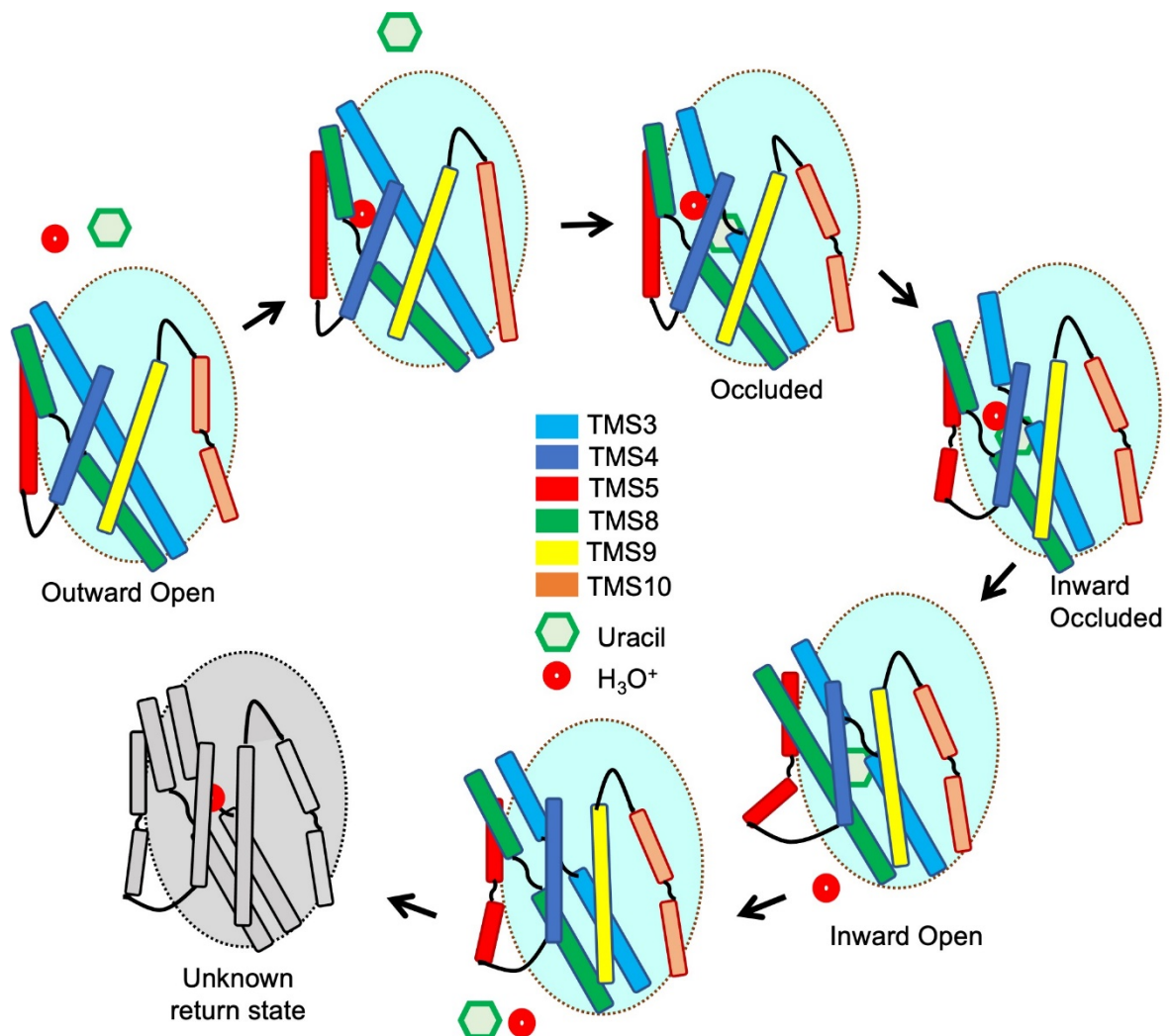
Na<sup>+</sup> unbinding happens after 800ns of unbiased MD simulations and is observed in 12 out of 50 runs of 1 $\mu$ s.<sup>178</sup> This is mainly determined by the number of water molecules that penetrate the functional sites of the transporter from the intracellular space. After abandoning the Na2 site, Na<sup>+</sup> is located between T269 (TMS5), N-terminal and D421 and D436 (TMS8). As suggested by computational and experimental studies, this process destabilizes the inward occluded state.<sup>181</sup> Internalization of hydronium implies neutralization of D28 and D26 and subsequent relocation of the N-terminal LID in agreement with our previously reported results.<sup>94</sup>

In the case that uracil is solely left inside the system, FurE is in the IO conformation and the substrate is prepared to leave the transporter. TMS5 (the inner gate) is open, TMS4, 8, 9 are completely shifted, while the upper part of TMS3 bends. H<sub>3</sub>O<sup>+</sup> release, disrupted the T336-H<sub>3</sub>O<sup>+</sup>- E51 interaction connecting the “hash” and “bundle” motives, allowing TMS8 to get distant from TMS1. Subsequently, the binding site residue positions are altered, and binding site interactions of uracil are not favored anymore. The E51-K199 interaction is also disrupted, causing TMS5 to bend in the maximum angle and consequently uracil is able to leave the binding site.

When FurE is in the *apo* form, obtains an energy minimum structure close to IO (**Figure 3.1.9**). This discovery agrees with the fact that Mhp1 has been crystalized also in the *apo* form conformation.<sup>165</sup> However, given that the FES minimum in the *apo* form is wide, the FurE transporter might undertake several alternative conformations between Occ and IO. The evidence arising from the FES, that the presence of hydronium stabilizes a FurE state close to Occ (**Figure 3.1.9C**) (e.g., TMS8 tends to assume the orientation of Occ) encourages the suggestion that H<sub>3</sub>O<sup>+</sup> binding could be vital even for FurE’s backward transition to OO. In such a case, FurE could function as antiporter, as proposed for other transporters like SERT, DAT and LeuT.<sup>42,182,183</sup> The broad FES curves in both OO-to-Occ and Occ-to-IO transitions and the flexibility witnessed in TMS10 and TMS5, supports the possibility that the new symporting cycle appears when the substrate binds in low probability conformations, shifting the equilibrium again to the Occ and finally to OO states. However, our calculations are not conclusive and further investigation is required.

Overall, in this study, we investigated the function of the FurE proton symporter and revealed molecular details of the conformational changes, fundamental for uracil/H<sup>+</sup> symport, using Metadynamics and mutational analysis. In **Figure 3.1.15** we represent a schematic representation of the mechanism. By analytically assessing the differences in the energy landscape of the transporter conformational transition under different combinations of proton

and substrate binding, we created a model relating anionic and cationic residues with specific proton binding and internalization and showed how proton translocation is required for uracil transport. Moreover, theoretical simulations identified residues critical for specific phases of uracil/ $H^+$  transport, which were validated by mutational analysis. Overall, the  $H_3O^+$ /uracil binding and transport shape the energy landscape by eliciting induced-fit conformational changes that cause consecutive movements of specific TMS in the hash and bundle domains, associated also with opening and closing of outer (TMS10) and inner (TMS5) gates. Our results assume that the hash motif helices are flexible and can tilt upon substrate binding in the Op-to-Occ conformational rearrangement, while in the Occ-to-IO case they exhibit important relative movement along with the opening of the inner gate and coordinated TMS8 and TMS4 shifts. In addition, we demonstrate that a crucial step for uracil transport is the intracellular release of  $H_3O^+$ , assisted by the flexibility of N-terminal LID sequence. This observation complements with critical details our understanding of the nucleobase/ $H^+$  coupling mechanism.



**Figure 3.1.15:** Schematic illustration of the FurE's "hash" motif helices and outer and inner gates (TMS10 and TMS5) relative to the bundle motif which is considered immobile during uracil and  $\text{H}_3\text{O}^+$  internalization. In the Outward Open (OO) state, FurE is in *apo* form.  $\text{H}_3\text{O}^+$  binding results in side chain rearrangement but it does not cause any change in the tertiary structure. Uracil binding prompts the closing of the TMS10 outer gate (salmon) and the kink and tilt of TMS8 (green) and TMS3 (cyan), respectively reaching the Occluded (Occ) state.  $\text{H}_3\text{O}^+$  moves toward the TMS5 inner gate (red), which slightly bends, while TMS3 and TMS8 also exhibit structural alterations, initiating the Inward Occluded (IOcc) state. Once  $\text{H}_3\text{O}^+$  is released in the intracellular environment, TMS5 bends more, whereas TMS8 is no longer tilted and moves away from the "bundle". TMS4 and TMS9 are shifted by TMS5, 8 and 3 bending introducing the Inward Open (IO) conformation. After the release of both  $\text{H}_3\text{O}^+$  and uracil, TMS5 slightly returns to the previous bend position. An inward-facing unidentified return state probably introduced by a  $\text{H}_3\text{O}^+$ , is represented in grey.

## 3.2 NCS2/NAT family transporters: Structural and Functional characterization of the UapA transporter.

### 3.2.1 Foundation of the study

NCS2/NAT proteins are  $\text{H}^+$  or  $\text{Na}^+$  symporters responsible for the uptake of purines, pyrimidines or related metabolites in bacteria, fungi, and some plants. In particular, the UapA transporter of *Aspergillus nidulans* is specific for the transport of xanthine and uric acid in the fungi cells. The dimeric UapA as described above, consists of 14 transmembrane segments folded in a rigid core domain and a flexible gate domain and its biological function is based on the "elevator" transportation mechanism, which implies the relative motion of the gate towards the core. The entire transportation process is divided in four protein states: Outward-Open (OOp), Outward-Occluded (OOc), Inward-Occluded (IOc), and Inward-Open (IOp). The substrate approaches the OOp-dimer, and the transporter progressively evolves into the IOp conformation allowing the molecule to enter the cytoplasm. Importantly, a crystal structure of the IOp conformation was recently resolved and several mutations have been reported, identifying key residues in the transportation mechanism. However, the exact translocation pathway remains elusive, and the transportation mechanism is still unclear.

The key objective of this work was to generate a theoretical model describing the structural alterations of UapA by utilizing cutting-edge Molecular Dynamics simulations including Metadynamics in order to correlate computational and experimental data and gain insight into this large-scale and very complex phenomenon. Apart from the IOp conformation, the other checkpoint steps of the conformational pathway, namely OOp, OOc and IOc, were built using targeted MD simulations using as templates the crystal structures of the Band3, Bor1, UraA homologous transporters, respectively. The large-scale conformational changes can be computationally reproduced with reasonable computational cost using Metadynamics

and specifically, the Path Collective Variable approach in the space of the root mean square deviation of selected protein's backbone.

### 3.2.2 Construction of the three UapA models

The key step in Metadynamics is the identification of the initial and final structures of a protein conformational change as well as the selection of the collective variable. In this aspect in order to create the path, between the outward open and inward open conformation we had to create the four major structures Inward-open, inward-occluded, outward-occluded and outward-open. To this end, apart the inward-open determined by X-ray crystallography<sup>58</sup>, we used as templates three transporters who had the best ranking similarity with UapA according to p-Blast and had been structurally resolved with X-ray crystallography.

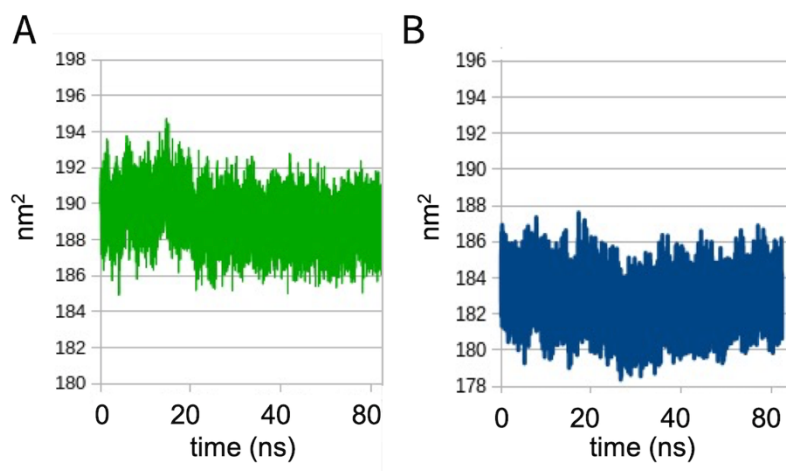
UraA was used as query for the Inward-Occluded<sup>56,57</sup> conformation, Bor1<sup>49,50</sup> for the Outward-Occluded and Band3<sup>51</sup> for Outward-Open. Unfortunately, the length of the helices of these query transporters was not identical to that of UapA. Targeted Metadynamics was the method chosen to create the models as the most appropriate because this way it was possible to keep the length of the UapA helices intact. In detail, the important, conserved residues were kept aligned as anchors and the rest were accordingly matched. A stable force was applied on the Ca carbons of the UapA crystal structure directing them towards the corresponding ones of the query crystal structure. The newly produced UapA models in IOcc, OOcc, and OOp conformation were inserted in a lipid bilayer solvated in explicit waters and were equilibrated and subjected to 100ns stabilization MD simulations.

### 3.2.3 The UapA crystal structure

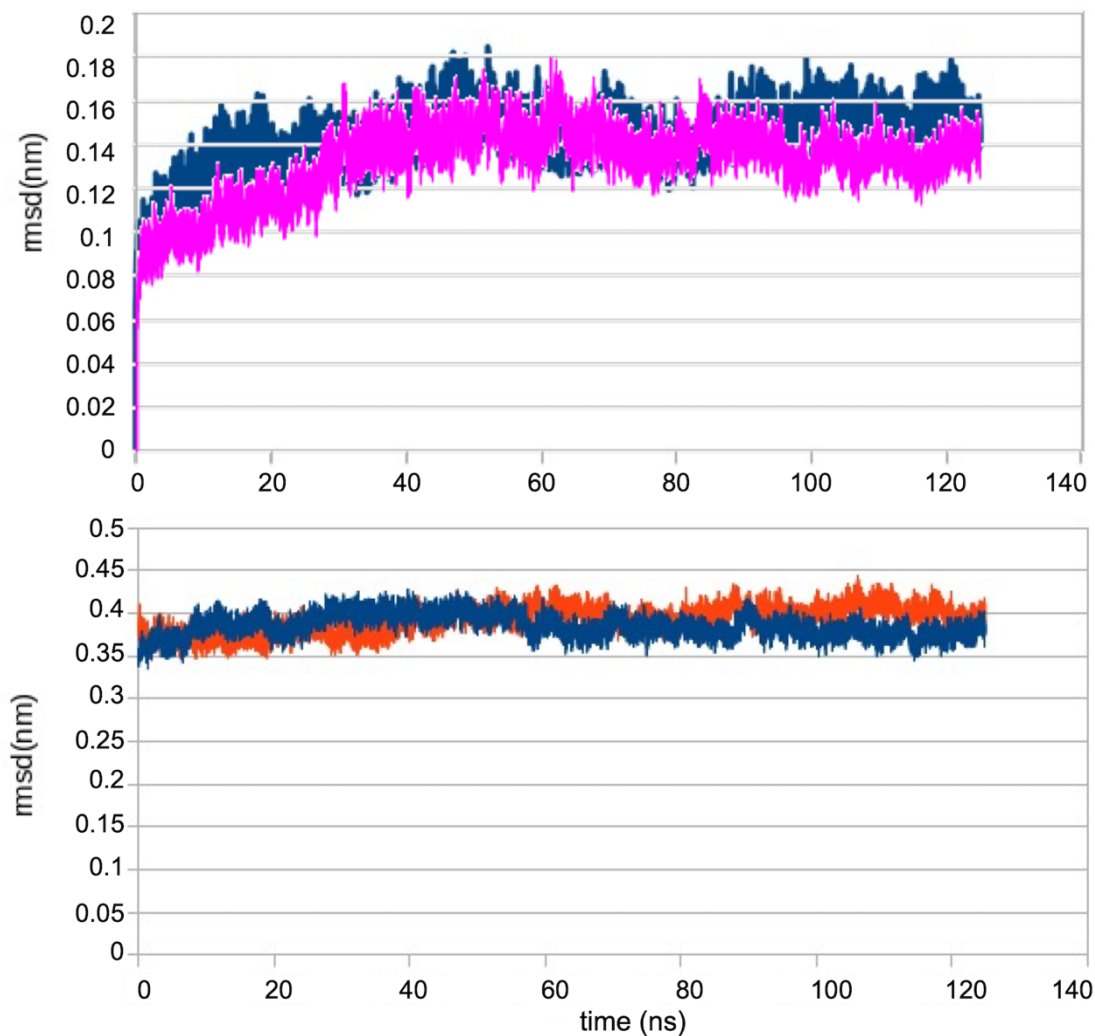
In order to evaluate the constructed UapA models, a series of studies were made to characterize the UapA crystal structure. Standard MD of 100 ns was executed, and the solvent accessible area was calculated for each protomer. As displayed in **Figure 3.2.1** the solvent accessible area in protomer/chain A obtains higher values compared to protomer/chain B.

Moreover, the UraA and UapA crystal structures' RMSD of the Ca carbons of the two protomers was compared in a 125 ns MD (**Figure 3.2.2**), indicating that they act in a quite similar way. This led to investigate more this behavior using Principal component analysis on the movement of each protomer. As a result, the protomers were found to be asynchronous and to behave differently, revealing an important aspect of the "elevator" mechanism. During MD

simulations, the xanthine-substrate leaves the binding site of UapA in less than 20 ns, in agreement with previous work<sup>58</sup>.



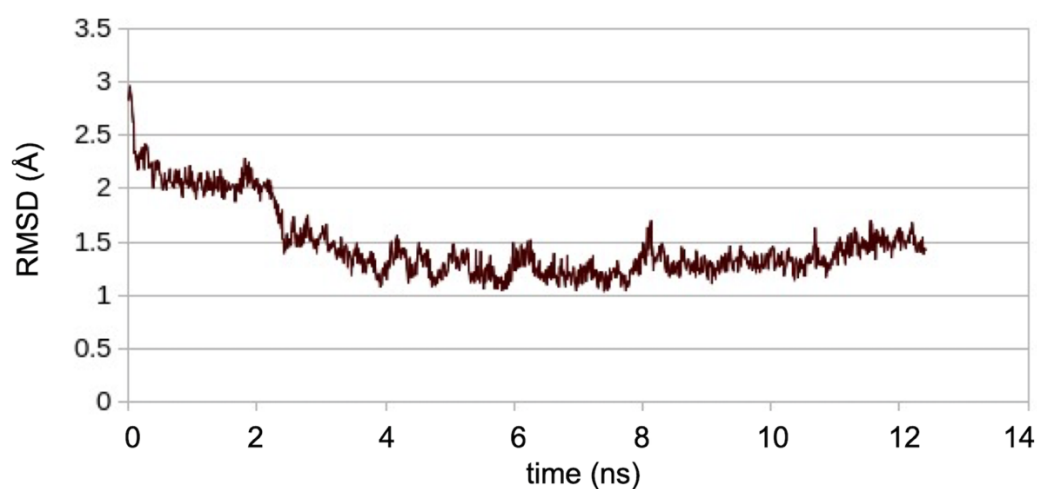
**Figure 3.2.1:** Solvent accessible area in protomer A (green) and B (blue).



**Figure 3.2.2:** The RMSD of UapA (upper panel) and UraA (lower panel) during 125ns MD simulation. In UapA the two protomers are colored in blue and magenta, while in UraA in blue and orange.

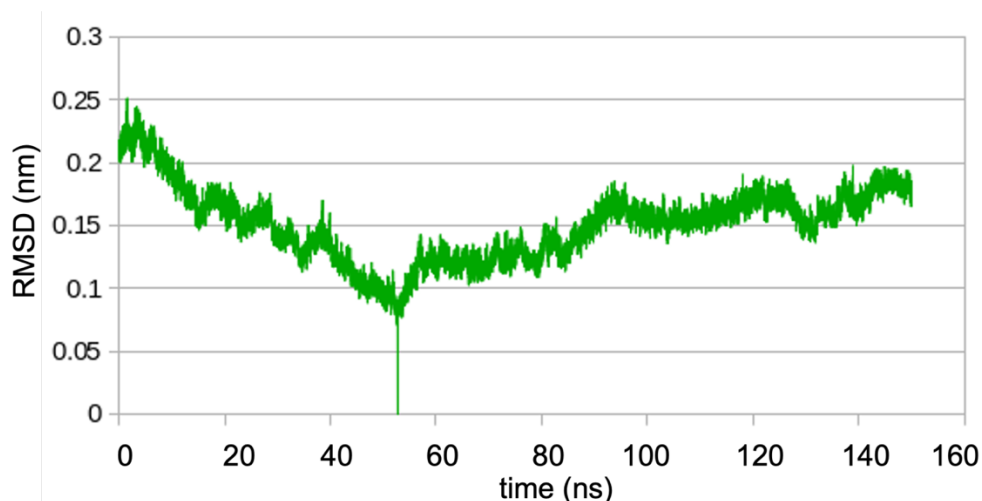
### 3.2.4 The Inward-Occluded Model structure

The construction of the Inward-Occlude model was initiated by applying a stable force on Ca atoms of the UapA crystal structure towards the corresponding atoms of UraA query. The RMSD of these atoms during this steering process is presented in **Figure 3.2.3**, calculated on the average structure. In detail the value of the force increased steadily for 4 ns, then it was constant for 6 ns, and it went back to zero for the rest 2 ns. The final IOcc model was the central structure obtained from the structures produced during 1-3 ns (**Figure 3.2.3**). The pka of important binding site acidic residues was calculated using PROPKA 3.0 and was found equal to 6.8 for E356 and 8.5 for D360. Thus, E356 was negatively charged and D360 neutral during the simulations.



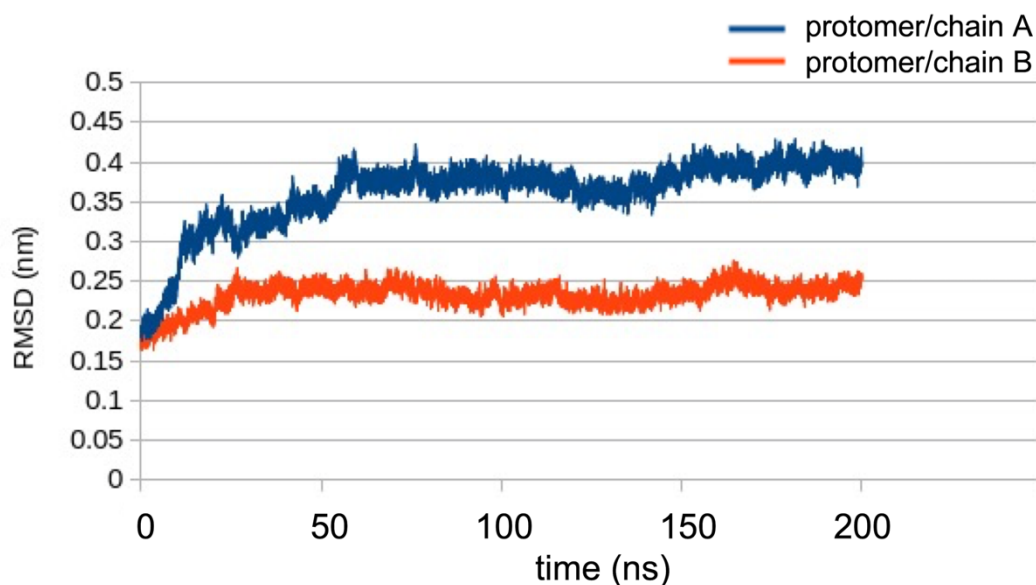
**Figure 3.2.3:** RMSD of the Ca atoms of the UapA during the steering process to Inward Occluded conformation.

Following the steering process, the IOcc UapA model underwent a series of computational tests in order to assess the stability of the structure, the first being the RMSD calculation during 150 ns MD (**Figure 3.2.4**). The structure was rather stable as the average RMSD was 0.16 nm and the maximum value marked was 0.25 nm.



**Figure 3.2.4:** RMSD calculation of the Ca of the inward Occluded structure of UapA.

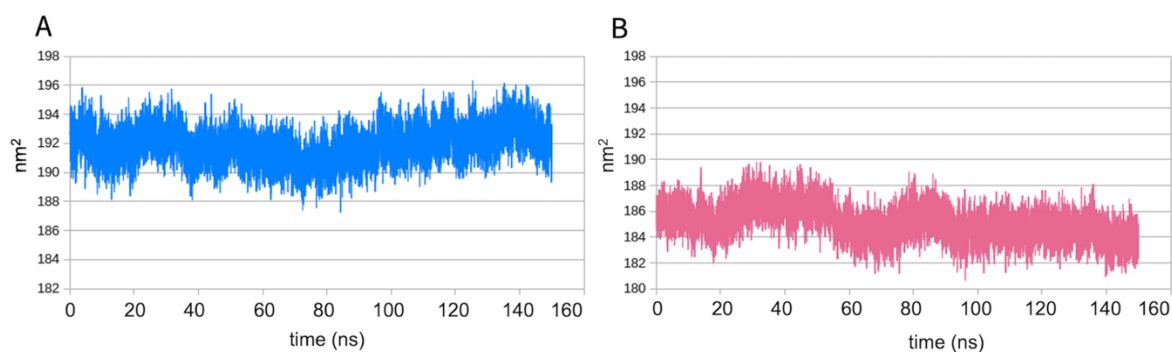
Additionally, each protomer/chain was steered to the model and the reference UraA structure and then standard MD of 200 ns followed. The two protomers deviate in different level from the query (**Figure 3.2.5**). Protomer A has an average RMSD of 0.37 nm, while protomer B 0.23 nm. This result is indicative of the different behavior of each protomer in the dimeric UapA and UraA transporters, in accordance with results presented above concerning the crystal structure of UapA.



**Figure 3.2.5:** RMSD of UapA Inward occluded structure calculated separately for each protomer upon the corresponding UraA protomer.

These findings are also supported by the solvent accessible surface area calculated (SASA), presented in **Figure 3.2.6**. Protomer A is more accessible to the solvent compared to protomer B.





**Figure 3.2.6:** Solvent accessible area calculated for each one of the two protomers of the UapA Inward occluded structure. (A) Protomer A is depicted in cyan and (B) protomer B in pink.

### 3.2.5 The Outward-Open model structure

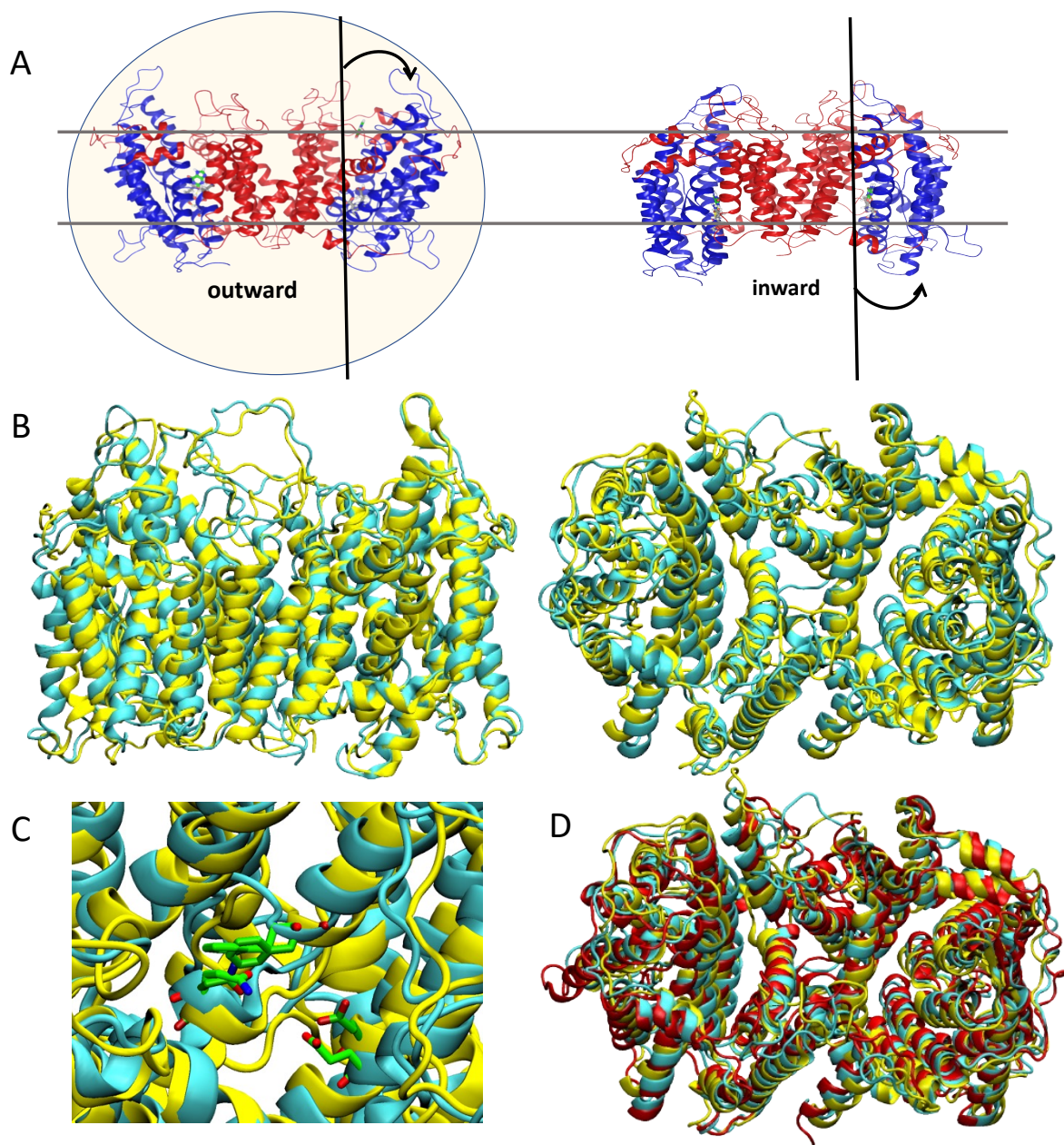
The Outward-Open and Outward-Occluded models were also constructed as described in the Methods section, using as templates the Band3 transporter and Bor1, respectively. One of the most important states, the Outward-Open conformation, has not been yet resolved with X-ray crystallography nor with any other method, as in this state happens the first step of the substrate transportation. Our model structure for the Outward-Open UapA (**Figure 3.2.7A**) is rather open towards the extracellular and the substrate binding site is exposed to the solute. Later on, we will present the conformational comparison between the 4 stages, and we will show that the Outward-Open conformation probably has a key role in specificity, as the interactions of important residues are modified.

The model structure of the Outward-Open conformation underwent a standard MD simulation of 100ns during which the structure was rather stable. In **Figure 3.2.7B** UapA Outward-Open model structure and centroid structure of the Outward-Open UapA 100ns standard MD were aligned. It is rather clear that the two structures are very well aligned and differ in TMS2,7,14 with an average RMSD of 1.4 Angstrom (**Figure 3.2.8**), which is rather low for a big protein like UapA. The same holds when these two structures are aligned with the one at the end of 100ns simulation (**Figure 3.2.7D**). Moreover, the residues of the binding site namely E356, Q408, F155 match perfectly between the model and the centroid following the good alignment of the helices (**Figure 3.2.7C**). These facts signify that the model constructed for UapA in Outward-Open conformation is very stable and hence quite accurate.

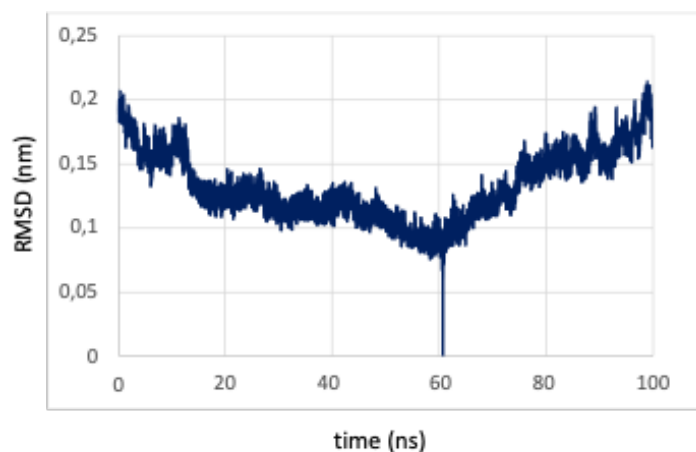
Inspecting visually the 100ns MD which also contains xanthine constrained in the binding mode of the crystal structure, we noted some important residue interactions. At the

cytoplasmic part of UapA R481 (TMS13) of the one protomer interacts constantly with E366 (TMS8) of the other protomer forming a salt bridge (**Figure 3.2.9A**), an interaction which is absent in the Inward-Open crystal structure, in which case we have found that R481 plays a major role being the residue responsible for the translocation of xanthine from the binding site to the intracellular solute, serving as an inner gating element<sup>58,76</sup>. In addition to that we have seen that in the apo Inward-Open systems, Q408 that in this case cannot form any bond in the binding site, rotates towards TMS6 and quite often interacts with R481 of the TMS13 of the other protomer which is present in this cavity. These observations lead to the conclusion that for the substrate to be stabilized in the binding site during the Outward-Open to Occluded conformational change, R481 should not be in the binding cavity, thus D366 locks it in a stable interaction. This interaction seems that could also stabilize the closed state of the intracellular part of the transporter and thus the Outward-Open conformation as well as contributes to the dimer formation since it interacts with cross-protomer residues. This way R481 role could explain and support the fact that UapA is a functional dimer. To support these, deletion of R481 or relocation of its position by the insertion of two amino acids (Ala-Gly) before it, resulted in UapA instability and increased turnover<sup>8</sup>. Combination of R481 substitutions with substitutions in the proposed Outward-facing gating elements (e.g. F528S or T526M -TMS14-) resulted in an additive effect in enlarged specificity. Interestingly, the arginine in this residue position is conserved only in the fungal NATs, whereas bacterial, plant and animal homologues, have Leu, Ile, Thr or Lys residues. Another interesting electrostatic interaction is the one between R417 (TMS10-core) and E286 (TMS6-gate) (**Figure 3.2.9B**), which is present only in the Outward-Open conformation because as seen in **Figure 3.2.9C**, TMS10 has moved upwards by two helical turns compared to the Inward-Open conformation in which case the two residues are in greater distance as the transporter is open to the cytoplasm and the distance between the core and gate domain is maximum. So, R417 seems to regulate the core-gate interaction, distance and sliding and as a result the conformational change. Specific substitutions of R417 proved crucial in determining the high affinity of UapA for uric acid compared to xanthine, by increasing the binding affinity for uric acid<sup>2,127,128,184</sup>. The residues of the binding site, E356, Q408 form with xanthine the interactions spotted in the crystal structure (**Figure 3.2.9D**). As only exception arises F155 which does not form T-stacking interaction with the xanthine but is rather pointing towards the extracellular and rarely interacts with F528 (TMS14), which is pointing towards the binding site (**Figure 3.2.9D**). This behavior of F155 is very rational, considering that the substrate is not expected to strongly bind in the binding pocket in the Outward-Open conformation. Thus, it seems that the last interaction to shield the binding mode

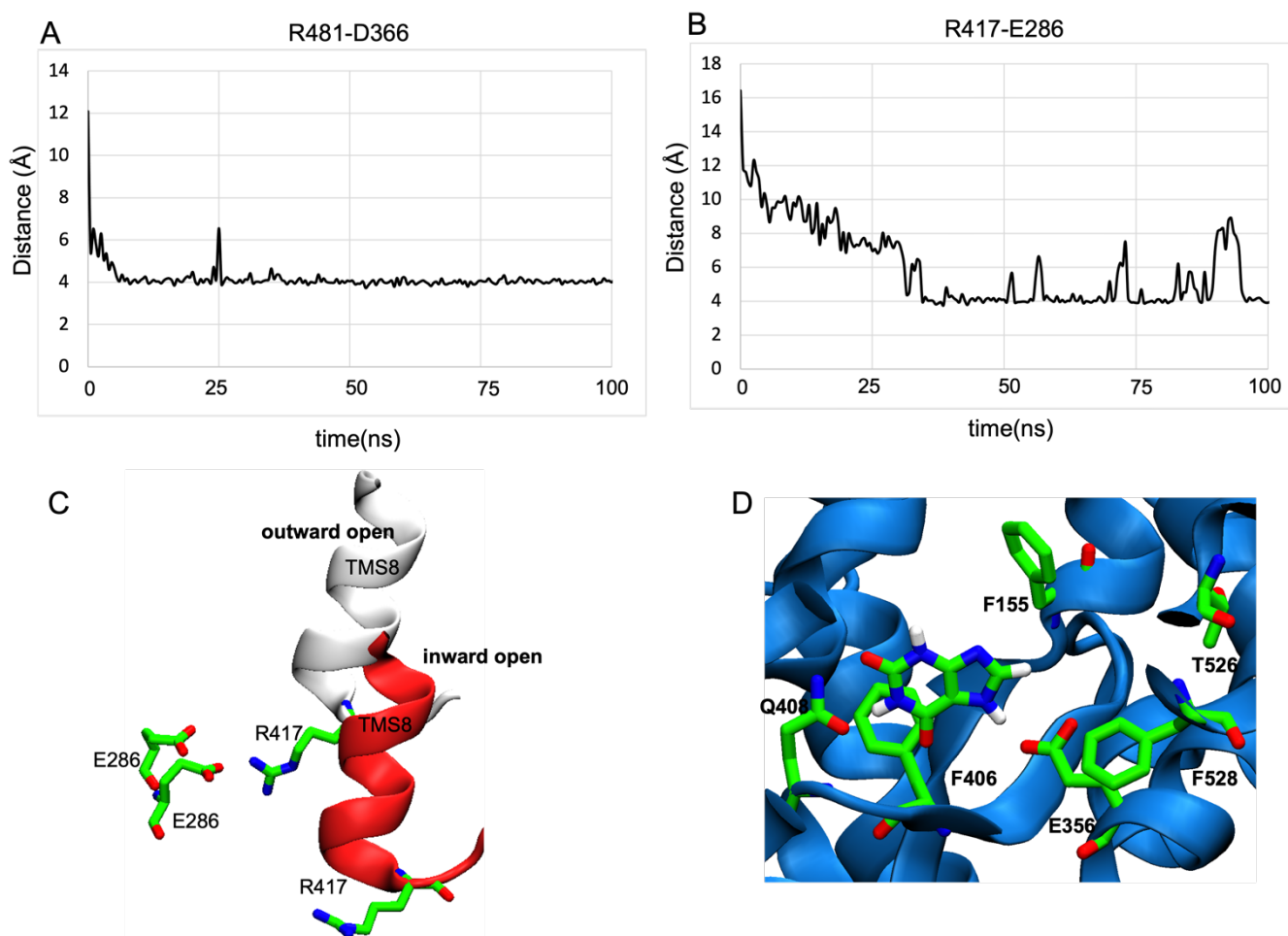
of xanthine in the binding site in the Occluded state is that of F155, which prevents the substrate from escaping to the extracellular solute. It has been suggested that F528 and Q408 of the NAT signature motif interact functionally<sup>185</sup> as the mutation F528S was a suppressor mutation of the UapA-Q408E mutant. Moreover, T526M, T526L substitutions when combined with mutations in Q113 resulted in growth on 2 mM adenine. In detail, growth of the double mutated UapA on high concentration of 2 mM of adenine or hypoxanthine, did not considerably increase the relative affinities for these non-natural substrates<sup>1,8</sup>. Based on these findings and their topology, it has been proposed that these residues act as outward-facing gating elements that select the purines allowed to access the substrate binding site of UapA, which could explain why specific substitutions of these residues might lead to enlarged substrate specificity<sup>8,126</sup>.



**Figure 3.2.7:** (A) The Outward-Open conformation of UapA derived from Homology Modeling, compared to the Inward-Open derived from X-ray crystallography. (B) Initial UapA Outward-Open model structure and centroid structure of the Outward-Open UapA 100ns standard MD aligned. (C) The binding site residues, E356, Q408, F155, of the Outward-Open model structure (yellow) and the centroid of the 100ns MD of the Outward-Open UapA model structure (cyan). (D) Superposition of the Outward-Open model structure (yellow), the centroid of the 100ns MD of the Outward-Open model structure and the Outward-Open UapA model after 100ns of MD.



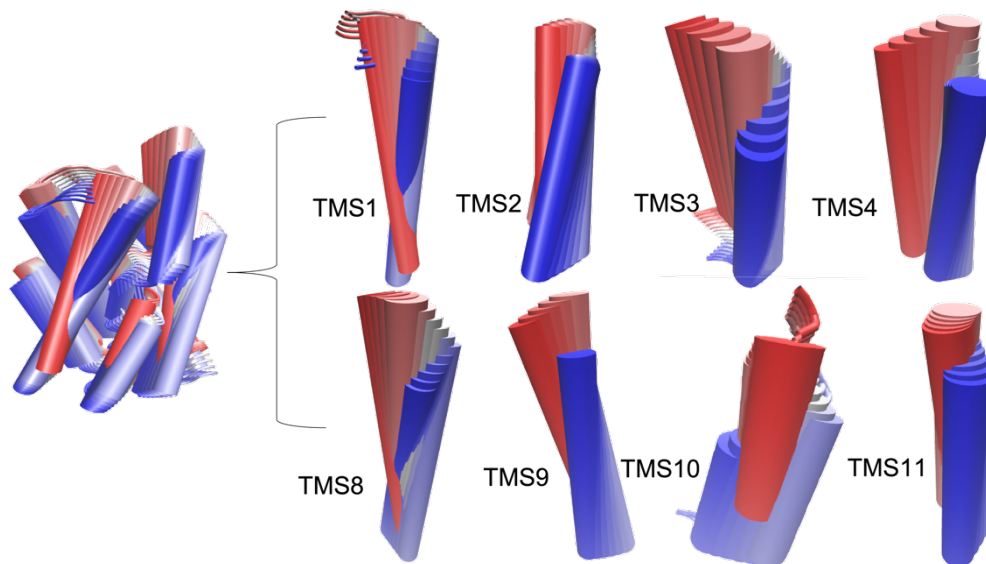
**Figure 3.2.8:** The RMSD of the Ca atoms of the helices calculated for the 100ns stabilization MD of the model UapA in Outward-Open conformation on the centroid structure of the same simulation.



**Figure 3.2.9:** (A) The distance calculated between C<sub>g</sub> of D366 and C<sub>z</sub> of R481 during the 100ns MD simulation of the Outward-Open UapA model structure. (B) The distance calculated between C<sub>d</sub> of E286 and C<sub>z</sub> of R417, during the 100ns MD simulation of the Outward-Open UapA model structure. (C) Comparison of the position of R417 in the Outward-Open and Inward-Open conformation, which regulates the formation of a salt-bridge between R417 and E286. (D) Binding site residues' topology in the Outward-Open UapA as derived from 100ns Md simulation.

### 3.2.6 Comparison of the four UapA models

A comparison of the UapA models and the relative movement of the helices highlights the “elevator” mechanism (**Figure 3.2.10**). As it can be seen the transmembrane segments

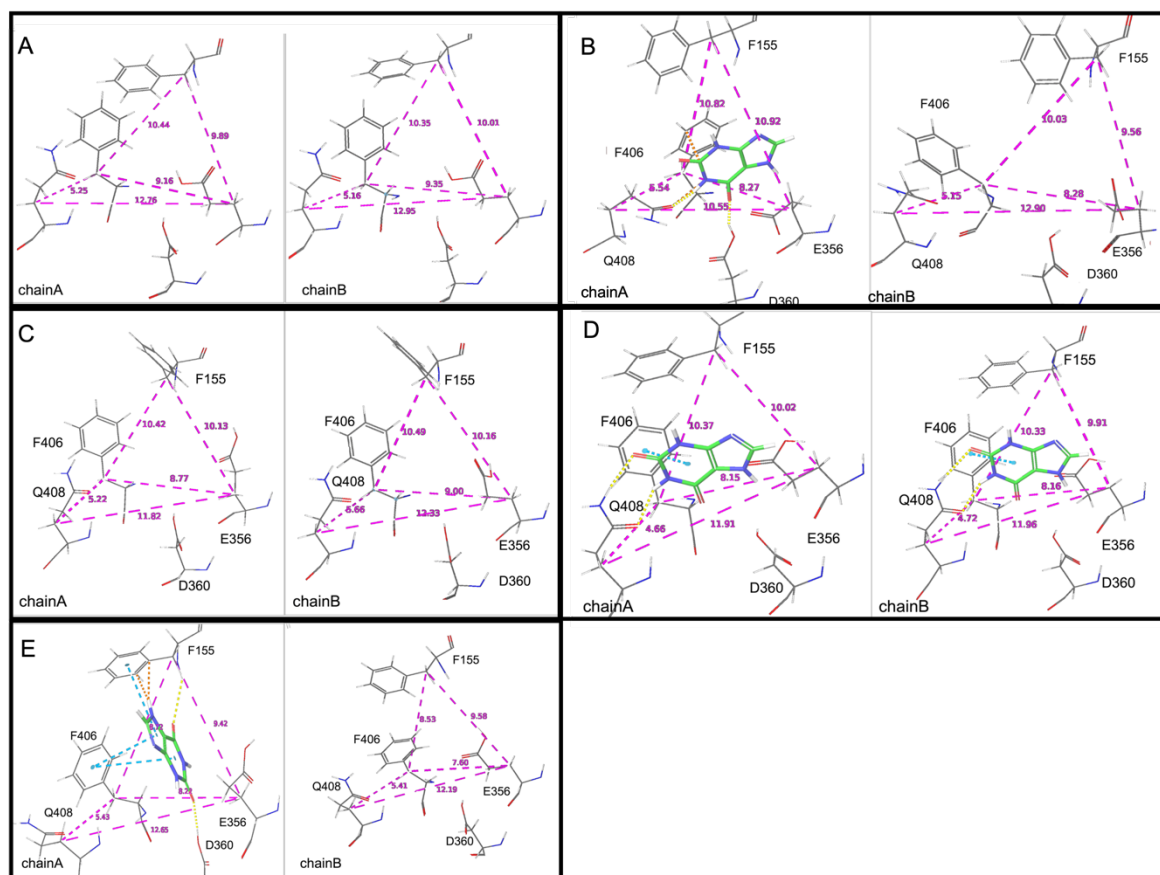


**Figure 3.2.10:** The conformational change of UapA from Outward-Open model (red) to Inward-Open (blue) crystal structure passing through the Occluded state, depicted on the core domain of a single protomer.

including the binding pocket (TMS8, TMS10, TMS3) are the ones displaying the larger conformational rearrangement.

Hereafter follows a comparison of the binding sites of the three conformational states that the substrate can be present or absent initiating two different states of the binding site, with respect to the distances of the Cb atoms (**Figure 3.2.11**). The distances are measured in Angstrom.



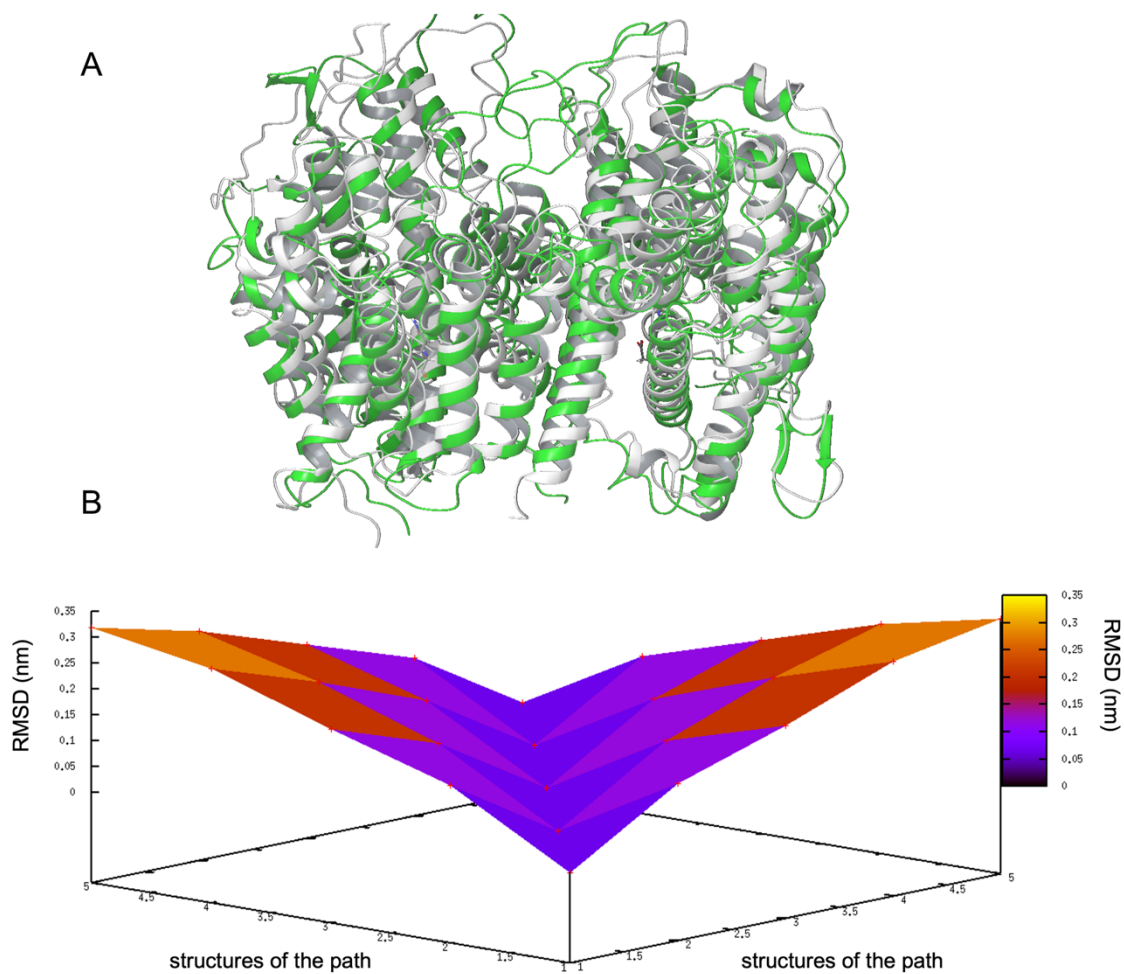


**Figure 3.2.11:** Comparison of the binding sites of the four conformational states with respect to the distances of the Cb atoms. (A) crystal structure of UapA with both substrates in, (B) UapA crystal structure with one substrate in and one absent, (C) UapA outward occluded without substrates (D) UapA outward occluded with both substrates present, (E) UapA outward open with one substrate present and one absent.

### 3.2.7 Metadynamics simulations for the IOp-to-IOcc transition step.

#### 3.2.7.1 The Free Energy Surface of the IOp-to-IOcc transition step.

As described above, starting from the UapA crystal structure the atoms were steered on the corresponding atoms of UraA template using targeted MD (**Figure 3.2.12A**). The intermediate structures were extracted from the steering process through RMSD calculation so that they create an optimal path of equally distant structures (**Figure 3.2.12B**). The method was parametrized adequately for these systems after several trials.



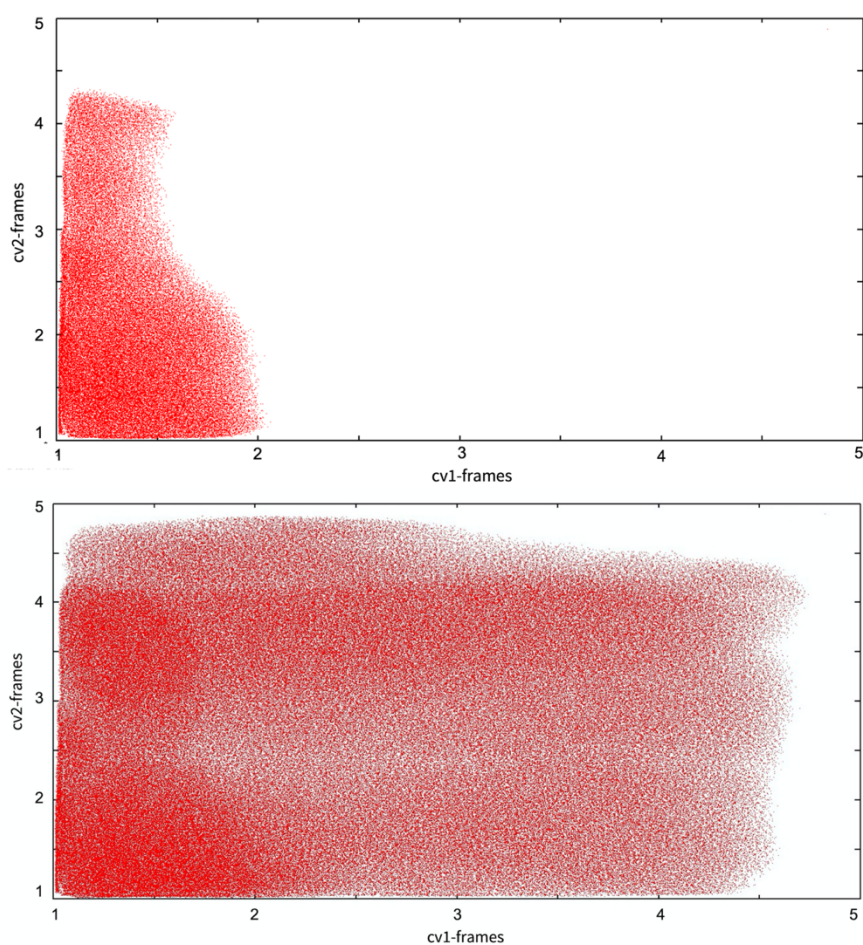
**Figure 3.2.12:** (A) UapA Inward occluded structure (green) superimposed to the UraA crystal structure (white). (B) The RMSD of the intermediate structures/frames between Inward Open and Inward Occlude structure of UapA, that were chosen to form the path connecting these two conformations.

One path was introduced for each chain/protomer of the dimeric transporter separately as the RMSD calculations presented above, surface accessible area analysis (SASA) and principal component analysis (PCA) on both the UapA crystal structure and the IOcc model, indicated that the two chains are asynchronous and behave differently. The progress along each path was the CV chosen for this Metadynamics simulation (p1.sss, p2.sss), so instead of aligning the whole dimeric structure on the other, each protomer was aligned on the corresponding of the template during the simulation. The width of the gaussians was calculated according to the fluctuation of each CV during a 100ns classical MD of the crystal structure (inward open conformation) and followed by a RMSD analysis to decide. The  $\lambda$  was equal to 135 for protomer A and 136 for protomer B, as indicated by the RMSDs between the five frames that identify the path. The ligand was docked in the binding pocket and an upper wall



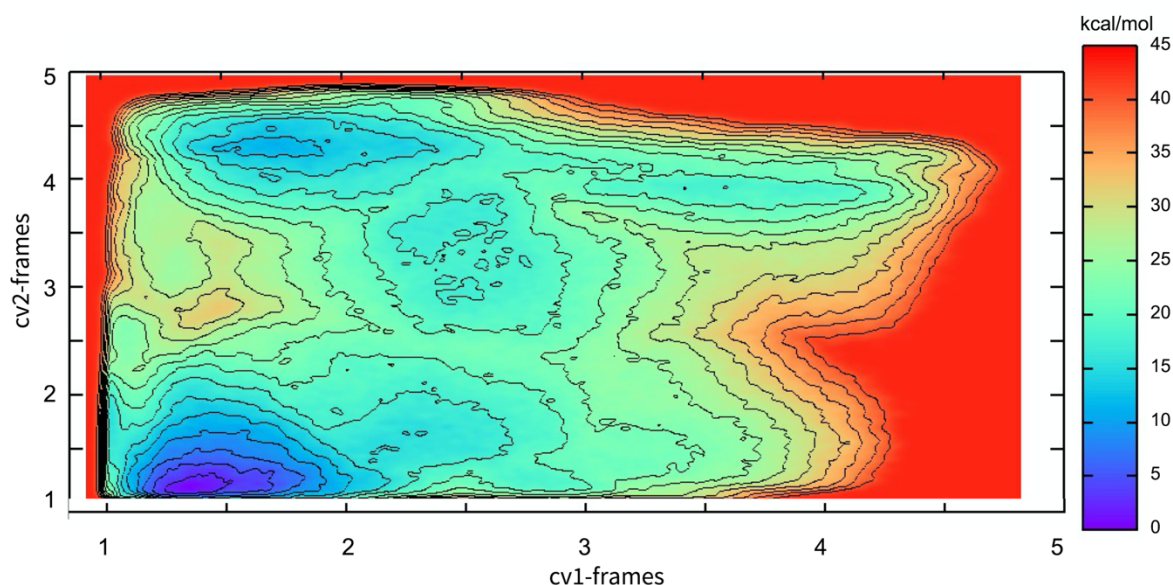
was applied in the crystallographic binding mode, in order to hamper the exit of the substrate from the inward open transporter. The steps were parametrized and finally validated for these transporters. The calculations were run on multiple nodes with GPU acceleration. All calculations were performed on a system that the dimeric protein was embedded in a membrane bilayer of DPPC lipids, ergosterol and explicit waters (TIP3).

It is worth noting that the progress along the path due to the implementation of the bias potential, was not identical for the two protomers, supporting their asynchronous movement (**Figure 3.2.13**).



**Figure 3.2.13:** The progress along the path in the start of metadynamics simulation (upper panel) and close to the end (lower panel) due to the implementation of the bias potential indicating that the two protomers perform an asynchronous movement as protomer B moves faster along the path.

The CV space of this 160000-atom system was fully investigated until it reached convergence and a semi-qualitative Free Energy Surface (FES), in a 1024 ns trajectory. The FES derived from this calculation is presented in **Figure 3.2.14**.

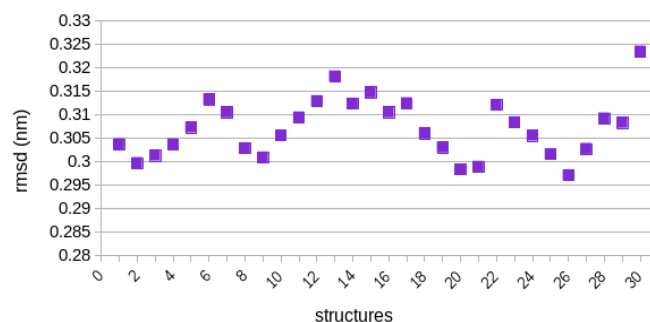


**Figure 3.2.14:** The Free Energy Surface corresponding to the Inward Open to Inward Occluded transition of UapA structure.

It appears that in the case studied, the global minimum of the Free Energy surface when xanthine is included in both protomers, is the Inward Open state, corresponding to all the points in the space  $f(cv1) = cv2 = [1...1.15]$  for  $cv1 = [1.3..1.4]$ . The next minimum with higher energy corresponds to the points belonging to  $f(cv1) = cv2 = [4.3, 4.4]$  for  $cv1 = [1.5, 1.8]$ . More explicitly, the next step corresponds to a structure with the one protomer in Inward Open state and the other in Inward Occluded. This fact confirms the previously mentioned result, that the two protomers are not synchronized. Other intermediate states are observed in the FES area  $[2.3-2.6, 1.4-1.6]$  and  $[2.4-2.6, 3.4-3.9]$  leading to the Inward Occluded state. In the case studied, in which only xanthine is included as a substrate, the Inward Occluded state is not conformationally identical to the UraA crystal structure (frame 5). It emerges closer to the frame 4 which has an RMSD of 0.05 nm (calculated on the Ca atoms) compared to frame 5 (as spotted on **Figure 3.2.12B**).

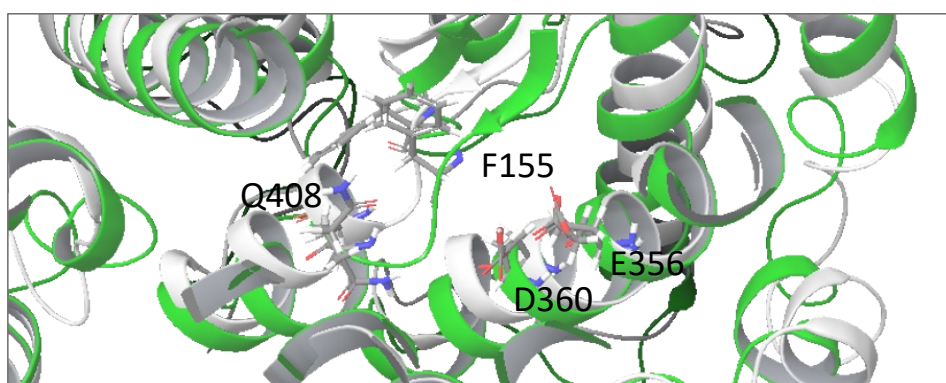
### 3.2.7.2 The Inward Occluded Structure through Metadynamics

Proceeding to the study of the Inward Occluded structure given by the FES, we extracted the trajectory frames narrowing down to the ones belonging to the  $[x:y]$  coordinates:  $[4.40-4.44 : 4.25-4.30]$ . The RMSD of these frames compared to the Ca atoms of the helices of the Inward Occluded modeled dimeric structure are represented in **Figure 3.2.15**. The one with the lowest RMSD was chosen as the Inward Occluded structure derived from the FES.



**Figure 3.3.15:** The RMSD of the frames derived from the FES compared to the Inward Occluded modeled dimeric structure as a result of the steering process.

Comparing the modeled structure for Inward Occluded and the one derived from the metadynamics calculations (**Figure 3.2.17**), we notice an RMSD of 0.29 nm on the overall structure, which is very low considering the huge dimeric transporter. Visual inspection of the aligned structures shows a great resemblance and no important displacement of the atoms of the helices. The FES derived IOcc structure was subjected to MD stabilization of 100ns and the plot of the RMSD compared to the average structure denotes that it is a quite stable structure.



**Figure 3.2.17:** The binding site of the modeled Inward Occluded structure (white) aligned on the FES derived one (green).

The highest energy of the Inward Occluded state compared to the Inward Open state, does not imply that UapA could not reach a state almost identical to UraA, but merely denotes that Inward Occluded structure is not favored with this substrate stoichiometry. In any case, further investigation should set light in this aspect.

Further, we surveyed the behavior of the important residues in the essential FES minima. In the Inward Open state, all binding site residues namely E356, F155, Q408, F406 interact steadily with the ligand. Though, the ligand upper wall is often affected meaning that

the ligand is not stable enough in the binding pocket in this conformation and attempts to escape towards the intracellular solute. This finding agrees with the results of previous studies in UapA crystal structure indicating that xanthine instantly abandons the binding pocket of Inward Open UapA, assisted by R481. In the Inward Occluded state, frames have been spotted displaying a T-stacking interaction between the F155 and F528. The latter even if it is away from the binding site, has been shown to play important role in specificity. All the other binding site residues interact constantly with the ligand.

### **3.2.8 The NAT signature motif and its structural and functional role**

The NAT signature motif is a conserved amino acid motif located in TMS10 of all NATs and includes residues critical for substrate binding and specificity (Q408, R417) or transport catalysis (N409, G411, T416), as shown in systematic mutational analysis on the UapA transporter. A sixth residue of the motif, F406, can also moderately contribute to specificity according to random genetic screens<sup>8</sup>. Generally, the five residues (Q408, R417, N409, G411, T416) are irreplaceable for wild-type UapA transport activity, however specific conserved substitutions might still possess low activity or substrate binding, though with modified kinetics<sup>1,76,127</sup>. In all nucleobases specific UapA homologues the NAT signature motif corresponds to the consensus sequence T/S/A/V405-F/Y/S/L406- A/S/T/G/V407-Q/E408-N409-X-G411-X-X-X-X-T416-R/ K/G417. Crystallography and Homology Modelling indicate that the NAT motif, located in the core domain, has a key role in the substrate binding site of UapA<sup>58,126</sup> or UraA<sup>56,57</sup>.

In our metadynamics calculations, the NAT signature motif behaves accordingly to the classical MDs of the crystal structure. Q408 forms a bidentate H-bond with the xanthine-substrate. A407 and F406 also interact with xanthine via backbone H-bonds and p-p stacking, respectively, while N409 and T416 seem to be involved in a dynamic H-bond network with specific residues in TMS1, TMS3 and TMS8.

### **3.2.9 Conclusion**

UapA transporter has been a great challenge to crystallize as it is a transmembrane dimeric protein, resulting in obtaining one crystal structure in inward-open conformation with xanthine in the binding pocket. Although this fact provided much information, yet the mechanism undergoing for substrate internalization remains unknown. Computational

approaches, like standard Molecular Dynamics are also not sufficient to simulate the whole conformational change, as that is computationally extremely demanding (1ms scale). Path-Collective variable metadynamics approach was a useful tool managing to highlight the UapA conformational transition from Inward-Open to Inward-Occluded state. The Free Energy Surface obtained, revealed that the two UapA protomers are not synchronized, so each one can be in different stage of the conformational cycle. That also means that it is not necessary that both protomers have the substrate in the binding pocket at the same time. This fact raises the hypothesis that the presence of the one substrate in the binding pocket could prevent the presence of the other to the other protomer, or that as long as a second substrate binds to the other protomer, the release of the first is favored. According to the results presented in this study, when both substrates are bound, the Inward-Open conformation is favored, while an Inward-Occluded state is less possible considering the higher energy values attained for this state in the specific stoichiometry of the system. In other words, it is rather possible that the occluded state of the one protomer is energetically favored when the other protomer is empty, while as long as another ligand approaches to bind in the second protomer, the Inward-Occluded state is less preferred compared to the Inward-Open. Further simulations with different substrate stoichiometries would set more light in this matter.

The Outward-Open model structure as resulted from Homology modeling and Targeted Molecular Dynamics, led to significant information about gating elements and about the elevator mechanism. In particular, the salt bridges R417-E286 and R481-D366, in the lower part of UapA seem to stabilize the outward conformation and regulate by their absence the substrate internalization.

### **3.3 Context-dependent cryptic roles of specific residues in substrate selectivity of the UapA purine transporter**

Adapted from Kourkoulou, A., Zantza, I., Foti, K., Mikros, E. & Diallinas, G. *J. Mol. Biol.* (2021).

#### **3.3.1 Foundation of the study**

As stated before, a noticeable feature of the NAT family members is the NAT signature motif located in TMS10<sup>127,184</sup>, in which are included residues critical for substrate binding and specificity namely Q408, F406, R417, or transport catalysis namely N409, G411, T416. In the UapA crystal structure, the xanthine binding mode includes a bidentate between xanthine and Q408, one backbone H-bond with A407 and  $\pi$ - $\pi$  stacking interaction with F406<sup>58</sup>. Further,

N409 and T416 are involved in a dynamic intramolecular H-bond network with specific residues in TMS1, TMS3 and TMS8. G411 has a different role associated with UapA-mediated transport as G411V is capable of binding, similar to the crystal structure, but not transporting the substrate<sup>127</sup>. Finally, R417 has been linked to uric acid binding affinity.

Our team has previously shown that the NAT motif appears in two major specificity-depending versions. In ascorbate-specific NATs the F406 in UapA is substituted by a serine (P406S) residue, while the Q408 by a proline (Q408P). Mutant versions of UapA introducing residues present in the L-ascorbate transporters proved to have basically no transport activity for either purines or L-ascorbate. At the same time, mutations in residues outside the major substrate binding site, selected by direct genetic screens, enlarge UapA specificity so that it can transport, besides xanthine and uric acid, other purines and uracil. These findings reveal that residues outside the main substrate binding site, are also involved in specificity regulation. Interesting examples are R481<sup>58</sup> of the other protomer which is near xanthine-binding site, T526 and F528 in TMS14 but also in the vicinity of the binding site. Mutations in these residues showed none or very little measurable affinity for hypoxanthine and adenine<sup>1,8,125</sup>. In this study we initiated an innovative approach to reveal the role of variable NAT signature motif versions on specificity via Molecular Dynamics and Docking calculations on UapA accompanied by mutational analysis to shed light into observations raised from the functional analysis of the mutants.

### **3.3.2 Context-depended role of the NAT signature motif in determining UapA specificity**

Seven UapA versions mutated in the first four residues of the NAT signature motif were studied with respect to their function, namely UapA-TY**A**Q, -T**F**AE, -T**Y**AE, -**S**YSQ, -**S**YSE, -**S**SSQ and -**S**SSP, where in bold are substitutions relative to the TFAQ (T405-F406-A407-Q408) motif present in the wild-type UapA. The broad specificity profile of the various NAT signature motif versions was investigated by Kourkoulou et al., using for the first time in UapA a  $\Delta 7$  strain, previously used to reveal cryptic functions of NCS1 transporters<sup>95,96,169</sup>, which genetically lacks all seven major endogenous nucleobase-related transporters (UapA, UapC, AzgA, FcyB -high-affinity cytosine/purine transporter<sup>168,169</sup>-, FurD and FurA - high-affinity uracil and allantoin transporters<sup>95,186</sup> - and CntA - general nucleoside transporter<sup>167</sup>). The previously used  $\Delta$ ACZ strain which lacks only UapA, UapC and AzgA nucleobase transporters<sup>96</sup>, but this  $\Delta 7$  strain does not transport any purines, pyrimidines, nucleosides, allantoin and nucleobase toxic analogues in a detectable level. As a result, the UapA mutant

versions genetically introduced in this strain could be analyzed for their ability to transport any purine-related substrate without background “noise”. Moreover, UapA mutants with too little transport activity can be measured due to the use of the strong *gpdAp* promoter.

Among the constructed mutants, UapA-TFAE, -TYAQ and -SSSQ have been previously functionally analyzed in a genetic background (*uapA- azgA- uapC-*) that permitted testing of purines only. Thus, it was known that residues at positions 406 and especially 408 (F/Y406 or Q/E408) are critical for UapA specificity. UapA-TFAE sequence, which is functional, has acquired the ability to bind (but not to transport) hypoxanthine/guanine, albeit with relatively high affinity. The UapA-SSSQ mutant has been shown to lack any measurable purine transport capacity<sup>8</sup>. Finally, this analysis showed that UapA-TYAQ has acquired ability for recognition of all natural purines, but with extremely low binding affinities ( $K_i \geq 1 \text{ mM}$ )<sup>187</sup>. It is interesting to note that in fungal and bacterial NATs that transport uric acid and xanthine 406 and 408 positions are occupied by a Phe and a Gln residue, respectively, while bacterial uracil transporters have a Tyr and a Glu at these positions. **SYSE** and **SYSQ** are found in many metazoan NATs, while **SSSP** sequence is found only in ascorbate-specific NATs. The rest of the mutants studied were constructed and studied to inspect the individual role of the corresponding residues.

The experimental part included growth tests and comparative uptake assays, performed by A. Kourkoulou. **Table 3.3.1** presents the results concerning the transport function and specificity of UapA NAT motif mutants. The comparative uptake assays with radiolabeled xanthine and growth tests, were in accordance. Based on these, UapA-SYSQ was found to be able to recognize and transport UA, X, AD and 5FU, and to at least bind with low affinity also HX and U, indicating it as the most efficient and broad-specificity UapA version. In contrast, UapA-SYSE, has lost its transport capacity for purines, but acquired higher capacity for 5FU accumulation, differing at only a single residue with UapA-SYSQ (i.e. Q408E). A negative epistatic interaction between Tyr406 and E408 is assumed since loss of transport capacity is mainly due to the combination of these residues, while these two mutations alone do not lead to total loss of purine uptake. These results undoubtedly support those complex interactions of residues in the NAT signature motif determine UapA specificity, making it rather hard to define them.

**Table 3.3.1:** Transport function and specificity of UapA NAT motif mutants

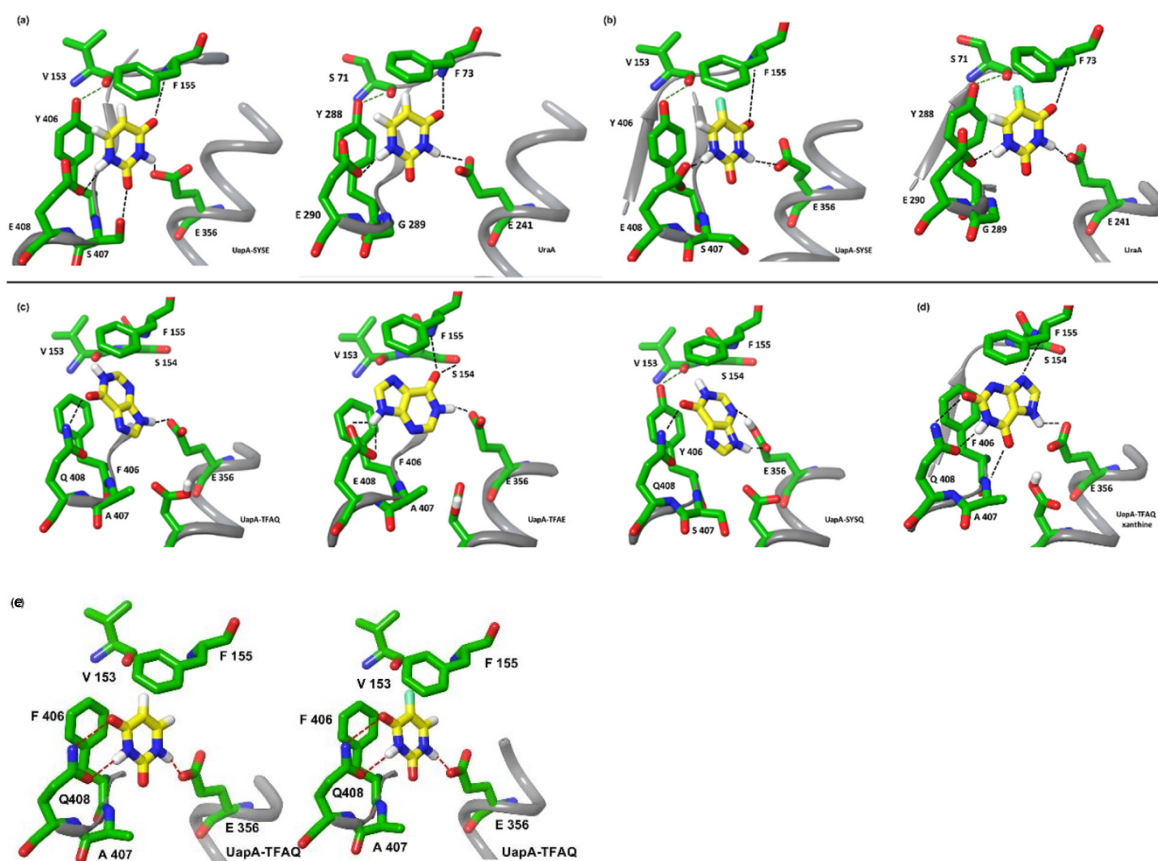
NAT motif version	UA	XA	AD	HX	5FU (U)	5FC (C)	5Fud (Ud)	AA	Comments on UapA function
TFAQ	+++	+++	-	-	-	-	-	-	Wt function/specificity
<u>TYAQ</u>	+++	+++	+	+	++	-	-	-	increased <i>binding/transport</i> promiscuity
<u>TFAE</u>	++	++	+/-	++	+/-	-	-	-	increased <i>binding/transport</i> promiscuity, reduced affinity and transport of UA, XA
<u>TYAE</u>	-	-	-	-	+/-	-	-	-	loss of purine <i>binding/transport</i> , but gain of moderate 5FU transport
<u>SSSQ</u>	+/-	-	-	-	+	-	-	-	loss/reduction of purine <i>binding/transport</i> , but gain of 5FU transport
<u>SYSE</u>	-	-	-	-	+++	-	-	-	loss/reduction of purine transport but gain of efficient 5FU transport
<u>SYSQ</u>	+++	+++	++	+/-	++	-	-	-	increased promiscuity for purines and 5FU
<u>SSSP</u>	-	-	-	-	+/-	-	-	-	loss of purine, some gain of very low 5FU transport

Underlined are amino acid replacements in mutant versions of UapA. +++ signifies wild type-like substrate binding and transport activity. - stands for 5% transport capacity of wild-type UapA and no indication of binding of nucleobases (up to 1 mM) or toxic nucleobase analogues (50–100  $\mu$ M). In between values, +/-, +, ++ reflect compromised binding/transport (10–50% of wild-type UapA). UA is uric acid, XA is xanthine, AD is adenine, HX is hypoxanthine, , 5FU is 5-fluorouracil, U is uracil, 5FC is 5-fluorocytosine, C is cytosine, 5Fud is 5-fluorouridine, Ud is uridine, AA is L-ascorbic acid (1%). This table is based on functional studies (growth tests, subcellular localization uptake measurements and transport competition assays) of different mutants expressed in a D7 genetic background.

### 3.3.3 Computational analysis gives insights on the specificity modifications of the various NAT signature motif versions

To better understand the specificity alterations of UapA, we performed molecular simulations of protein-substrate interactions (**Figure 3.3.1**). The models of UapA-SYSE, UapA-SYSQ, and UapA-TFAE have been constructed using as template the crystallized UapA structure in the inward-open conformation. Important differences were observed in the modelled structures, related to the binding cavity of the transporter due to the presence of the F406Y mutation but also to the calculated pKa of acidic residues critical for substrate binding (e.g. the invariable E356 and D360 in TMS8 and E408 in TMS10 of UapA-SYSE or UapA-TFAE).





**Figure 3.3.1:** UapA mutants-substrate interactions as resulted by Molecular Simulations. (A) UapA-SYSE-uracil and UraA-uracil interactions. The UapA-SYSE-uracil interaction is stabilized by the formation of four H-bonds (depicted as dashed lines) between N1H and N3H with E408 and E356 carboxylates, respectively, C2=O with S407 hydroxyl group and C4=O with F155 backbone, as well as, a p-p stacking formed between uracil with both F155 and Y406. Notice also that Y406 hydroxyl group forms a H-bond with V153 backbone. The UraA-uracil crystallographic structure is also presented on the right for comparison. (B) UapA- SYSE-5FU and UraA-5FU interactions. In both structures 5FU is oriented similarly to uracil with the F atom located in a hydrophobic cleft and C-F bond parallel to both F155 and Y406 rings. (C) Hypoxanthine interactions with UapA, UapA-TFA E and UapA- SYSQ. Hypoxanthine forms two H-bonds with the UapA binding cavity, one between N9H and E356 carboxylate and the second between C6=O and Q408 amide group. In UapA-TFA E, hypoxanthine N9H and N1H both form H-bonds with E408, and E356 carboxylates, respectively, while C6=O forms H-bond with S154 hydroxyl group and F155 backbone. In UapA-SYSQ hypoxanthine interaction E356 is protonated. Hypoxanthine H-bond interactions with the binding cavity are between N9H, N3 and E356, as well as, between C6=O and Q408. (D) Wild-type UapA-xanthine crystal structure is included for comparison. (E) Wild type UapA with docked uracil and 5FU. Both ligands are bound in a mirror image compared to UapA-SYSE and UraA.

More specifically, the calculated pKa of E356 rises from 7.5 in the wild-type (UapA-TFAQ) to 10.0 in the UapA-SYSE mutant, while that of D360 drops from 9.4 to 7.4. Similar pKa values to UapA-SYSE were calculated in UapA-SYSQ. Thus, in both UapA-SYSE and UapA-SYSQ E356 is considered protonated, while D360 is considered negatively charged (given transport assays are performed at pH 6.8). On the other hand, E408 in both UapA-SYSE and UapA-TFAE exhibits a theoretical pKa of 6.8. Alterations in the acidic nature in side

chains of UapA-SYSQ and UapA-SYSE as compared to wild-type (UapA-TFAQ) or UapA-TFAE are expected to modify the polar nature of the substrate binding site, and thus might affect substrate binding and transport. The other important observation related to the binding cavity of these mutants versus the wild-type UapA concerns the formation of a hydrogen bond when F406 is mutated to tyrosine between the phenolic hydroxyl and the V153 backbone carbonyl group (see **Figure 3.4.1A, B**). This hydrogen bond stabilizes the main part of the binding cavity between TMS3 and TMS10, which is known to form the characteristic beta-sheet structure in the middle of the transmembrane domain where the substrate is embedded. This stabilization does not seem to affect transport activity, but apparently contributes to enlarged specificity as UapA-TYAQ recognizes 5FU, hypoxanthine and adenine significantly better than UapA-TFAQ.

In a second step, we used Induced Fit Docking (IFD) calculations in order to elucidate the increased apparent affinity of UapA-SYSE for uracil and 5FU observed. The energetically most favourable binding mode of uracil is remarkably similar to the crystal structure of uracil bound in UraA, its natural transporter, as presented in the left and right panels of **Figure 3.3.1A**. In UapA-SYSE, uracil is well stabilized between the two acidic side chains E408 and E356, forming also  $\pi$ - $\pi$  stacking interactions with F155 and Y406 and an H-bond with S407 side chain. The latter corresponds to a glycine in UraA in this position and to an alanine in UapA, conducting them unable to form this interaction via the residue side chain. A quite similar binding mode was encountered for 5FU, in both UapA-SYSE and UraA in IFD calculations (**Figure 3.3.1B**, left and right panels, respectively). Notice that the F atom is in a hydrophobic cleft in the binding pocket. In this cleft, the C-F bond is parallel to both F155 and Y406 rings, while the orientation of the pyrimidine ring is not affected. MM-PBSA and FEP calculations were also conducted in order to explain the experimental result showing that UapA-SYSE has high capacity and transports 5FU but not uracil. However, according to these calculations, the theoretical binding affinities of both 5FU and uracil are almost the same ( $\Delta G_{5FU} = -16.7$  kcal/mol,  $\Delta G_{URA} = -16.9$  kcal/mol,  $\Delta \Delta G = -0.23$ ). Significantly, in wild-type UapA, which does not seem to transport uracil or 5FU, these ligands are not similarly positioned in the binding pocket, compared to the docking result for UraA or UapA-SYSE, as in their best fit position seem to interact via a mirror-image orientation (**Figure 3.3.1E**). Thus, it becomes evident that uracil/5FU binding mode in UraA or UapA-SYSE, both of which recognize and can transport these ligands, is distinct from that in wild-type UapA, which cannot transport these pyrimidines. However, our docking calculations are not able to explain the reasons that UraA recognizes uracil and 5FU with very high affinity (sub micromolar concentration), despite

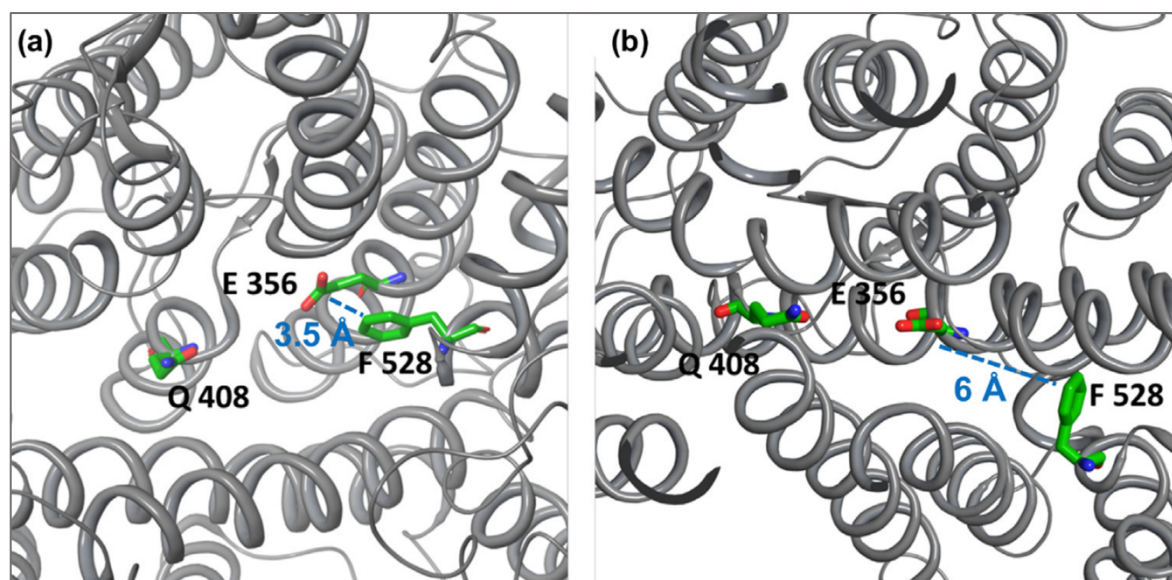
similar binding modes, whereas UapA-SYSE recognizes these two substrates with moderate affinities (148–259  $\mu\text{M}$ , dependent on the combination of different F528 substitutions and SYSE). A profound explanation for this difference is that binding affinities cannot be well predicted by docking approaches, while a probable one is that residues outside the binding site, distinct in UraA and UapA-SYSE, might be essential for substrate recognition.

In an effort to further rationalize the specificity of UapA mutants against hypoxanthine, IFD calculations were performed considering wild-type (UapA-TFAQ) and mutant variants UapA-SYSQ and UapA-TFAE. We need to stress again here that according to functional assays, hypoxanthine does not bind to the wild type (UapA-TFAQ), whereas it binds to UapA-TFAE (with relatively high affinity: 60–70  $\mu\text{M}$ ) but is not transported and is recognized by UapA-SYSQ (with low affinity:  $\sim 1$  mM). In **Figure 3.3.1C**, are presented the lowest energy binding modes of hypoxanthine derived by IFD calculations. It is rather clear that none of them is similar to the xanthine-UapA binding mode where Q408 forms two H-bonds with N1H and C2=O (**Figure 3.3.1D**). In the wild-type UapA, the N9H of hypoxanthine interacts with E356, but not the N1H, while a H-bond is formed between Q408 and C6=O forcing the hypoxanthine scaffold in a entirely different orientation compared to the orientation of the physiological substrates (e.g. xanthine). In UapA-TFAE, the orientation of hypoxanthine enables the two positive charges on N9H and N1H42 to interact with the two negative charges of E408 and E356, respectively, while the C6=O group forms a H-bond with S154. This can explain the relatively high binding affinity of hypoxanthine in this UapA mutant. Nonetheless, this ligand is not transported, possibly due to its topological disorientation in the binding site. Finally, in UapA-SYSQ, E356 is protonated according to pKa calculations. Due to that protonation, it can now interact with N9H and N3 while C6=O interacts with the amide group of Q408. This could explain the binding, albeit low affinity, of hypoxanthine in UapA-SYSQ, as compared to the wild-type UapA in which we encountered fewer interactions for hypoxanthine.

### **3.3.4 F528 dynamically interacts with the substrate binding site**

It is still under investigation in our group, the functional role of all different UapA residues studied herein and the possible structural explanations. Mutations of F528 were isolated several times in this and previous work. To rationalize the key-element role of F528 in affecting specificity either alone or in combination with substitutions in the NAT signature motif, we constructed the model of the outward conformation of UapA by homology modelling using as query the Band3 crystal structure (anion exchanger of human erythrocyte). Band3,

functions as bicarbonate ( $\text{HCO}_3^-$ ) transporter and is the human structural homologue of UapA. Band3 has been crystallized in the outward conformation offering a unique model, for the outward conformation of the category of transporters functioning according to the elevator mechanism. The relative position of F528 to the position of E356 and Q408 in the modelled outward conformation and the inward crystal structure of UapA is depicted in **Figure 3.3.2**. It is clear that, in the outward topology the two residues interact through p-electrons being in very close proximity. The distance of the two side chains, as shown in the 100ns MD simulations, remains stable at approximately 3.50 Å, indicating a quite stable interaction. In contrast, in the inward crystal structure this distance increases to ~6.00 Å, suggesting a dynamic interaction between the substrate binding site and F528, able to explain its critical role in specificity.



**Figure 3.3.2:** UapA outward conformation. (A) The model constructed using Band3 as template is shown mainly focusing on the interaction between F528 and E356. The E356 carboxylate group remains almost parallel to the F528 phenyl ring (3.5 Angstrom) during the 100 ns MD simulation. (B) The corresponding distance in the inward conformation is much higher (pdb structure 5i6c).

### 3.3.5 Conclusions

The NAT signature motif in TMS10 is crucial for substrate binding and specificity and as we showed, it contributed to the evolution of substrate specificity in the NAT family. In this study, we analyzed the NAT motif versions found in several NATs from a functional point of view by conducting a series of virtual molecular docking calculations on the constructed mutant homology models of UapA and by evaluating experimentally the results expressing in a novel genetic system a series of these variant UapA mutants. In UapA-SYSE, uracil is well stabilized

in a docking position, similar to the UraA, not encountered in wild-type UapA, although it is not transported, while according to experimental results, SYSE led to loss of transport capacity for purines but gained a capacity for toxic 5FU accumulation in the context of UapA. We also showed that binding of hypoxanthine in UapA-SYSQ is energetically favored compared to wild-type (UapA-TFAQ) which agrees with the promiscuous profile of SYSQ revealed by growth tests and uptake assays. Finally, computational analysis proposed that F528 interacts dynamically with the substrate binding site and specifically with E356, explaining the role of F528 in specificity regulation indicated by mutational analysis.

### **3.4 NCS2/NAT family transporters (UapA, XanQ, SmLL9, AncXanQ): Structure, function, specificity, mechanism of transportation and crucial amino acids as key factors**

#### **3.4.1 The case of SmLL9 transporter**

*Adapted from “Specificity profile of NAT/NCS2 purine transporters in Sinorhizobium (Ensifer) meliloti “ by Maria Botou, Vassilis Yalelis, Panayiota Lazou, Iliana Zantza, Konstantinos Papakostas, Vassiliki Charalambous, Emmanuel Mikros, Emmanouil Fletmetakis, Stathis Frillingos, Molecular Microbiology (March 2020)*

##### **3.4.1.1 Foundation of the study**

SmLL9 is a transporter specific for xanthine and uric acid of the rhizobial species, *Sinorhizobium (Ensifer) meliloti* which belongs to the NAT/NCS2 family and the NAT/COG2233 Cluster 1<sup>179</sup> including also two thoroughly studied homologs, the xanthine-specific XanQ of *E. coli*<sup>113,188</sup> and the crystallographically resolved xanthine/uric acid transporter UapA from *Aspergillus nidulans*<sup>58,76</sup>. Interestingly, all homologs of *S. meliloti* cluster along with well-known purine nucleobase transporters, NAT/NCS2 homologs, from *Escherichia coli* or other proteobacteria<sup>113,188–190</sup>. However, most other rhizobia contain only homologs of the NCS1 family exhibiting very low similarity to the known purine transporters or lack NAT/NCS2 homologs. SmLL9 is the first bacterial homolog that transports with high affinity both xanthine and uric acid, among other functionally known homologs. The XanQ and UacT from *E. coli*, previously well studied, are specific for xanthine and uric acid respectively. The sequence homology of SmLL9 and XanQ is very limited. In detail, it they

exhibit 29.8% identity based on MUSCLE alignment and 26.2% identity, 43.9% similarity, 18.3% gaps based on the EMBOSS Needle tool. Still, both transporters preserve the functionally crucial residues in XanQ<sup>188</sup> (Table 3.4.1.1). These observations prompted us to use molecular dynamics to study the structure-function relationships of SmLL9 in relation to XanQ and to experimentally confirm the results with site-directed mutagenesis at important amino acid positions.

### 3.4.1.2 SmLL9 purine nucleobase transporter of *S. meliloti* has enlarged specificity profile

As a first step, the Prof. Frillingos' team, analyzed the transport activities of the NAT/NCS2 transporters of *S. meliloti* strain 1,021 for several putative substrates namely adenine, guanine, hypoxanthine, xanthine, uric acid, uracil, thymine, and cytosine. It was found that none of them was able to transport any pyrimidine nucleobase, but only purines. In detail, homologs belonging to NAT/COG2233, are high-affinity transporters of xanthine and/or uric acid (except SmWQ0 which seems unable to transport any nucleobase), and homologs that belong to AzgA-like/COG2252 are high-affinity transporters of adenine and/or guanine/hypoxanthine. Experiments in the Frillingos' lab have shown that SmLL9 is a high-affinity transporter for xanthine ( $K_M = 3.3 \mu\text{M}$ ), while a lower-affinity transporter for uric acid ( $K_M = 27.3 \mu\text{M}$ ). The protonophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) abolishes the uptake activity of xanthine in SmLL9, indicating a dependency on H<sup>+</sup> symport.

Firstly, and to give an insight into the design of the computational experiments, a range of purine analogs as putative competitors were used to compare SmLL9, that transports multiple purine substrates, to the xanthine-specific XanQ<sup>191</sup> for competition against xanthine transport (known appropriate *E. coli* transporters were used as controls). SmLL9 recognizes with high-affinity uric acid, as well as 8-methylxanthine, and 7-methylxanthine which are not ligands for XanQ. This highlights that SmLL9 displays broader specificity toward analogs with diverse substitutions on the imidazole of the purine substrate. These findings were subjected to further investigation through computational calculations.

### 3.4.1.3 Homology modeling of SmLL9 and comparison with XanQ

As encountered above, SmLL9 is the first bacterial homolog shown to transport with high affinity both xanthine and uric acid, as the previously well characterized homologs from *E. coli* XanQ and UacT are specific for xanthine or uric acid, respectively. It is worth

noting that SmLL9 is classified in a different clade compared to the one including XanQ and UapA<sup>188</sup>. However, XanQ and SmLL9 share almost 30% identity and 43.9% similarity and most of the crucial residues with respect to function in XanQ (based on Cys-scanning analysis<sup>188</sup>) are retained in SmLL9 (**Table 3.4.1.1**).

**Table 3.4.1.1:** Amino acid differences of SmLL9 at residues that are functionally important in XanQ. The information on the functional roles of amino acid residues of XanQ is based on the Cys-scanning analysis data published previously (references included in the table).

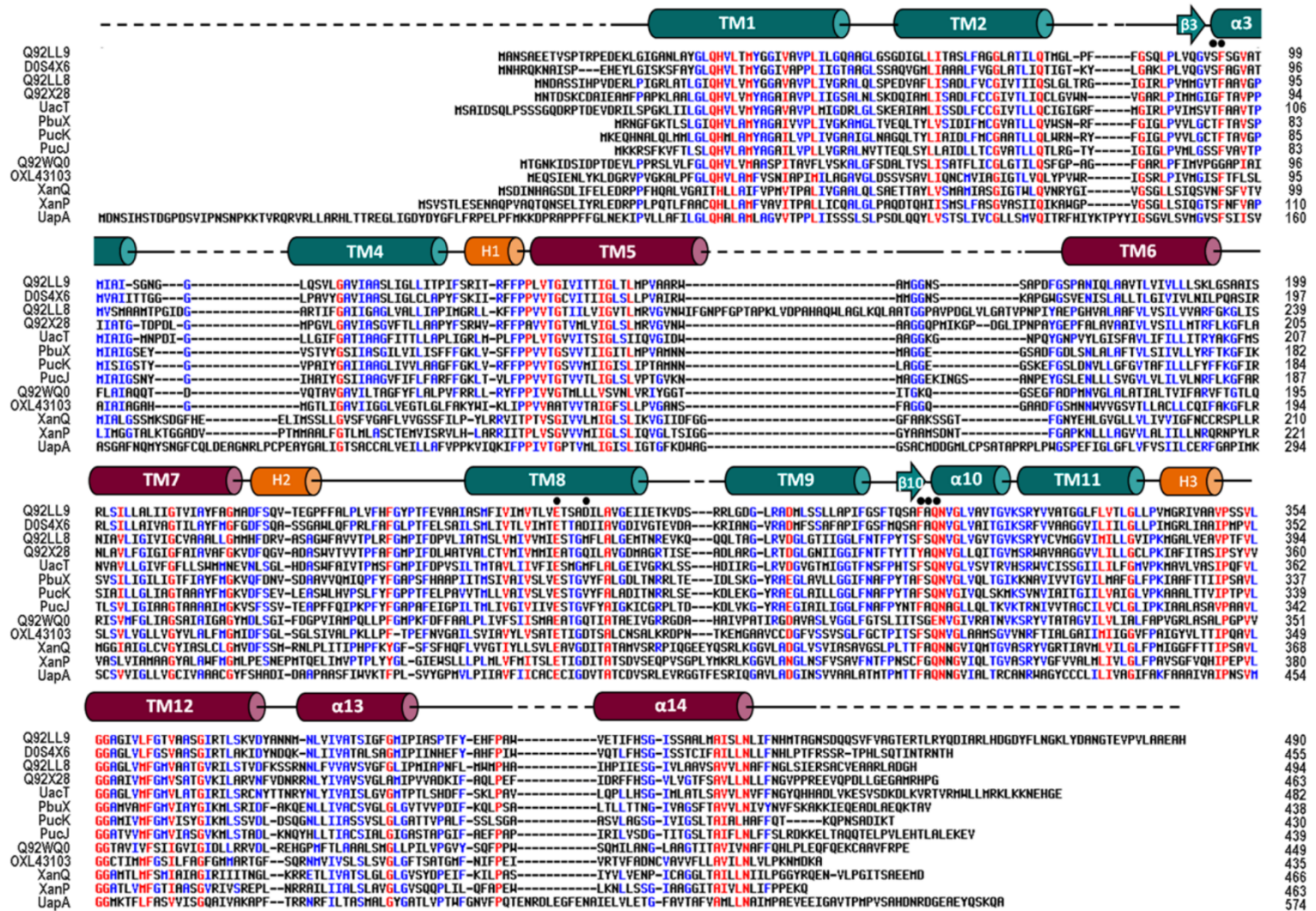
Residue (SmLL9)	Residue (XanQ)	Importance of residue in XanQ (key points)	Reference (XanQ)
His31	His31	H-bonding at the periphery of the binding site important for high affinity binding of xanthine (H31-N325)	Karena & Frillingos, 2009
Gly68	Ser68	H-bonding affects the substrate-binding residue Phe94 at the xanthine binding site (S68-S91)	Karena <i>et al.</i> , 2015
Gly69	Gly69	Replacement by A retains high affinity for xanthine but with broader ligand profile	Karena <i>et al.</i> , 2015
Leu70	Ile70	Replacement by C retains high affinity for xanthine but with broader ligand profile	Karena <i>et al.</i> , 2015
Ala71	Gly71	Irreplaceable for function (by A or C)	Karena <i>et al.</i> , 2015
Gln75	Gln75	Crucial in a H-bond network that is essential for stability and function in the core domain; irreplaceable individually but replaceable when retaining properties and geometry of a central H-bond pair (Q75-D304)	Karena <i>et al.</i> , 2015
Gly83	Gly83	Crucial for folding/insertion in the membrane	Karena & Frillingos, 2011
Ser84	Ser84	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Karena & Frillingos, 2011
Leu86	Leu86	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Karena & Frillingos, 2011
Pro87	Leu87	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Karena & Frillingos, 2011
Leu88	Ser88	Confers to a H-bonding network in core domain that is essential for stability and function	Karena <i>et al.</i> , 2015
Gln90	Gln90	Confers to a H-bonding network in core domain that is essential for stability and function	Karena <i>et al.</i> , 2015
Ser93	Asn93	Part of a H-bond network at the binding site; crucial for the xanthine specificity of XanQ	Karena & Frillingos, 2009 Karena & Frillingos, 2011
Phe94	Phe94	Confers to high affinity binding of xanthine; its backbone amide N forms H-bond with xanthine	Karena & Frillingos, 2011
Thr142	Ser151	Replacement by A, T or C retains high affinity for xanthine but with broader ligand profile	Karena <i>et al.</i> , 2015
Gly143	Gly152	Replacement by A retains high affinity for xanthine but with broader ligand profile	Karena <i>et al.</i> , 2015
Ile146	Val155	Replacement by Ile impairs activity; V155M leads to altered specificity profile	Karena <i>et al.</i> , 2015
Gly150	Gly159	Confers flexibility at the interface between the gate and core domain; irreplaceable for function	Karena <i>et al.</i> , 2015
Gly212	Gly221	Crucial for folding/stability in the membrane	Karena <i>et al.</i> , 2015
Ser251	Val261	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Mermelekas <i>et al.</i> , 2010
Glu262	Glu272	Xanthine-binding residue; irreplaceable	Karena & Frillingos, 2009
Ala265	Gly275	Sensitive to inactivation by <i>N</i> -ethylmaleimide.	Mermelekas <i>et al.</i> , 2010



		Xanthine binding increases accessibility of a cysteine (G275C) to SH-alkylating reagents	
Asp266	Asp276	Carboxyl group-276 is needed in the transport mechanism. Irreplaceable by Q, H, M, V, I, G, C, A; replaceable with low activity by N, T. D266E is highly active but affects pH-activity dependence and substrate specificity. Contributes to a H-bond network at the binding site that might be important for proton coupling and/or specificity.	Mermelekas <i>et al.</i> , 2010 Karena & Frillingos, 2011
Ala269	Ala279	Sensitive to inactivation and/or disruption of folding/stability in the membrane by bulky replacement (C, N, V) but not by T, S or G.	Mermelekas <i>et al.</i> , 2010
Val270	Thr280	Sensitive to inactivation and/or disruption of folding/stability in the membrane by bulky replacement (C, N, V, Q) but not by A, S or G.	Mermelekas <i>et al.</i> , 2010
Ile274	Ser284	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Mermelekas <i>et al.</i> , 2010
Asp290	Asp304	Crucial in a H-bond network that is essential for stability and function in the core domain; irreplaceable individually but replaceable when retaining properties and geometry of a central H-bond pair (Q75-D304)	Karena & Frillingos, 2009 Karena <i>et al.</i> , 2015
Met291	Gly305	Confers to folding/stability in the membrane	Mermelekas <i>et al.</i> , 2010
Thr304	Pro318	Crucial for folding/insertion in the membrane	Karatza <i>et al.</i> , 2006
Ser306	Thr320	Contributes to a H-bond network at the binding site; activity impaired with Cys-replacement	Karatza <i>et al.</i> , 2006 Karena <i>et al.</i> , 2015
Ala307	Thr321	Contributes to H-bond network that is essential for stability and function in the core domain; activity impaired with Cys-replacement	Karatza <i>et al.</i> , 2006 Karena <i>et al.</i> , 2015
Phe308	Phe322	Contributes to xanthine binding by aromatic stacking with substrate; sensitive to inhibition by <i>N</i> -ethylmaleimide	Karatza <i>et al.</i> , 2006 Karena <i>et al.</i> , 2015
Ala309	Ala323	Sensitive to inactivation by <i>N</i> -ethylmaleimide; backbone amide N forms H-bond with xanthine Xanthine binding increases accessibility of a cysteine (A323C) to SH-alkylating reagents	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Gln310	Gln324	Xanthine-binding residue; irreplaceable	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Asn311	Asn325	Irreplaceable; crucial H-bonding interaction at the periphery of the binding site (H31-N325). Xanthine binding hinders accessibility of a cysteine (N325C) to SH-alkylating reagents	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Val312	Asn326	Sensitive to inactivation by <i>N</i> -ethylmaleimide; accessible to membrane-impermeable MTSES <sup>-</sup>	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Gly313	Gly327	Sensitive to inactivation by <i>N</i> -ethylmaleimide; accessible to membrane-impermeable MTSES <sup>-</sup>	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006

Leu314	Val328	Sensitive to inactivation by <i>N</i> -ethylmaleimide; accessible to membrane-impermeable MTSES <sup>-</sup>	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Val315	Ile329	Sensitive to inactivation by <i>N</i> -ethylmaleimide	Karatza <i>et al.</i> , 2006
Thr318	Thr332	Sensitive to inactivation by <i>N</i> -ethylmaleimide; T332N alters specificity for xanthine analogs	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Gly319	Gly333	Sensitive to inactivation by <i>N</i> -ethylmaleimide; accessible to membrane-impermeable MTSES <sup>-</sup> . G333R alters specificity for xanthine analogs	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Ser322	Ser336	Sensitive to inactivation by <i>N</i> -ethylmaleimide; accessible to membrane-impermeable MTSES <sup>-</sup>	Georgopoulou <i>et al.</i> , 2010 Karatza <i>et al.</i> , 2006
Pro342	Pro354	Important for folding/stability in the membrane	Mermelekas <i>et al.</i> , 2010
Gly355	Gly369	Confers flexibility at the interface between the gate and core domain; irreplaceable for function	Karena <i>et al.</i> , 2015
Gly356	Gly370	Confers flexibility at the interface between the gate and core domain; irreplaceable for function	Karena <i>et al.</i> , 2015
Phe362	Phe376	Contributes to xanthine coordination at the inward-open conformation; activity diminished in the absence of the phenyl group	Karena <i>et al.</i> , 2015
Gly363	Ser377	S377T leads to inactivation; S377C retains high affinity for xanthine with broader ligand profile	Karena <i>et al.</i> , 2015
Gly369	Gly383	Confers flexibility at the interface between the gate and core domain; irreplaceable for function	Karena <i>et al.</i> , 2015
Ile391	Leu404	Irreplaceable for function (by A or I); L404C retains low transport activity that is restored to wild-type levels by <i>N</i> -ethylmaleimide adduct	Karena <i>et al.</i> , 2015
Pro396	Gly409	Irreplaceable for function in the gate domain	Karena <i>et al.</i> , 2015
Ser417	Asn430	Sensitive to inactivation by <i>N</i> -ethylmaleimide; 2-thioxanthine protects N430C from alkylation. Irreplaceable by Thr but replaceable by Ser.	Papakostas <i>et al.</i> , 2008 Papakostas & Frillingos, 2012
Ile419	Ile432	Sensitive to inactivation by bulky replacements (F, W, L, M) but not by Q, N or A, S, T, V.	Papakostas <i>et al.</i> , 2008

SmLL9 was used as an input in HHpred search tool for homology detection and structure prediction, which ranks UapA<sup>58</sup> (PDB ID: **5I6C**) first with 100% probability ( $p = 2.5e-40$ ), 24% identity, and 0.38 similarity. The same holds for XanQ with 100% probability ( $p = 6e-43$ ), 30% identity, and 0.54 similarity. Both transporters also exhibit homology to UraA<sup>57</sup> (PDB ID: **5XLS**) with 100% probability ( $p = 8.9e-40$ ), 24% identity and 0.42 similarity for SmLL9, and 100% probability ( $p = 2e-37$ ), 24% identity and 0.38 similarity for XanQ. The sequence alignment of SmLL9 and other known NAT/NCS2 homologs is depicted in **Figure 3.4.1.1**.

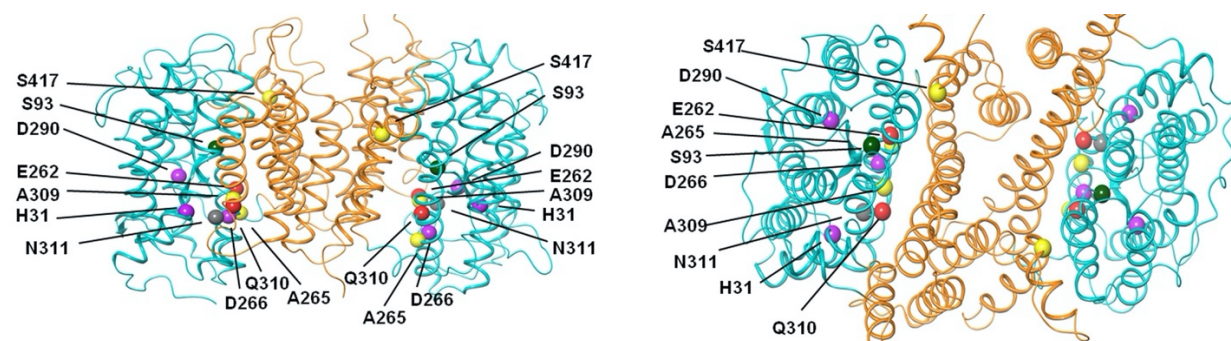


**Figure 3.4.1.1:** Sequence alignment of SmLL9 and other known NAT/NCS2 homologs in Cluster 1 (C1\_Xanthine-Uric Acid). The alignment shown was prepared with MultAlin (Corpet, 1988). High consensus sites (occurring at >90% of the sequences) are shown in red and low consensus sites (>50%) in blue. The approximate positions of alpha-helices and beta-sheets shown on top of the sequences and the xanthine-binding residues (indicated as small dots) are according to the UapA structure PDB no. 516C (Alguet et al., 2016). The alignment includes the relevant homologs of *S. meliloti* analyzed in this study (Q92LL9, Q92LL8, Q92X28, Q92WQ0), two homologs (from *Acinetobacter calcoaceticus*, DOS4X6, and *Prevotella copri*, OXL43103) discovered and annotated as xanthine transporters in a riboswitch-based functional mining study of a gut metagenomic library (Genee et al., 2016), and the well-known xanthine and/or uric acid-transporting homologs UacT, XanQ and XanP from *E. coli* K-12 (see text and Table S3), PbuX, PucK and PucJ from *Bacillus subtilis* (Schultz et al., 2001; Saxild et al., 2001), and UapA from *Aspergillus nidulans* (see text). The UniProt accession numbers not given in the figure are: Q46821 (UacT), P42086 (PbuX), O32140 (PucK), O32139 (PucJ), A0A22912P1 (PXL43103), P67444 (XanQ), P0AGM9 (XanP), Q07307 (UapA).

Structural models of both SmLL9 and XanQ were constructed based on homology modeling and specific protein-substrate interactions were monitored. The structural models were generated as monomers based on homology to the UapA crystal structure of *Aspergillus nidulans* which is in the Inward Open conformation. In order to construct the dimeric form of

both models, a 3D superposition on the UapA transporter dimer was performed (**Figure 3.4.1.2**). Each structure was subjected to 40 ns of molecular dynamics simulation in the *apo* form using GROMACS software<sup>134</sup> and charmm36 force field<sup>135</sup>. The protein was embedded in DPPC lipid bilayer which was solvated in explicit water. The RMSD of the C $\alpha$ -carbons of the  $\alpha$ -helices did not exceed 1 Å, affirming the stability of the theoretical model.

According to the UapA template the 3D models of SmLL9 and XanQ are dimeric and each protomer consists of 14 transmembrane segments (TMS) (Figure 3.2.1.2). A core domain of (TMS 1–4 and 8–11) and a gate domain (TMS 5–7 and 12–14) combined form the protomer. The transmembrane helices have the same length except TMS9 which is longer for XanQ by one turn (4 amino acids). The loops connecting the TMS are also of the same length on average except the loop between TMS3–4 in XanQ which is longer by 10 amino acids, and the loop between TMS4–5 which is shorter by three amino acids. Three exoplasmic loops of both bacterial homologs are shorter compared to UapA, by at least 10 amino acids in TMS3–4, 15 amino acids in TMS5–6, and 12 amino acids in TMS13–14.

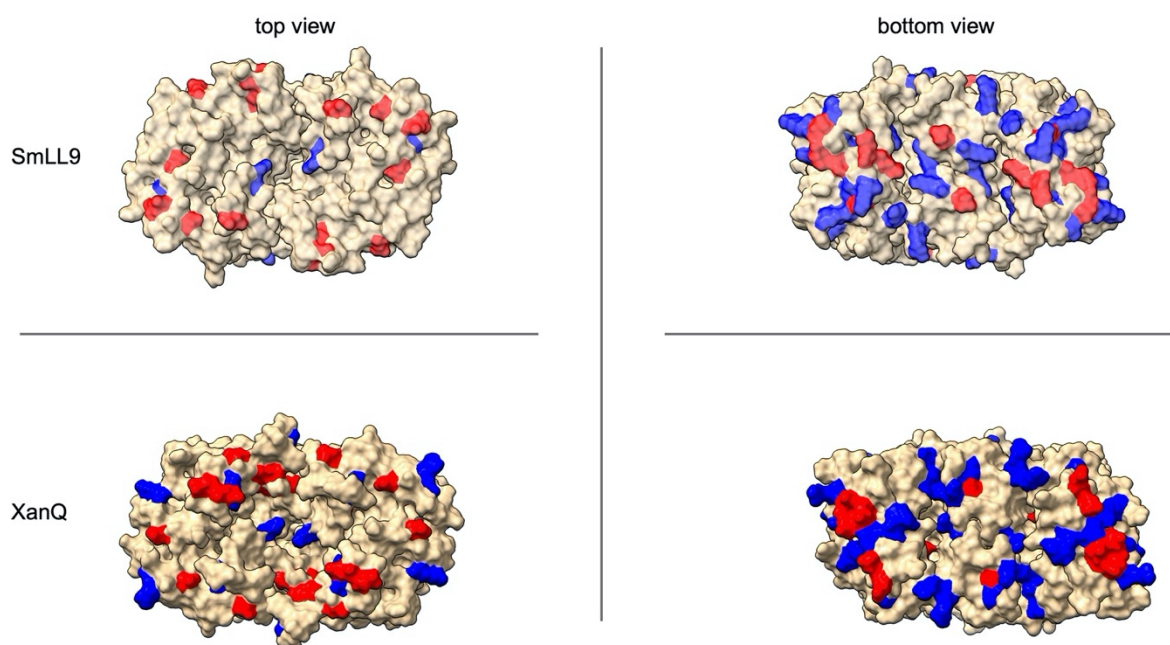


**Figure 3.4.1.2:** Homology modeling and site-directed mutagenesis of SmLL9. (a) Dimeric structural model of SmLL9 as viewed parallel to the membrane plane (*left*) and from the cytoplasmic side (*right*). Different colors highlight the core and the gate domain of each protomer. Residues found to be important with site-directed mutagenesis are shown as spheres in red (irreplaceable, involved in substrate binding), gray (irreplaceable, involved in H-bond interactions in the vicinity of the binding site), purple (H-bond interactions in the core domain), green (crucial for specificity), and yellow (sensitive to inactivation by site-directed alkylation).

Remarkably, in UapA there is a pair of cysteines shown to be rather important for UapA and possibly other fungal homologs, as it forms a disulfide bond that is needed for intracellular folding in the loop between TMS3–4 which does not exist in these bacterial homologs<sup>58</sup>, while they also lack the 10-residue segment between these cysteines. Concerning the cytoplasmic loops, the loop between TMS4–5 in SmLL9 is shorter by one residue relative to UapA resulting in the distortion of the small helix between TMS4 and TMS5 found in the UapA crystal structure.



The distribution of the ionized residues in both SmLL9 and XanQ models is reasonable (**Figure 3.4.1.3**), as they are allocated on the protein surface, most of them at the periplasmic and cytoplasmic sides or alongside the pore of the protein. Positive charges face mostly the cytoplasm while negative charges the periplasm.



**Figure 3.4.1.3:** The distribution of the ionized residues in SmLL9 and XanQ models at top and bottom view.

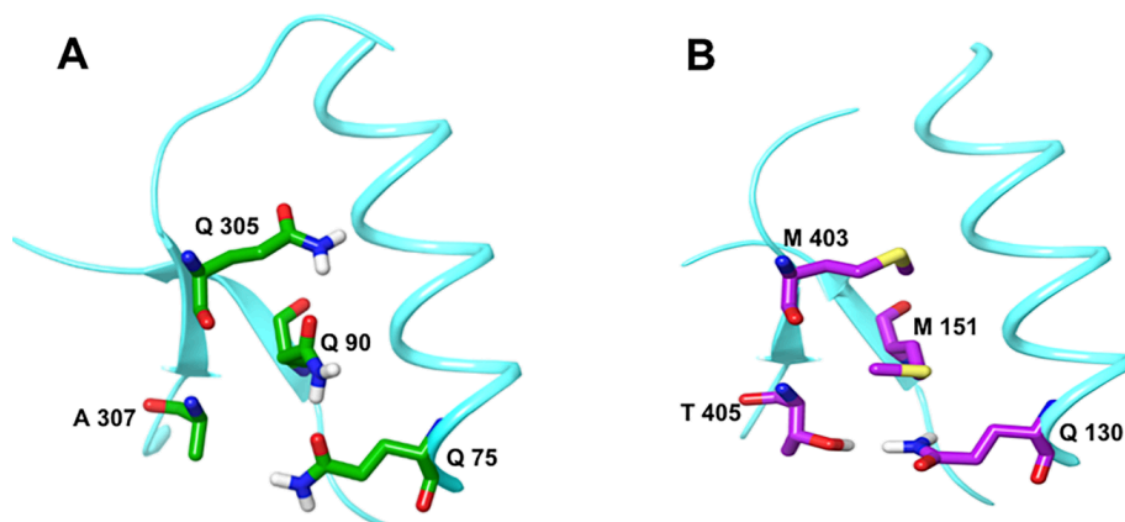
The pKas of the carboxylates in the binding cavity were calculated for both models using PDB2PQR version 2.0.0 server<sup>192</sup>. D266 in SmLL9 and the corresponding D276 in XanQ had pKa values higher than 8 and they were considered as protonated in the docking calculations. In the case of XanQ the pKa of E272 was 7.5 and thus both the protonated and the charged carboxylate form were used in the docking calculations which were performed twice. On the other hand, the corresponding to E272 of XanQ in SmLL9, E262, had a pKa value equal to 7 and was considered negatively charged in all calculations.

High similarity between SmLL9, XanQ, UapA, and UraA is observed concerning residues crucial for the binding and transport of substrate, specificity, kinetics, and transporter function in UapA and UraA that are located in the vicinity of substrate (4.5 Å) in TMS 1, 3, 8, and 10, after 3D alignment of the constructed models and the UapA and UraA crystal structures (**Table 3.4.1.2**).

**Table 3.4.1.2:** Residues existing in the substrate binding cavity of Q92LL9, XanQ, UapA and UraA transporters in the vicinity of the crystallographic ligand, aligned after superimposition of 3D homology models (SmLL9, XanQ) and crystal structures (UapA, UraA).

SmLL9	XanQ	UapA	UraA
Q310	Q324	Q408	E290
E262	E272	E356	E241
D266	D276	D360	H245
S93	N93	S154	S72
F308	F322	F406	Y288
F94	F94	F155	F73
L361	M375	L459	L341
V42	P42	P97	V35
S306	T320	T404	T286
A307	T321	T405	T287
M35	I35	M90	M28
V365	I379	V463	I345

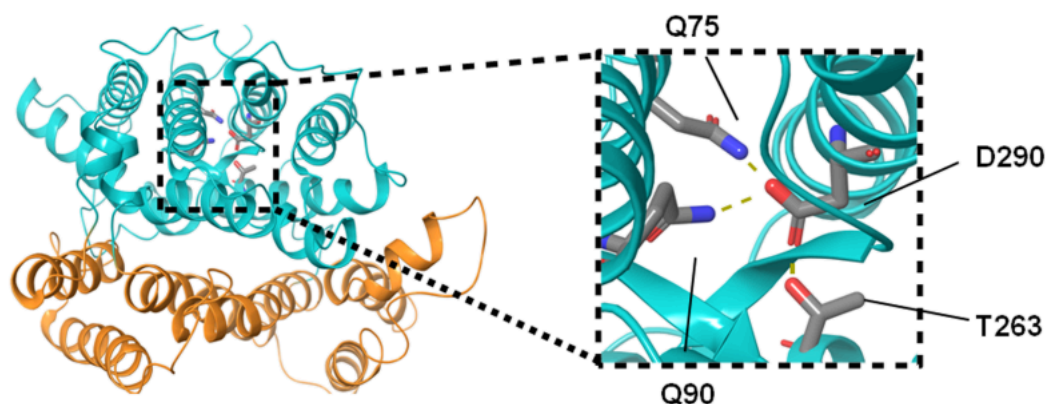
In SmLL9 core domain, a hydrogen bond network formed connects TMS 2, 3, 8, 9, and 10, similar to the one described for XanQ. The network consists of the side chains of Q90, Q75, Q305, D290, and T263 and is divided in two parts, one between Q90-Q75-Q305 (**Figure 3.4.1.4A**) and one between D290, T263, and the backbone of Q305.



**Figure 3.4.1.4:** The Q305-Q90-Q75 network in SmLL9 and its counterpart in UapA. (A) In SmLL9, Q75 (TMS2), Q90 (TMS3) and Q305 (TMS10) are aligned along the membrane axis forming a H-bond network of interactions between TMSs 2, 3 and 10. (B) In UapA, the corresponding interactions are formed between Q130 (TMS2) and T405 (TMS10) as well as between M151 (TMS2) and M403 (TMS10).

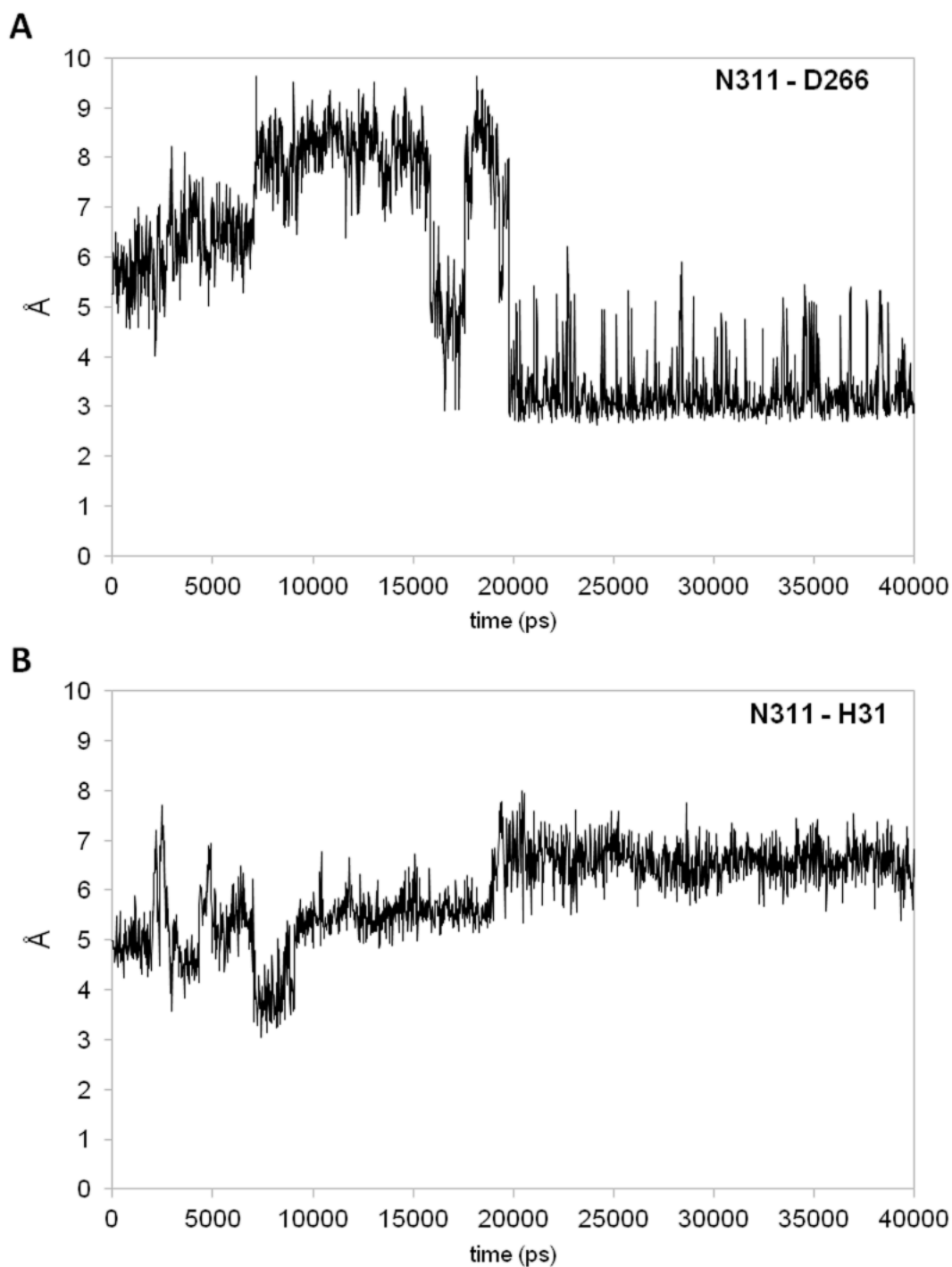
During the MD simulations the D290 side chain forms strong H-bonds with Q75 and Q90 being oriented toward the center of the core domain preserving the interaction with T263

(**Figure 3.4.1.5**). The residues corresponding to Q90 and Q305 in UapA<sup>58</sup> are M151 and M403 which seem to form stabilizing hydrophobic interactions as the two S-CH<sub>3</sub> moieties are almost parallel and about 4.6 Å apart.



**Figure 3.4.1.5:** The H-bond network stabilizing the core domain in SmLL9. Shown in the left is the extracellular view of the modeled structure with core domain in teal and gate domain in orange. The inset in the right highlights the H-bond network present during the major part of the MD simulations where the D290 side chain reorients pointing to the central cavity of the core domain and forming strong H-bonds with Q75, Q90 and T263 connecting TMS8 and TMS9 with TMS2 and TMS3.

In addition, there is a weak H-bond interaction between Q130 and T405 which is absent in SmLL9 because the corresponding residues are Q75 and A307 respectively (**Figure 3.4.1.4B**). Thus, the H-bond network of the core domain of the bacterial homologs seems to include different types of interactions which are stronger than the ones in UapA. Moreover, N311 side chain in SmLL9 exhibits a relative flexibility in MD simulations, as it frequently changes orientation forming interactions with H31 on the one side and D266 on the other side (**Figure 3.4.1.6**).



**Figure 3.4.1.6:** Distances between A) residues N311 and D266, B) residues N311 and H31 of SmLL9. Distances have been monitored throughout the 40 ns MD simulation between A) the amide nitrogen N311 (ND2) and D266 carboxylate (OD1) and B) the amide oxygen N311 (OD1) and imidazole nitrogen (NE2) H31.

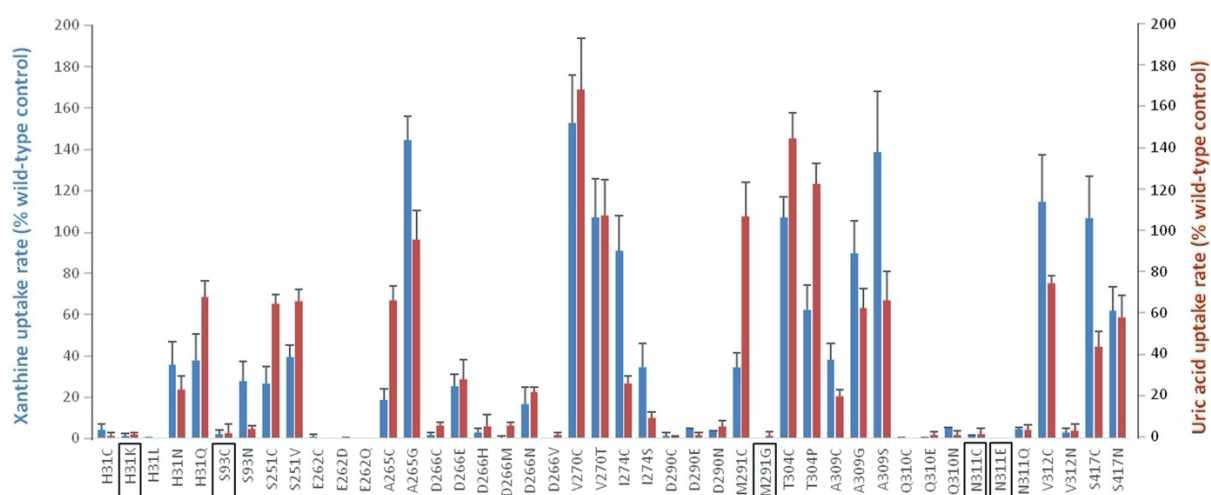
Concerning the binding of the substrates in the binding site, induced fit docking (IFD) calculations indicated that SmLL9 and XanQ bind xanthine in a similar manner, in contrast to uric acid which is bound in XanQ in a reverse orientation and less optimally compared to SmLL9. The results from the IFD calculations are presented later on.



### 3.4.1.5 Key binding-site residues are conserved between SmLL9 and XanQ preserving their functional role

Aiming to address the role of key amino acid residues of SmLL9 and understand the way it functions as well as its broader specificity compared to XanQ, we performed site-directed mutagenesis on SmLL9 at positions corresponding to important residues of XanQ (Table 3.4.1.1). These residues were either involved in substrate binding or participated in H-bond interactions at the periphery of the binding site or belonged to the crucial in the coordination of substrate binding TMSs 8 and 10 (Figure 3.4.1.2). Residues of SmLL9 that are conserved in XanQ were replaced with amino acids with similar side-chain properties. In the case that the similar side-chain mutants retained significant activity, they were subjected to more extensive mutagenesis. Residues differing from XanQ were substituted by the corresponding amino acid occupying the same position in XanQ. In addition, taking advantage of the fact that SmLL9 does not contain any native Cys in its sequence, all the residues examined were mutagenized also to Cys.

Mutagenesis data reveal that most SmLL9 mutants display remarkably similar properties with the corresponding XanQ mutants studied previously (see Table 3.4.1.1). The substrate-binding residues E262 (TMS8) and Q310 (TMS10) are strictly irreplaceable concerning function, since mutants E262C, E262D, E262Q, Q310C, Q310E, Q310N are inactive for either xanthine or uric acid transport (Figure 3.4.1.7).



**Figure 3.4.1.7:** Site-directed mutagenesis of SmLL9. *Escherichia coli* T184 expressing the corresponding mutants from pT7-5/SmLL9-BAD were assayed for uptake of [<sup>3</sup>H]-xanthine (1 μM) (blue bars) or [<sup>14</sup>C]-uric acid (4 μM) (red bars) and the deduced transport rates were expressed as percentages of the rate obtained with wild-type SmLL9. Values obtained with cells harboring pT7-5/-BAD alone had been subtracted from the sample measurements in all cases. Each bar represents the mean of at least three determinations with SD shown. Mutants shown in boxes attain negligible protein levels in the *E. coli* membrane.

The same has been shown for the corresponding XanQ mutants at E272<sup>193</sup> and Q324<sup>184</sup>, differing only because Q324N and Q324E retain some activity but are grossly impaired in the affinity for xanthine or xanthine analogs<sup>184</sup>. The residues D290 (TMS9), which is central in a H-bond network stabilizing the core domain in XanQ<sup>188</sup>, and N311 (TMS10), which is involved in interactions with H31 and D266 at the periphery of the binding site are also irreplaceable, as mutants D290C, D290E, D290N, N311C, N311E, N311Q are inactive (**Figure 3.4.1.7**). The same has been shown for the XanQ mutants corresponding to D304<sup>193</sup> and N325<sup>184</sup>. Interestingly, N311C and N311E are not detected in the *E. coli* membrane, pointing to a more austere role of N311 in affecting the protein stability/folding relative to N325 in XanQ. The H-bonding Q or N can replace H31 (TMS1) residues, but C, L, or K cannot (**Figure 3.4.1.7**). Mutant H31K is not detected in the membrane. Similar data have been obtained with the corresponding XanQ mutants, with the difference that H31C and H31L in XanQ display low transport activity but impaired affinity for xanthine or xanthine analogs<sup>193</sup>. D266 (TMS8) is replaceable only by E or N and irreplaceable by C, H, M, or V (Figure 3.2.1.7), in close resemblance with the properties of D276 in XanQ<sup>194</sup>. As a difference arises that D276E in XanQ displays high activity, in contrast to D266E in SmLL9 which transports at a rate of 20% relative to wild type (**Figure 3.4.1.7**) pointing to a more crucial role of this carboxyl group in SmLL9. In the uric acid transporter UacT of the NAT/COG2233 Cluster C1\_Xanthine-Uric Acid<sup>195</sup>, all residues corresponding to the above six crucial residues of SmLL9 were found to be irreplaceable<sup>189</sup>. Of the corresponding residues in the fungal UapA, the two substrate-binding residues E356 (TMS8)<sup>1</sup> and Q408 (TMS10)<sup>127</sup> are also essential and D388 (TMS9) is also functionally irreplaceable<sup>1</sup>, but the other residues are less stringently related with activity: N409 (TMS10) is replaceable by Ala, Ser or Gln with retention of wild-type properties, and only N409D results in loss of function<sup>127</sup>, H86 (TMS1) is linked with defects in folding and targeting to the plasma membrane as shown with the low-activity mutants H86A, H86K, or H86D, while H86N is active and indistinguishable in function from wild type<sup>196</sup>, and D360 (TMS8) is replaceable by Ala retaining wild-type properties<sup>126</sup>.

S93 (TMS3) is irreplaceable by C while S93C is not detected in the membrane, whereas S93N retains the significant xanthine uptake activity but marginal transport activity for uric acid (**Figure 3.4.1.7**). Kinetic analysis shows that S93N transports uric acid with low affinity (3-fold higher  $K_M$  relative to wild type) and low efficiency (10-fold lower  $V_{max}/K_M$  for uric acid than for xanthine). In the xanthine-specific XanQ, the corresponding Asn93 is also irreplaceable by C with low levels of the N93C mutant in the membrane<sup>197</sup>, whereas N93S and

N93A display high transport activity for xanthine and marginally detectable transport activity for uric acid as well<sup>197</sup>. In the related uric acid transporter UacT, replacement of the corresponding T100 by Ala results in significant ability to transport xanthine in addition to uric acid<sup>189</sup>. In the fungal UapA, replacements of the corresponding S154 with A or N lead to defects in transport activity and kinetic changes relative to wild type indicative of a role of S154 in specificity. The affinity for xanthine in S154N, shows a 5-fold improvement without affecting the affinity for uric acid. The affinity for uric acid for S154A shows 7-fold decrease without affecting the affinity for xanthine. These mutants are defective in transport of both substrates but their relative transport efficiencies for xanthine and uric acid have not been tested<sup>198</sup>.

Among the other residues, high transport activities for both uric acid and xanthine (**Figure 3.4.1.7**) and close to wild-type kinetics is displayed by most mutants, with few exceptions mentioned below. Met291 (TMS9) is irreplaceable by G and M291G is not detected in the membrane, but M291C is highly active (Figure 3.2.1.7), albeit with significantly lower affinities (higher  $K_M$ ) relative to wild type for both uric acid and xanthine; in XanQ, G305 at the corresponding position has been annotated as important for folding/stability and mutant G305C is not detectable in the membrane<sup>194</sup>. T304 (TMS10) is replaceable by C or P with retention of high transport activity for both xanthine and uric acid (**Figure 3.4.1.7**), although T304P displays significantly lower affinities (higher  $K_M$ ) relative to wild type; in XanQ, Pro318 at the corresponding position is crucial for stability and detection of the protein in the membrane<sup>184</sup>. V312 (TMS10) is replaceable by C with high activity and small differences from wild type in kinetics but irreplaceable by N (**Figure 3.4.1.7**). The corresponding N326 in XanQ is replaceable by C but sensitive to inactivation by alkylation with *N*-ethylmaleimide<sup>184,199</sup>, and the corresponding V320 in the uric acid transporting homolog UacT is irreplaceable by N<sup>189</sup>.

The mutants with Cys replacement in the natively C-less SmLL9 background allow the examination of the effects of alkylation of these mutants by sulfhydryl-specific reagents on the transport activity. In case an alkylation by the small and relatively hydrophobic *N*-ethylmaleimide inhibits an active Cys-mutant, it would be indicative of involvement of the corresponding position in the substrate translocation pathway. Thus, if the presence of the substrate changes this inhibitory effect, it would indicate a position sensitive to the conformational alternations of the transporter or it would show direct involvement in substrate binding<sup>112,200</sup>. Of the nine active Cys-mutants of SmLL9, V270C, I274C, M291C, and T304C are not inhibited by *N*-ethylmaleimide, S251C is inhibited with  $IC_{50} = 169 \mu M$ , V312C with  $IC_{50} = 47 \mu M$ , and three mutants are inhibited with  $IC_{50} = 14\text{--}20 \mu M$  which remains unchanged (A309C, S417C) or increases 3-fold (A265C) in the presence of 1 mM substrate (xanthine).

The corresponding alkylation-sensitive mutants of XanQ display very similar responses to *N*-ethylmaleimide; V261C is the least sensitive ( $IC_{50} = 97 \mu\text{M}$ )<sup>194</sup>, N326C is inhibited with  $IC_{50} = 50 \mu\text{M}$ <sup>199</sup>, A323C with  $IC_{50} = 34 \mu\text{M}$ <sup>199</sup>, N430C with  $IC_{50} = 10 \mu\text{M}$ <sup>201</sup>, G275C with  $IC_{50} = 15 \mu\text{M}$ <sup>194</sup>. A thorough comparison of the effects of the *N*-ethylmaleimide of mutants A265C (TMS8) and A309C (TMS10) on the xanthine transport activity with the corresponding effects in XanQ, illustrates that the *N*-ethylmaleimide sensitivity is, for all mutants, in the same  $IC_{50}$  range but the SmLL9 mutants respond differently to the binding of xanthine. For instance, the sensitivity of A265C decreases, whereas there is an increase in the sensitivity of G275C in XanQ; the sensitivity of A309C remains the same, while the sensitivity of A323C in XanQ increases.

#### 3.4.1.6 Mutation of S93 to Asn converts SmLL9 to a xanthine-selective transporter

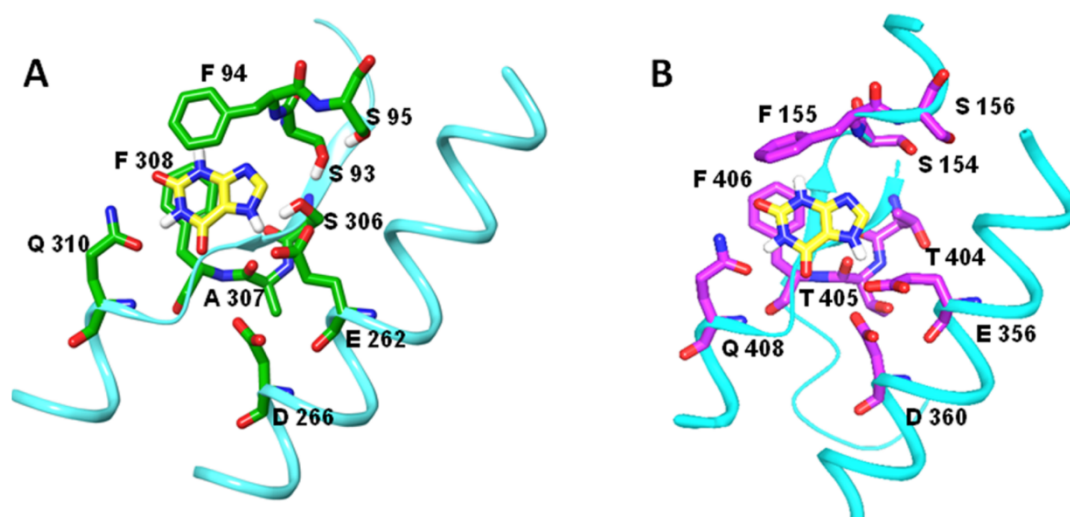
As derived by the transport kinetics results, there are significant differences in affinity ( $K_M$  value) or efficiency ( $V_{\max}/K_M$  value) in several mutants compared to the wild type for either xanthine or uric acid uptake or both. Notably, S93N (XanQ has N in the corresponding position) mutation transports the uric acid with 10-fold lower efficiency relative to xanthine. Compromised efficiencies for uric acid were detected for A265G, A309G, and A309S (6- to 7-fold lower  $V_{\max}/K_M$  relative to xanthine), whereas D266E and D266N display 10- to 12- and 15-fold lower efficiencies for both the uric acid and xanthine, respectively, compared to the wild type. The same is true for A309C and I274S. The xanthine transport efficiency in these cases is 12- and 15-fold lower and the uric acid efficiency is 8- and 27-fold lower compared to the wild type. Six mutants (H31Q, S93N, D266E, D266N, A309G, V312C) display specifically low affinity for uric acid (>3-fold higher  $K_M$  relative to wild type), whereas M291C and T304P have low affinity for both uric acid and xanthine.  $K_M$  is 2.5- to 3.5-fold higher for xanthine and 11-fold higher for uric acid relative to wild type.

Following up on the above results, SmLL9 mutants were subjected to xanthine uptake assays in the presence of uric acid, 7-methylxanthine or 8-methylxanthine, which are not recognized by the xanthine-specific XanQ, in order to understand the effect of each mutation on SmLL9. Each one of the three ligands inhibit most of the mutants displaying  $K_i$  values similar to wild type, but S93N displays 12-fold lower affinity (12-fold higher  $K_i$ ) for uric acid. Four mutants (H31N, H31Q, A265C, T304P) are inhibited with 3- to 4-fold higher  $K_i$  by uric acid and five mutants have lower affinity for 7-methylxanthine with 2- to 2.5-fold higher  $K_i$ , like the A309C, or 8-methylxanthine with 3- to 5-fold higher  $K_i$ , like H31N, V312C or both in

the case of H31Q, S93N. Of the above mutants, S93N stands out as the one which shows reduced affinity for uric acid and, to a lesser extent, for the methyl analogs of xanthine at C8 or N7 and, also, very low efficiency for uric acid transport (**Figure 3.4.1.7**). The reciprocal mutant (N93S) in XanQ allows marginal uric acid uptake and low-affinity recognition of uric acid, 7-methyl and 8-methylxanthine<sup>197</sup>. A comparison of the effects of uric acid and 8-methylxanthine on SmLL9 and the relevant SmLL9 and XanQ mutants illustrates that the S93N replacement deteriorates the ability of SmLL9 to recognize uric acid and 8-methylxanthine imitating partly the properties of XanQ and, more precisely, mimicking the N93S mutant.

### 3.4.1.7 The important role of S93 in binding uric acid in SmLL9 transporter as revealed by Induced Fit Docking calculations (IFD)

Xanthine binds in SmLL9 similarly to UapA, as indicated by the IFD calculations (**Figure 3.4.1.8**).



**Figure 3.4.1.8:** Structural similarity between the SmLL9 and UapA binding sites. Xanthine substrate (A) in the binding cavity of SmLL9 as resulted from IFD calculations and (B) in UapA crystal structure.

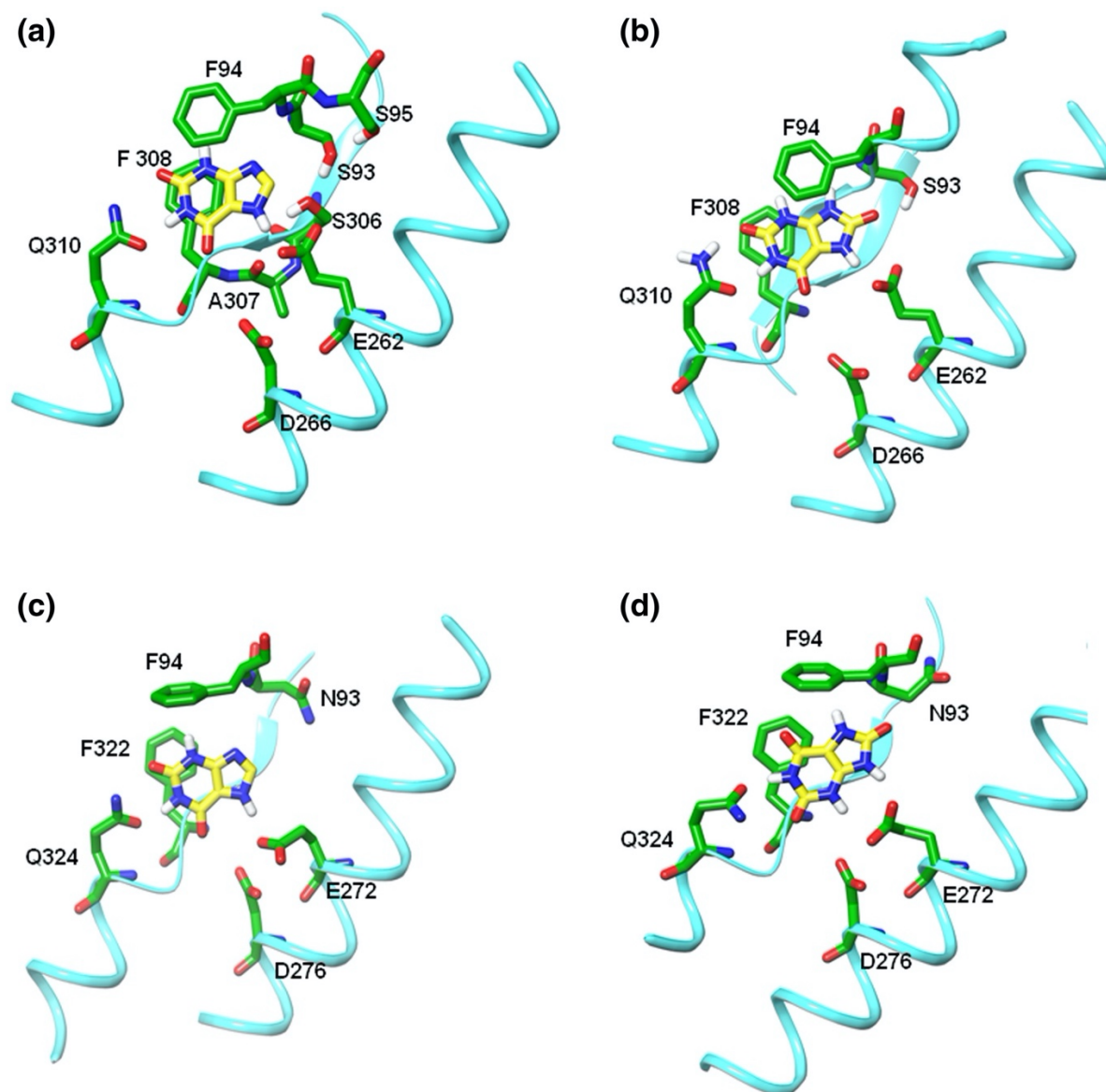
The O2 and N1 atoms of xanthine form two H-bonds with the amide of Q310 side chain while N7 an H-bond with E262. C6=O forms an H-bond with the A309 backbone NH. The stabilization of xanthine is enhanced by  $\pi$ - $\pi$  stacking interactions with F94 and F308 as in the crystal structures of UapA and UraA (**Figure 3.4.1.9A**). Further, the IFD calculations reveal the probable existence of an H-bond network in the vicinity of the substrate between S95-S93-S306 side chains and A307 backbone C=O.

In the case of XanQ, xanthine is bound slightly tilted compared to SmLL9, forming H-bonds between N7 and E272 carboxylate, between N1-H, C2=O and Q310 amide side chain

and between C6=O and the backbone NH of A323. Xanthine forms a  $\pi$ - $\pi$  stacking interaction with F322 and a T-stacking with F94 (**Figure 3.2.1.9C**). Additionally, it is interesting to note that during the MD simulations the N93 side chain is relatively flexible, interacting with different residues in the vicinity, namely S95, S269, T320, and E272.

Uric acid binds to SmLL9 with the same geometry and H-bonds as xanthine according to IFD calculations (**Figure 3.4.1.9B**) yet forming an extra hydrogen bond between C8=O and the S93 hydroxyl group connecting thus the substrate with the above mentioned S95-S93-S306-A307 H-bond network.

In contrast, when uric acid is docked inside XanQ binding cavity, it binds in a reverse orientation as N7 and N3 are directed toward E272, and C6=O is oriented toward F94 inversely to the corresponding xanthine carbonyl group which points to the E272 side chain (**Figure 3.4.1.9D**). Uric acid appears to form two H-bonds between N1H and C2=O and Q324 amide. However, the angle between the two planar systems, namely the Q324 amide and the substrate aromatic system, is  $\sim 45$  degrees, less optimal for the formation of strong H-bonds. Moreover, the C8=O group is not able to interact with the side chain of N93, which faces toward the opposite side of the binding cavity.



**Figure 3.4.1.9:** Structural basis of the interaction of SmLL9 and XanQ with xanthine and uric acid (UA). Low energy structures of transporter-substrate complexes as resulted from the induced-fit docking analysis. (a) SmLL9-xanthine. (b) SmLL9-UA. (c) XanQ-xanthine. (d) XanQ-UA. The H-bond network between S95-S93-S306 is illustrated in (a)

Overall, the docking calculations suggest that uric acid does not fit well in the binding cavity of XanQ and also the latter displays weaker interaction with uric acid compared to xanthine. Uric acid cannot be accommodated in the binding cavity and for that adopts a reverse orientation of the purine moiety compared to xanthine. The angle of the H-bond with Q324 is not optimal so the interaction is weak and moreover, there is no interaction between C8=O and the N93 side chain due to spatial and geometry reasons. It is interesting that, although the major difference between XanQ and SmLL9 in the binding cavity is related to residue 93, this difference is enough to produce weaker XanQ-uric acid interaction and generate specificity in

favor of xanthine. The presence of the H-bond between C8=O of uric acid and S93 hydroxyl group is necessary for the molecule to bind to SmLL9 (**Figure 3.4.1.9B**). This interaction stabilizes the pose which is the energetically most favorable and geometrically suitable according to the UapA-xanthine based model. This agrees with the mutagenesis data that show that the S93N replacement leads to impairment of the uric acid binding and uptake activity.

### 3.4.1.8 Conclusion

SmLL9 displays substantially a broad substrate profile unlike other corresponding homologs in *E. coli* K-12, which use separate transporters for xanthine or uric acid and adenine or guanine/hypoxanthine<sup>190</sup>. SmLL9 was subjected to in silico calculations and site-directed mutagenesis to reveal key factors for its function and its difference from well experimentally-studied xanthine-specific XanQ of *E. coli*<sup>113</sup>, given that the 3D homology models of both the transporters can be constructed based on the structurally resolved homologs UapA and UraA. The in silico studies reveal that SmLL9 and XanQ share several major functionally important residues which have similar roles in both transporters and Mutagenesis data, support these findings. The structural models constructed indicate that the xanthine-binding residues E262/272 and Q310/324, are involved in important H-bonds at the periphery of the binding site residues H31/31, D266/276, D290/304, N311/325. but also residues of the binding-site that are highly sensitive to alkylation by *N*-ethylmaleimide with similar IC<sub>50</sub> values namely A265/G275, A309/A323. According to the IFD calculations, xanthine is bound to both transporters with equivalent interactions, while, SmLL9 has the additional ability to utilize uric acid on a very similar molecular background with the xanthine-specific XanQ. Xanthine transport kinetics of SmLL9 resemble those of XanQ as indicated by the very similar  $K_M$  values and comparable  $K_i$  values for inhibition by oxypurinol and other analogs, supporting the IFD results.

In SmLL9, S93 appears to play a key role in transporter's additional ability to recognize and transport uric acid. The IFD analysis for the binding-site interactions of SmLL9 and XanQ with xanthine and uric acid highlighted that S93 in SmLL9 is necessary as it forms a H-bond with C8=O of uric acid, an interaction that is absent in XanQ. Furthermore, the presence of Asn93 instead of S93 in XanQ is enough to weaken the interactions with uric acid in the binding cavity of XanQ (**Figure 3.4.1.9B,D**) consistent with the strict specificity of XanQ relative to SmLL9. Confirming the IFD results, the replacement of S93 with Asn leads to a specific >12-



fold lower affinity for uric acid compared to wild type, and hampers uric acid transport (**Figure 3.4.1.7**).

The importance of S93 from previous findings on other homologous NAT/NCS2 transporters and its specificity role is supported and even extended by our results. A Ser at this position is conserved in TMS3 at the binding site region in all dual-specificity xanthine/uric acid NAT/NCS2 transporters. This Ser was shown to impact specificity as indicated in mutagenesis studies for XanQ and UapA, though, these studies do not provide strong evidence that this Ser is important for binding and transport of uric acid. The crystal structure of UapA is in complex with xanthine and the binding of uric acid has not been studied with in silico analysis. The relevant UapA mutants S154N and S154A are defective in transport of both xanthine and uric acid, although they have different substrate profiles compared to the wild-type, they retain significant affinity for uric acid. The relative transport efficiencies for the two substrates are not specified yet. In XanQ, the relevant mutant N93S hardly binds and transports uric acid in addition to xanthine<sup>193,197</sup>. In UacT, the relevant mutant T100S shows very low uptake activity for uric acid and hardly detectable ability to transport xanthine in addition to uric acid<sup>189</sup>. Thus, the SmLL9 case presented here is the first one to show that a Ser at this position is important for the recognition and transport of uric acid.

In UapA, amino acids participating in gating elements, have been also suggested to influence specificity by acting as “selectivity gates” controlling the access of substrates to the binding site<sup>63,76</sup>. In silico and mutational analysis studies have shown that recognition of atypical substrates like hypoxanthine and adenine can be achieved, without affecting the recognition and transport of the physiological substrates, xanthine or uric acid<sup>1,58,126,128</sup>. However, there is poor evidence for such gating elements in bacterial NAT/NCS2 members. Very few mutations outside the binding site of XanQ allowed recognition of non-wild-type ligands in addition to xanthine, thus there has been no evidence so far for high-affinity recognition or transport at a significant rate for such non wild-type ligands<sup>188</sup>.

### **3.4.2 The case of AncXanQ transporter**

*As adapted from “Identification of New Specificity Determinants in Bacterial Purine Nucleobase Transporters based on an Ancestral Sequence Reconstruction Approach” by Ekaterini Tatsaki EleniAnagnostopoulou Iliana Zantza Panayiota Lazou Emmanuel Mikros Stathis Frillingos, Journal of Molecular Biology (December 2021)*

### 3.4.2.1 Foundation of the study

The NAT/NCS2 (Nucleobase-Ascorbate Transporter/Nucleobase-Cation Symporter-2) family, the most common nucleobase transporter family in microorganisms as mentioned earlier, is present in almost all major phyla of organisms and all domains of life<sup>113,195,202</sup>, though in primates the group is represented by two L-ascorbate transporters homologs, SVCT1 and SVCT2, SLC23A3 which is a functionally unknown member and a nucleobase-transporter related pseudogene, namely SNBT1/Slc23a4<sup>7,203</sup>. Few homologs have been functionally characterized. It is interesting that although all known nucleobase-related eukaryotic homologs are of broad specificity<sup>2,95,204,205</sup>, most known bacterial homologs display distinct and highly specific substrate profiles<sup>113,190,206,207</sup>.

The folding of NCS2 family transporters is known from the structurally solved uracil permease UraA from *Escherichia coli* in an occluded conformation<sup>56,70</sup>, and xanthine/uric acid permease UapA from *Aspergillus nidulans*, solved in inward-open conformation. In addition, they share structural homology with transporters of at least two additional families (AE/SLC4, SulP/SLC26)<sup>50-52,55</sup> and appear to function as homodimers. Apart from these transporters, several transporters belonging to other families also perform under the elevator mechanism<sup>75,83</sup>.

As it has been already shown, specificity is regulated not only from changes in the binding pocket, but also from alterations of residues at the interface of the two domains, which seem to act by decreasing the restraints for the access of non-canonical substrates to the binding site<sup>63</sup> and/or uncoupling the elevator from high-affinity binding of substrates<sup>82</sup>.

Regarding the diverse substrate profiles and the molecular background of the different NAT/NCS2 specificity mechanisms, our knowledge is limited up to the present time to the *in-silico* calculations and the rigorous mutational analysis of the fungal UapA, and the bacterial XanQ and UraA and the comparative analysis of these homologs with other related transporters.

Here, we investigate the specificity profiles of bacterial NAT/NCS2 transporters by using docking and molecular dynamics (MD) calculations, accompanied with an evolutionary experimental strategy different from the aforementioned methods, established on ancestral sequence reconstruction. Ancestral sequence reconstruction has been used to analyze diverse groups of proteins<sup>208,209</sup>, yielding important conclusions on the molecular basis of the rise of different functions and specificities in extant proteins, including studies that have shown that few sequence changes and rather simple mechanisms lead to new functions. Application of this experimental strategy to transmembrane systems is limited to date, due in part to lack of

structural and functional knowledge on an adequate number of homologs in most cases. The expansion of available atomic-resolution structures of membrane transport proteins through current advances in crystallography and single-particle electron cryo-microscopy<sup>210-212</sup>, will certainly offer a basis for more systematic application of such evolutionary strategies to membrane transporters. In this case, we use ancestral protein reconstruction to study the evolution of bacterial xanthine permeases of the NAT/NCS2 family. Our results provide evidence that the current xanthine-specific transporters, derived from xanthine-guanine transporters of broader specificity and the shift from polyspecific transporters to xanthine-specific ones is associated with few changes outside the binding-site.

### 3.4.2.2 Structural model of AncXanQ transporter

As described in the previous section concerning SmLL9 transporter, the xanthine permease XanQ of *E. coli* K-12 is the only one among the functionally known permeases of NAT/NCS2 family that is specific for xanthine and does not transport any other purine or pyrimidine. Several well characterized transporters in the XanQ/UapA cluster (or Cluster C1\_Xanthine-Uric Acid) of NAT/COG2233 subfamily<sup>195</sup>, like the structurally resolved fungal UapA and the rhizobial SmXUacT<sup>207</sup>, as well as other uric acid-specific like the *E. coli* UacT<sup>189</sup>, and the rhizobial SmUacT1 and SmUacT2, transport both xanthine and uric acid. Albeit their sequence similarity in all key binding-site residues, the xanthine-specific XanQ differs from all the above transporters in few functionally relevant amino acid residues (see below).

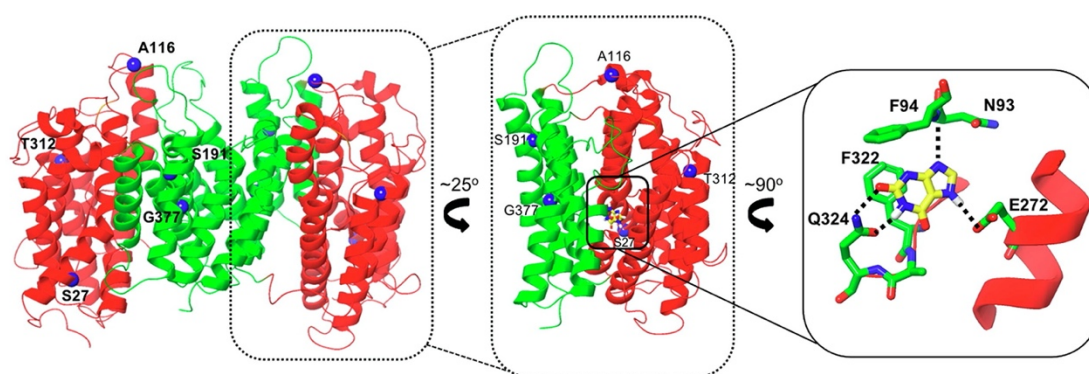
The homology model of XanQ has been constructed using UapA crystal structure as a template, as described in the previous section. Upon the XanQ model inspection and the relative sequence alignment of all the homologs in the XanQ clade (**Figure 3.4.1.1**), we noticed that apart being closely related to each other with >65% pairwise sequence identity, they share a signature-motif of four important residues absent in any other NAT/NCS2. With numbering according to XanQ, these are: Asn93, which is part of a H-bond network in the periphery of the binding site and has been linked with the preference for xanthine over uric acid<sup>113,197</sup>, A323, located in the substrate binding cavity and may also have a role in specificity<sup>189,199</sup>, Ser/G377, next to a conserved Phe/Tyr that contributes to substrate coordination at the broader area of the binding site<sup>188</sup>, and N430, which is distant from the binding site but may influence specificity indirectly<sup>126,201</sup>. Of these residues, Asn93 and N430 are always present in the XanQ clade but absent in any other NAT/NCS2 homolog, A323 is invariable in the XanQ clade and corresponding to Ala, Gly or Ser in other NAT/NCS2s, and S377 is uniquely present in a subset



**Figure 3.4.2.1:** Phylogenetic and sequence analysis of AncXanQ. (A) Phylogenetic analysis of the 20 XanQ orthologs representing one genome per XanQ-containing species derived from the analysis of fully sequenced genomes retrieved from the IGM/M database at JGI. (B) Characteristic amino acid residues found in all (top), all gamma-proteobacterial (middle) or all beta-proteobacterial XanQ orthologs (bottom) and their flanking sequence context. The residues shown in red and underlined (and numbered according to XanQ) are either important for function in XanQ and not conserved in other NAT/NCS2 groups (XanQ-clade signature) or distinguish between the gamma- and beta-proteobacterial orthologs and linked with specificity, as shown in the current study. (C) Amino acid occupation of the positions highlighted in B in the functionally known proteobacterial homologs of the XanQ/UapA cluster. (D) Homology model of AncXanQ based on the inferred amino acid ancestral sequence showing separately the region of the core domain (residues 22–135 and 259–352) and the gate domain (residues 151–229 and 368–445). Positions of amino acids predicted with a posterior probability of <90% (but >50%) or <50% are named and shown in orange or red, respectively. The AncXanQ sequence (shown in Supplement, Figure S2) was inferred using Maximum Likelihood statistics in MEGA7. (E) Sequence alignment of AncXanQ with homologs of the XanQ/UapA cluster. The alignment was prepared with MultAlin. High consensus sites (occurring at >90% of the sequences) are shown in red and low consensus sites (>50%) in blue. The positions of alpha-helices and beta-sheets shown on top of the sequences and the xanthine-binding residues (indicated as dots) are according to the UapA structure PDB no. 5I6C (Alguel *et al.*, 2016). The five amino acid residues analyzed in this study are indicated with an asterisk on top of the sequences. Apart from the functionally solved members of the XanQ clade (AncXanQ, AvpXanQ, XanQ, NmXanQ), the alignment includes UapA (Q07307), *E. coli* XanP (P0AGM9), and the XanP-paralogs AvDGQ0 of *Aeromonas veronii* (UniProt A0A1621446) and D0S6Q3 of *Acinetobacter calcoaceticus* (UniProt D0S6Q3).

So far there has been found none XanQ mutant or other homologs in the XanQ/UapA cluster (over 1000 site-directed mutants at more than 80% of the permease residues)<sup>189,207</sup>, with substantially altered profile from the xanthine-specific one<sup>113,188</sup>.

In order to discover the basis of this resilient substrate profile of XanQ, we explore the putative ancestral protein sequence of the XanQ clade namely AncXanQ and its properties regarding substrate specificity, the binding mode of xanthine, guanine and their analogs in the AncXanQ binding pocket, starting by constructing the 3D structure of the latter. The homology structural model of AncXanQ was constructed by prof. Frillingos' team, using as a template XanQ since the two sequences share high identity (sequence alignment of AncXanQ with homologs of the XanQ/UapA cluster depicted in **Figure 3.4.2.1E**). The XanQ model was constructed in an inward-open conformation because UapA crystal structure<sup>58</sup> was used as a template which had been crystallized in an inward-open conformation<sup>207</sup> (PDB 5I6C), thus the AncXanQ model was in an inward-open conformation, too. The alignment of these three transporters along with the AncXanQ model has a folding of 14 transmembrane segments (TMs), of which TMs 1–4 and 8–11 form the core domain and TMs 5–7 and 12–14 the gate domain (**Figure 3.4.2.2**). All transmembrane helices and interconnecting loops display the same length as in XanQ, due to the high identity (76%) shared between the two transporter sequences.



**Figure 3.4.2.2:** AncXanQ-5M approximates the profile of the xanthine-specific XanQ. (A) AncXanQ dimer derived from homology modeling. The core domain is shown in red and the gate domain in green. The positions of residues S27, A116, T312, S191 and G377 are presented as blue spheres, while the computational predicted mode of xanthine in the binding site is highlighted.

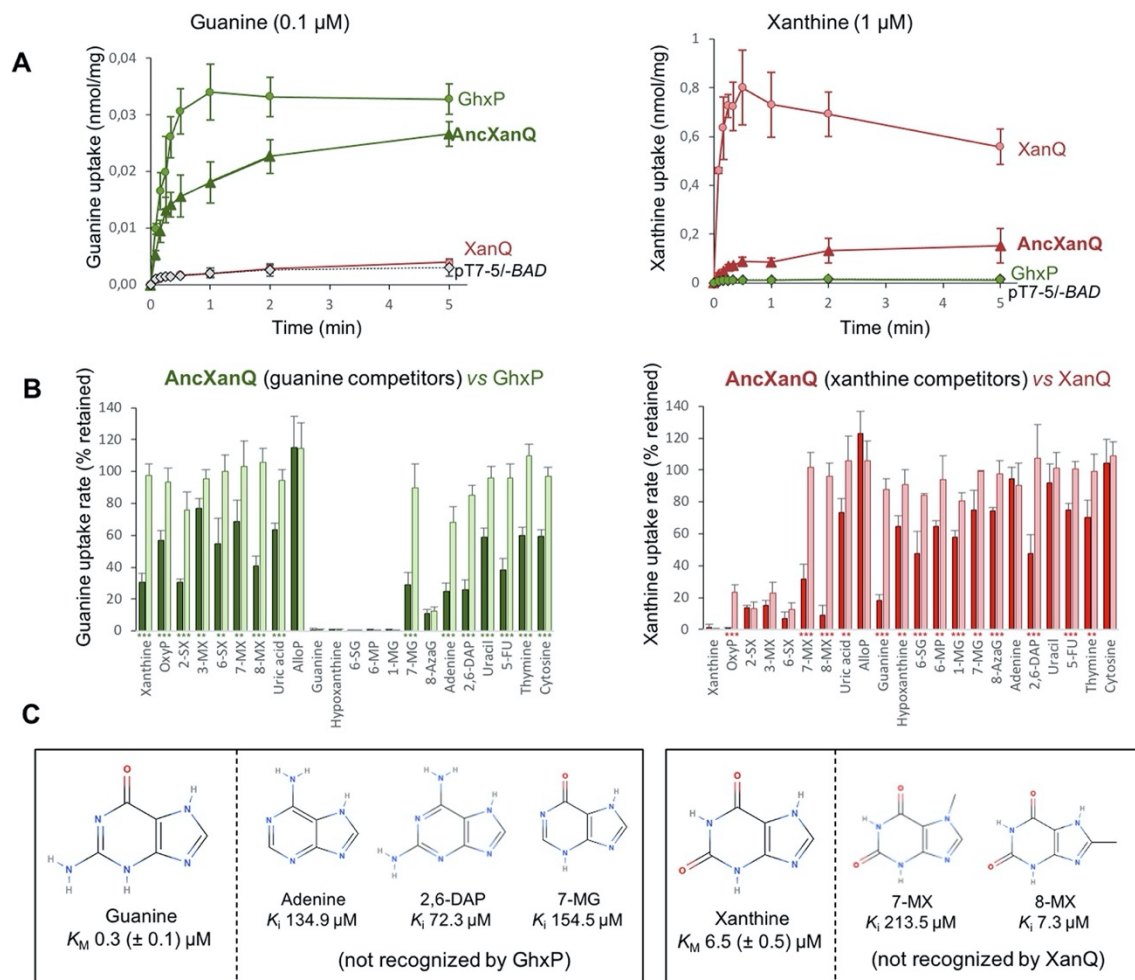
The initial apo AncXanQ model was subjected to 50 ns molecular dynamics simulation in the *apo* form embedded in DPPC lipid bilayer solvated with explicit water. The calculated RMSD of the C $\alpha$ -carbons did not exceed 1 Å, indicating the stability of the model structure. During the trajectory, we noticed the behavior of important amino acid side chains. F94 of the binding site, which has an important “gate-keeping role” as reported for UapA, fluctuates quite a lot. Q324 side chain instead of being directed towards the binding site, it is rotated interacting with backbone of neighboring residues.

Upon a close visual inspection of the alignment of AncXanQ, XanQ and UapA, one can notice very interesting similarities. In detail, the important binding site residue of UapA, E356 and D360, are in a well conserved motif in TMS8, E-X-GD, shared by the three transporters, highlighting the important role of this motif mostly in substrate binding and not specificity, as these transporters differ in specificity. Moreover, in TMS10 another one more important motif namely TFAQNNGV containing the UapA binding site residue Q408, is also conserved. at the end of TMS12, the important R481 in UapA as it mediates the release of xanthine to the intracellular environment according to standard MD calculations, is not conserved in XanQ or AncXanQ, albeit the R478, R479 are fully conserved, raising the hypothesis that this fact may display the difference of the mechanism of substrate release. Mutations in R481 do not lead to alterations of transport of physiological substrates but rather modulate the specificity of the protein allowing low-affinity transport of hypoxanthine or adenine.

### 3.4.2.3 Transport assay analysis of AncXanQ

Transport assays conducted by Frillingos research group, indicate that AncXanQ displays lower xanthine uptake activity but broader substrate profile than XanQ (**3.4.2.3A and B**). In detail, putative purine and pyrimidine substrates were tested (adenine, guanine, hypoxanthine, xanthine, uric acid, uracil thymine, cytosine). AncXanQ was found to transport guanine and xanthine, while XanQ transports only xanthine. AncXanQ transports xanthine with comparable affinity (KM) to XanQ (6.5  $\mu\text{M}$  and 4.1  $\mu\text{M}$  accordingly) but 5-fold lower efficiency ( $V_{\text{max}}/\text{KM}$ ) (0.37  $\text{ml}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$  and 1.83  $\text{ml}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$  accordingly). Regarding the guanine-transport activity, AncXanQ transports guanine with high affinity (KM 0.2  $\mu\text{M}$ ), which is higher than the one of GhxP, the major guanine-hypoxanthine transporter of *E. coli* K-12<sup>190</sup>, but lower capacity ( $V_{\text{max}}$  2.4  $\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ ), resulting in comparable efficiency ( $V_{\text{max}}/\text{KM}$ ) relative to the modern guanine transporter. Both XanQ<sup>191</sup>, and AncXanQ are inhibited by carbonyl cyanide m-chlorophenyl hydrazone (CCCP), denoting that they are proton-gradient dependent.



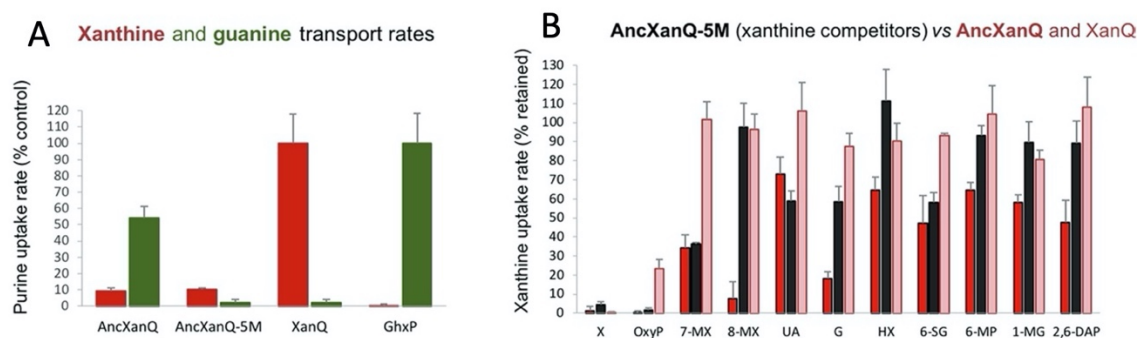


**Figure 3.4.2.3:** AncXanQ is a broad-specificity guanine/xanthine transporter. (A) *E. coli* JW4025 and T184 expressing the indicated permeases through pT7-5/-BAD were assayed for uptake of [<sup>3</sup>H]-guanine (0.1  $\mu\text{M}$ ) and [<sup>3</sup>H]-xanthine (1  $\mu\text{M}$ ), respectively, at 25 °C, as shown. Each data point represents the mean and standard deviation (SD) from triplicate measurements. (B) *E. coli* JW4025 expressing AncXanQ (dark green) or GhxP (light green) were assayed for initial rates of [<sup>3</sup>H]-guanine (0.1  $\mu\text{M}$ ) uptake and T184 expressing AncXanQ (dark red) or XanQ (light red) were assayed for initial rates of [<sup>3</sup>H]-xanthine (1  $\mu\text{M}$ ) uptake, in the absence or presence of the indicated non-radiolabeled nucleobases and analogs (1 mM). Transport rates are expressed as percentages of the rate obtained in the absence of competitor. Each bar represents the means of five to six determinations with standard deviation (SD) shown. The values obtained for the ancestral and the extant transporters were compared for each ligand using unpaired two-tailed t-test (as implemented in GraphPad Prism 8.0) and statistically significant differences are indicated with one ( $p < 0.05$ ), two ( $p < 0.01$ ) or three asterisks ( $p < 0.001$ ) under each pair of histogram bars. Values obtained with cells harboring pT7-5/-BAD alone have been subtracted from the sample measurements in all cases. (C) AncXanQ ligands that are not recognized by GhxP (on the left) or XanQ (on the right).

The xanthine transport activity of AncXanQ is inhibited by a range of analogs, with high affinity, including 7-methylxanthine, 8-methylxanthine, and guanine, which do not serve as XanQ ligands (Figure 3.4.2.4). Many purine nucleobases and analogs inhibited the transport activity of guanine in AncXanQ with high affinity. These purines include hypoxanthine (HX), 6-mercaptopurine (6-MP), 6-thioguanine (6-SG), 1-methylguanine (1-MG), and 8-azaguanine



(8-azaG), which are high-affinity ligands for GhxP<sup>190</sup>, but also adenine (A), 2,6-diaminopurine (2,6-DAP), 7-methylguanine (7-MG), and xanthine which are not ligands for GhxP (**Figure 3.4.2.4**). Overall, AncXanQ recognizes as ligands except guanine and xanthine, adenine and hypoxanthine, and all of their analogs tested, except allopurinol and uric acid. This profile is different than any modern NAT/NCS2 profile recorded so far for either a bacterial or a eukaryotic transporter.



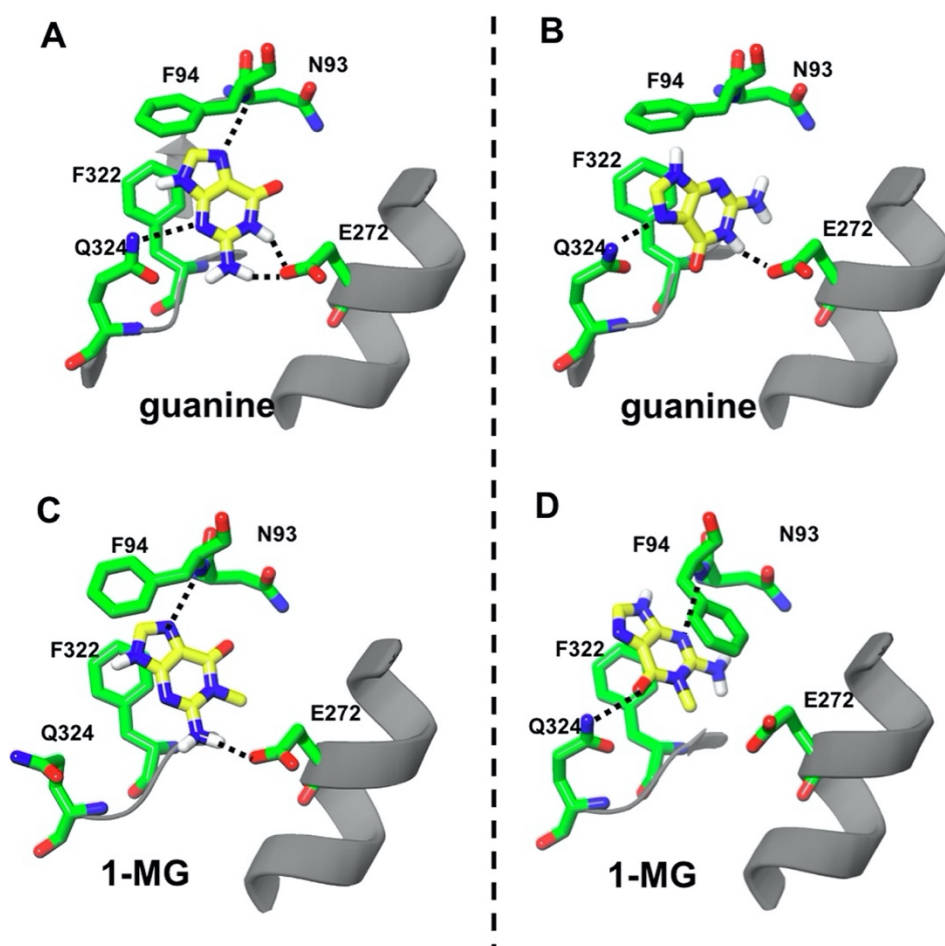
**Figure 3.4.2.4:** (A) Xanthine and guanine transport rates for AncXanQ and the quintuple mutant AncXanQ-5M (S27G/A116L/T312S/S191G/G377S) relative to the xanthine transporter XanQ and the guanine transporter GhxP. Initial rates of [<sup>3</sup>H]-xanthine (1  $\mu$ M) or [<sup>3</sup>H]-guanine (0.1  $\mu$ M) uptake were measured with *E. coli* T184 or JW4025, respectively, expressing the indicated permeases. The results are expressed as percentages of the value obtained for XanQ (xanthine uptake, red bars) or GhxP (guanine uptake, green bars). Values obtained with pT7-5/-BAD alone were subtracted from the sample measurements in all cases. Each bar represents the means of three determinations with standard deviation (SD) shown. (B) Comparison of the ligand inhibition profile of AncXanQ-5M with the profiles of AncXanQ and XanQ, highlighting differences in certain analogs. *E. coli* T184 expressing AncXanQ (dark red), AncXanQ-5M (black) or XanQ (light red) were assayed for initial rates of [<sup>3</sup>H]-xanthine (1  $\mu$ M) uptake, in the absence or presence of non-radiolabeled analogs (1 mM).

### 3.4.2.3 Binding mode of xanthine and guanine on AncXanQ

In order to characterize the binding mode of both the physiological substrates of AncXanQ, xanthine and guanine, induced fit docking calculations (IFD) were implemented using these ligands, in the binding area of the transporter. Remarkably, xanthine is bound in the binding pocket of AncXanQ with the same binding mode and forming the same interactions as in UapA and in XanQ in the lowest energy binding pose (**Figure 3.4.2.2**). In detail, N1H and C2=O form bidentate with Q324, N7H forms H-bond with E272 carboxylate, N9 forms H-bond with backbone carbonyl group and T stacking and p-p interactions with F94 and F322, respectively. N93 does not seem to interact with xanthine.

The selected binding poses showed to be the most energetically favorable are presented for guanine and 1-MG **Figure 3.4.2.5**. In the lowest energy pose (Figure 4(A)) guanine is stabilized in AncXanQ by H-bond formation of N1H and C2-NH2 with E272 carboxylate as well as N3 with Q324, N7 with F94 backbone carbonyl group and p-p and T stacking

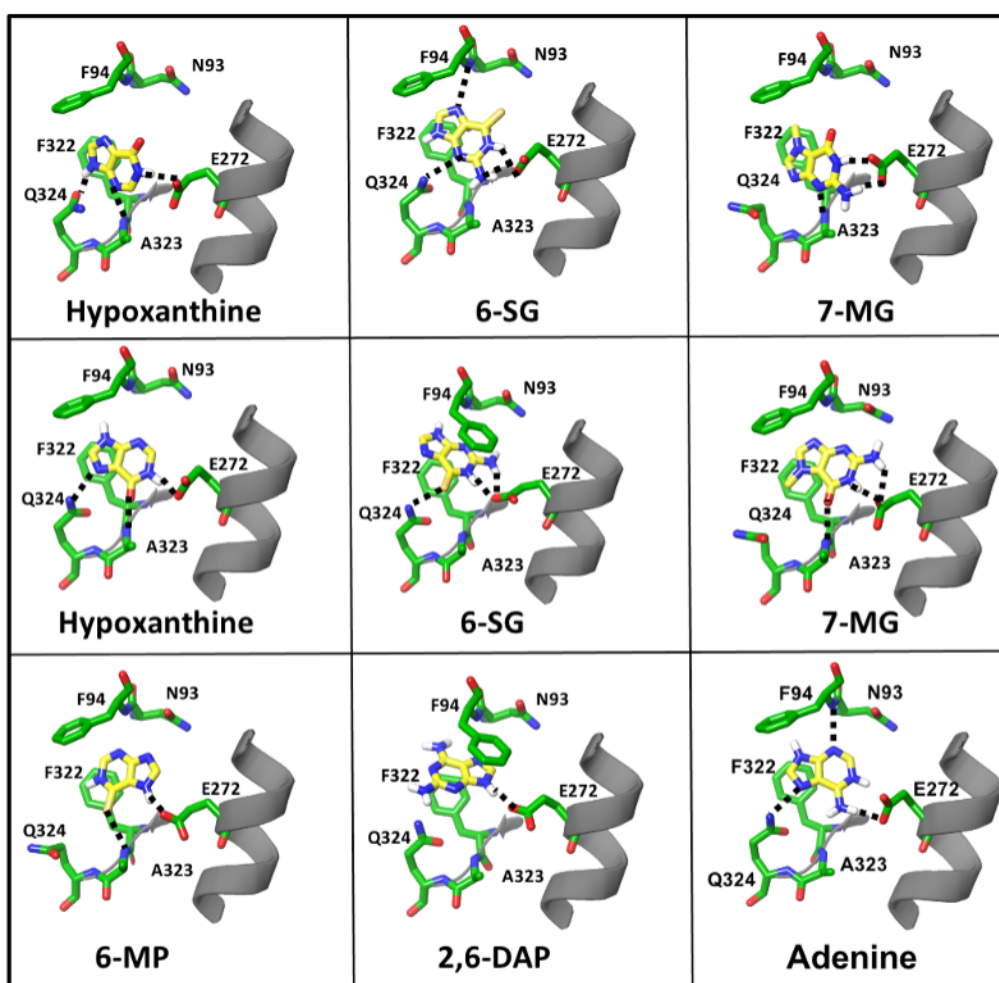
interactions with F322 and F94, respectively. In the lowest energy binding pose (**Figure 3.4.2.5C**), 1-MG is stabilized in the binding cavity by H-bond formation of C2-NH<sub>2</sub> with E272, N7 with F94 backbone and p-p and T stacking interactions with F322 and F94, respectively. However, in the 1-MG complex, there is a steric clash between the methyl group and the E272 carboxylate which implies a slight shift compared to guanine resulting in loss of the interaction with Q324 (**Figure 3.4.2.5C**). In the second lowest energy pose (**Figure 3.4.2.5B**), a H-bond is formed between guanine N1H and E272, C6 = O and A323 backbone, N7 and Q324, as well as p-p and T stacking interactions with F322 and F94, respectively. In the similar orientation but slightly shifted compared to guanine, 1-MG forms H-bonds between N3 and F94 backbone, C6 = O and Q324 and p-p stacking interactions with both F322 and F94 in the second lowest energy pose (**Figure 3.4.2.5D**). It is interesting to underline that guanine and consequently 1-MG, orient the pyrimidine moiety of the purine towards E272 in both cases.



**Figure 3.4.2.5:** Molecular Docking Simulations of the AncXanQ-guanine interactions. Common lowest-energy binding modes of guanine and 1-MG are presented. In the first binding mode, guanine (A) and 1-MG (C) are oriented so that their C2-NH<sub>2</sub> forms a H bond (dashed line) with E272 and their N7 a H bond with the F94 backbone. In the second binding mode, guanine (B) and 1-MG (D) are oriented so that their C6 = O forms a H

bond with the A323 backbone (guanine) or Q324 (1-MG) while N7 in guanine forms a H bond with Q324 and N3 in 1-MG forms a H bond with the F94 backbone. For further details, see text.

As shown in **Figure 3.4.2.6**, in the case of the other guanine analogs in the lowest-energy binding pose with AncXanQ, the orientation of the ligand and the binding interactions may vary. Explicitly, hypoxanthine, 6-SG, 7-MG, and adenine in the lowest-energy orientations share the same binding pose with guanine and 1-MG in the binding pocket, however 2,6-DAP and 6-MP obtain a binding orientation and binding interactions that highly resemble those of xanthine.



**Figure 3.4.2.6:** Binding modes of different inhibitors with AncXanQ. Hypoxanthine, 6-SG and 7-MG (first two rows) and adenine (third row, on the right) adopt two low-energy binding modes which are similar to the two alternative favored orientations and interactions of guanine and 1-MG (Figure 4). For 6-MP and 2,6-DAP (third row, on the left and in the middle), the adopted lowest-energy pose is similar to the favored binding orientation and interactions of xanthine (Figure 3.4.2.2).

#### **3.4.2.4 Combination of 5 mutations outside the binding site restricts specificity of AncXanQ to recognition of xanthine**

AncXanQ differs from XanQ at only five amino acid positions predicted to be in transmembrane segments, along with variations in conserved hydrophobic residues (Figure 3(A) and Supplementary Figure S3). These five amino acids are: S27 (Gly in XanQ); A116 (Leu in XanQ); S191 (Gly in XanQ); T312 (Ser in XanQ); and G377 (Ser in XanQ) and are located outside the binding site.

A mutant of AncXanQ changing all these five residues to the corresponding of XanQ (S27G/A116L/S191G/T312S/G377S) was constructed and named AncXanQ-5M and was expressed in the *E. coli* host membrane normally. AncXanQ-5M transports xanthine but not guanine, (**Figure 3.4.2.3B**), while according to ligand inhibition assays does not recognize 8-methylxanthine but recognizes 7-methylxanthine with high affinity (**Figure 3.4.2.3C**). Hence, AncXanQ-5M is mimicking the profile of XanQ as a xanthine-specific transporter.

#### **3.4.2.5 The role of Ser/G377 in relevance to the xanthine specificity of XanQ as derived from experimental procedures**

Along with AncXanQ and XanQ, two XanQ-homologs, AvXanQ (from *Aeromonas veronii*) and NmXanQ (from *Neisseria meningitidis*), were studied. AvXanQ and NmXanQ conserve the five-residue signature of XanQ and AncXanQ, respectively. Precisely, NmXanQ contains the five residues S27/A116/A191/T312/G377 which differs in one residue (A191) from AncXanQ (S191). NmXanQ was found to transport both xanthine and guanine. A191S-NmXanQ mutant, has the same functional properties with NmXanQ, revealing that the Ala/S191 site is not important for specificity. AvXanQ was specific for xanthine.

Based on the evidence that AncXanQ and NmXanQ are broad-specificity xanthine/guanine transporters whereas XanQ, AvpXanQ and AncXanQ-5M are xanthine-specific, it is affirmed that the combination S27/A116/Ser (or Ala)191/T312/G377 of the five residues among XanQ homologs denotes broad specificity for purines for this homolog whereas the combination G27/L116/G191/S312/S377 suggests a xanthine-specific homolog. Inspecting explicitly the Ser/G377 residue position, it is found located next to a conserved Phe/Tyr residue (F376 in XanQ) which has a role of substrate coordinator as it is presented in the crystal structures of UraA and UapA<sup>56-58</sup>.

The Ser/G377 specificity role was explored by conducting series of experiments using mutants of the AncXanQ, NmXanQ, and XanQ. In a first attempt, S377 was mutated to Gly in

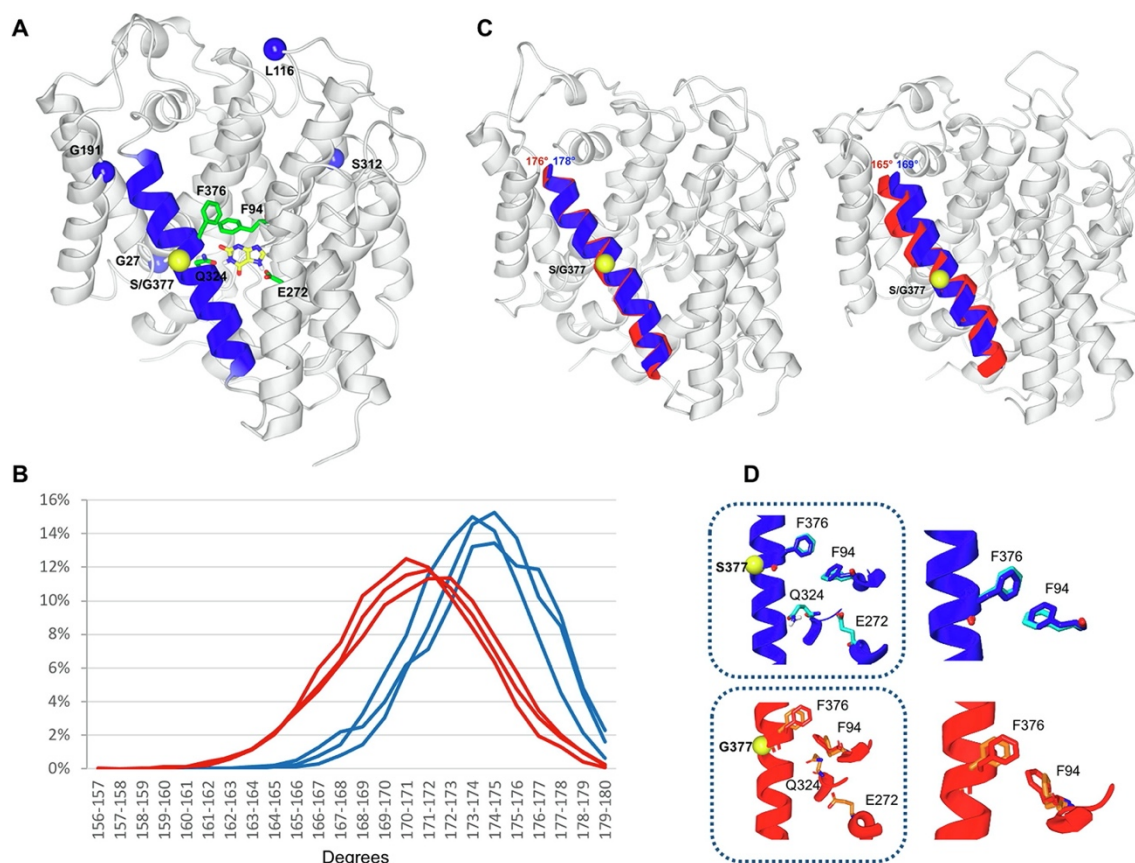
XanQ, introducing the XanQ-S377G mutant. XanQ-S377G was found to transport xanthine with similar kinetic properties as wild-type XanQ but, in addition, displays a high-affinity, albeit low-efficiency, transport for guanine, indicating that S377G mutation created a broad specificity mutant of XanQ. Also, in XanQ-S377G, xanthine uptake-activity is inhibited by guanine with a  $K_i$  42.3  $\mu$ M, but not by 7-methylxanthine, 8-methylxanthine and uric acid, and at the same time, xanthine, hypoxanthine, 6-thioguanine, 6-mercaptopurine, 1-methylguanine, and 7-methylguanine inhibited its guanine uptake-activity, while adenine, 2,6-diaminopurine, or 8-azaguanine did not, similarly to NmXanQ. However, XanQ-S377G does not recognize 3-methylxanthine differing from NmXanQ, AncXanQ, and wild-type XanQ.

Aside from S377G, the XanQ single mutants: G27S, L116A, G191S and S312T did not result in broaden specificity profile. Instead, they displayed the same affinity for xanthine and increased efficiency compared to the wild type, while xanthine transport is inhibited only by xanthine analogs which are high-affinity ligands of wild-type XanQ (S312T is also inhibited by 7-methylxanthine and 8-methylxanthine). The same holds for the G27S/L116A/G191S/S312T mutant and the combinations G27S/S377G and G27S/L116A/G191S/S312T/S377G which are also xanthine-specific. However, G191S/S377G has broadened guanine/xanthine profile similarly to S377G, whereas S312T/S377G is xanthine-specific like the wild-type XanQ. Thus, it emerges that the effect of S377G (broaden specificity) is suppressed by epistatic interactions with G27S and/or S312T. On the other hand, introduction of Ser in lieu of G377 was tested to investigate whether it suffices to establish specificity for xanthine in a broaden-specificity permease. Both AncXanQ-G377S and NmXanQ-G377S preserve the broad-specificity guanine/xanthine profile. The only difference obtained was that AncXanQ-G377S recognizes fewer xanthine and guanine analogs than wild-type in ligand inhibition analysis. Specifically, it recognizes guanine and guanine-related ligands of AncXanQ but does not recognize 7-methylxanthine, 8-methylxanthine, adenine, 2,6-aminopurine and 8-azaguanine which are high-affinity ligands for wild-type AncXanQ.

#### **3.4.2.6 Structural alterations in the XanQ S377G mutant**

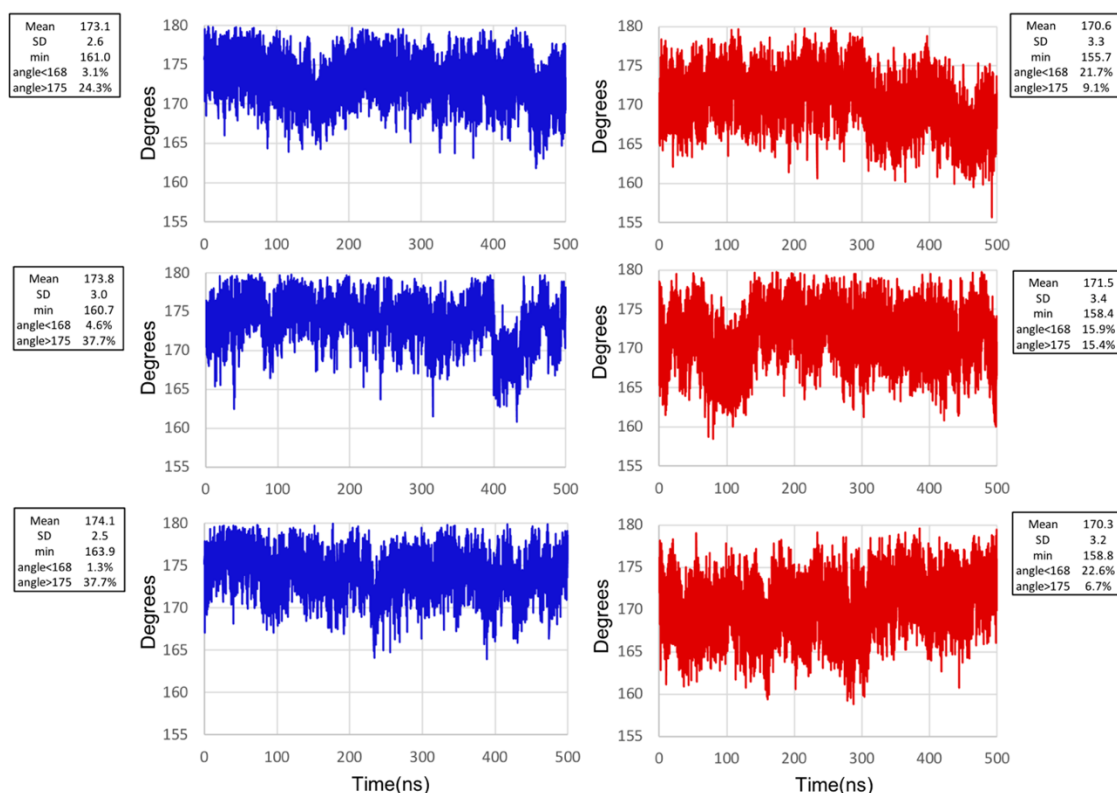
The specificity alterations of XanQ caused by the S377G mutation, where further computationally investigated. The already constructed model of XanQ was used as the wild-type structure and for the XanQ-377G a single replacement of S377 to Gly was made. Molecular Dynamics simulations of 500 ns were conducted on both the wild-type and the XanQ-377G which were embedded in lipid membrane solvated in explicit waters. During the

trajectory of the mutated structure, G377 serves primarily in TMS12 tilting. In detail, monitoring the angle formed between the C $\alpha$  atoms of residues 366, 377, and 384, it emerges that during the MD simulations, the distribution of angles in 377G-XanQ indicate that TMS12 bends more frequently towards 160 degrees compared to wild-type. The 5th and 95th percentiles of TMS12 tilting angle values are 165 and 176 in the mutant compared to 169 and 178 in wild type (**Figure 3.4.2.7B, Figure 3.4.2.8**).



**Figure 3.4.2.7:** Molecular Dynamics Simulations of the effect of S377G on XanQ. (A) Overall structural model of wild-type XanQ complexed with xanthine highlighting the key binding residues Q324 and E272, the interacting pair F376-F94 and TMS12 (in blue). The position of S/G377 is indicated with a yellow sphere and the positions of G27, L116, G191 and S312 are indicated with blue spheres. (B) Value distribution of TMS12 tilting angle formed between the C $\alpha$  atoms of residues 366, 377 and 384 in TMS12 in the wild-type XanQ (blue) and S377G XanQ (red), represented as population percentages per one-degree interval, as resulted from the 500 ns MD calculations after trajectory structure sampling every 60 ps (all data shown in **Figure 3.4.2.8**). Each one of the three lines of each color corresponds to the data from one of the three MD simulations for the relevant permease. (C) Structure snapshots of TMS12 presenting the 95th percentile (left) and the 5th percentile (right) of TMS12 tilting angle value distribution in the wild-type (blue) and S377G XanQ (red). (D) Interaction between F376 (TMS12) and F94 (TMS3) in the binding site of wild-type XanQ (blue/cyan) and S377G XanQ (red/orange). In wild type, a T-stacking interaction is observed between the F376-F94 phenyls in the 80% (blue) of the trajectory, while a second conformation exists covering the 3% of the trajectory where no interaction is observed (cyan). In the S377G mutant, the F94 dihedral monitored leads to phenyl orientations not suitable for interaction with F376, in the 45% of the trajectory (red), or exhibiting a weak interaction, in 43% of the trajectory (orange).





**Figure 3.4.2.8:** Fluctuation of helix 12 (TMS12) bending at residue 377 in wild-type XanQ and XanQ-S377G. The angle formed by the C $\alpha$  atoms of residues 366, 377 and 384 was monitored, in wild-type XanQ (blue) and S377G XanQ (red), during a 500-ns MD simulation. The MD simulation was run in triplicate, as shown, and the means, SD, minimum values, and percentages of angle values falling above 175 or below 168 degrees are indicated next to each MD run. Considering all runs, the mean ( $\pm$ SD) is 173.7 ( $\pm$ 0.5) for wild type ( $n=3$ ) and 170.8 ( $\pm$ 0.6) for the S377G mutant ( $n=3$ ), whereas the 5th and 95th percentiles are 169.0 ( $\pm$ 1.0) and 177.7 ( $\pm$ 0.6) for wild type ( $n=3$ ) and 165.3 ( $\pm$ 0.6) and 176.0 ( $\pm$ 1.0) for the S377G mutant ( $n=3$ ).

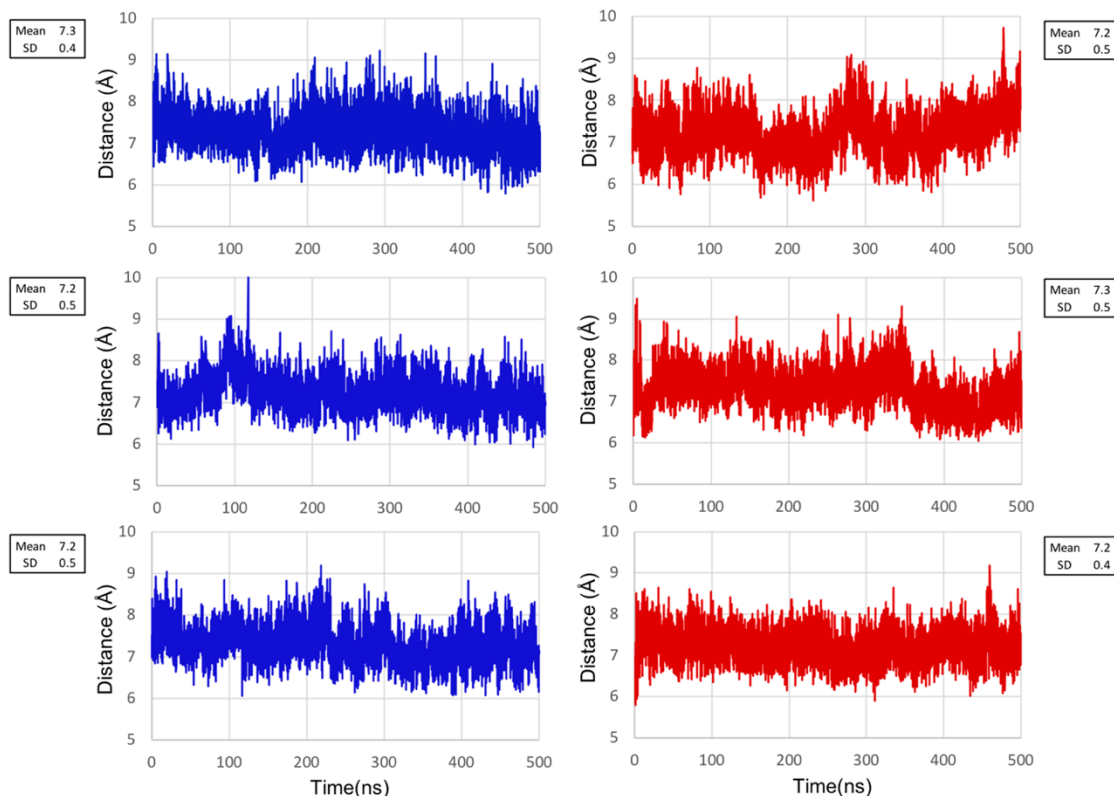
The RMSD of the 377 neighboring Ca residues (373 to 381) also support the previous finding. The average and the maximum RMSD values are 1.1 and 1.8 for the wild-type, while 1.6 and 2.3 for the XanQ-377G, respectively (**Table 3.4.1.1**). Moreover, G377 interacts appears to interact through backbone with T373 of the same TMS, while S377 interacts with T373 not only through backbone, but in addition it forms side chain-backbone H-bond involving the hydroxyl group of S377.

Table 3.4.1.1: The RMSD of the 377 neighboring Ca residues.

	XanQ wt	XanQ-377G
<b>RMSD</b> resid near 377		
resid 373 to 381 average	1.1 Å	1.6 Å
resid 373 to 381 max	1.8 Å	2.3 Å

Further, an interaction between F376 and F94 seems to lead an important role. Specifically, F94 lays in the middle of the substrate translocation pathway and interacts directly with the substrate (**Figure 3.4.2.7A**). This finding implies a probable gating role for F94 which seems to interact with F376, albeit with a different way in the mutant and in the wild type (**Figure 3.4.2.7D**). The F376 phenyl group was monitored during the MD trajectory and was found to have minor fluctuations, being relatively rigid in both wild-type and mutant. On the other hand, the F94 phenyl fluctuates between different conformations. Consequently, F94 leads the different type of interaction with F376. In wild-type XanQ, the dihedrals obtained between F94-F376 in the 80% of the trajectory correspond to a well-preserved T-stacking interaction, while the second most frequent F94 phenyl conformation is not valid for any interaction with F376 and represents only 3% of the trajectory (**Figure 3.4.2.7D**), upper panel). On the contrary, in 377G-XanQ simulation, the 88% of the trajectory corresponds to a non-interacting or weakly interacting setup of the two phenyl groups (**Figure 3.4.2.7D**, lower panel), although the distance between these phenyl groups appears almost the same in the mutant and the wild type (Supplement, **Figure 3.4.2.9**). In both wild-type and XanQ-377G, Q324 rotates in order to form H-bond interactions with TMS10 backbone.

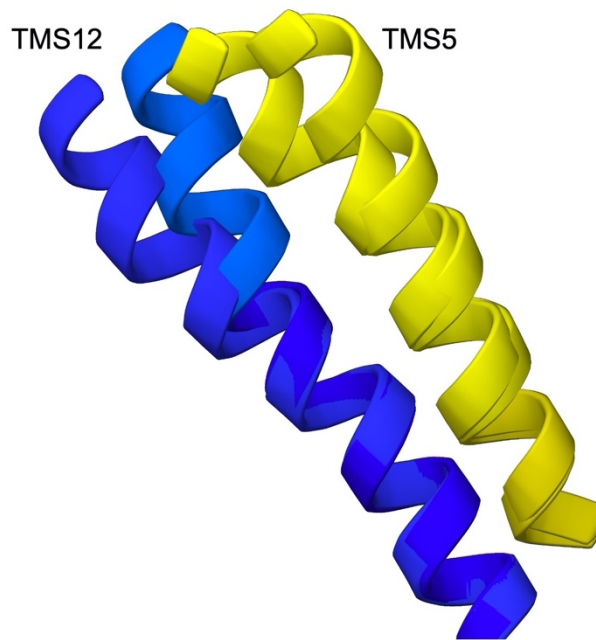




**Figure 3.4.2.9:** Fluctuation of the distance between F376-F94 in wild-type XanQ and XanQ-S377G. The distance between the C $\gamma$  atoms of F94 and F376 was monitored in wild-type XanQ (blue) and S377G XanQ (red), during a 500-ns MD simulation. The MD simulation was run in triplicate, as shown, and the means and SD values are indicated next to each MD run. Taking into account all runs, the mean ( $\pm$ SD) is 7.2 ( $\pm$ 0.1) for wild type ( $n=3$ ) and 7.2 ( $\pm$ 0.1) for the S377G mutant ( $n=3$ ).

UapA (xanthine/uric acid transporter) corresponding residues to F376, S377 and F94 of XanQ (xanthine-specific transporter) are F462, S463 and F155.

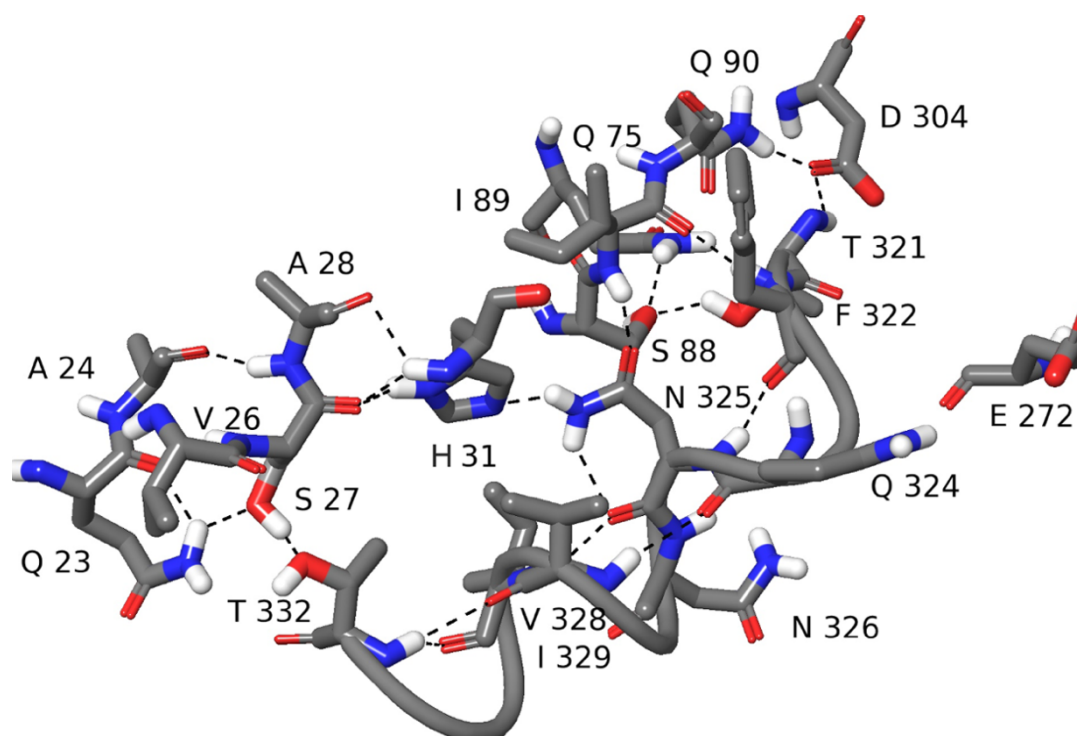
This bending of TMS12 affects also the neighboring TMSs. Explicitly, TMS5, another gate TMS at the interface of the dimer, which is located next to TMS12 moves towards TMS13-14 (**Figure 3.4.2.10**). TMS5, along with TMS12 is opposite to the binding site, implying that the translocation of both these helices, changes the environment in the vicinity of the binding site. This change could have an important mechanistic and steric impact on the binding of various purine molecules to the binding site, affecting specificity.



**Figure 3.4.2.10:** TMS12 (blue) bending affects the neighboring helix TMS5 (yellow) which also moves towards TMS13-14.

Concerning the potential epistatic effect of the mutation G27S on the S377G phenotype, our MD analysis indicates that by introducing a Ser in the position of G27 in TMS1, two more H-bond interactions occur, involving T332 (TMS10) and H31 (TMS1). These H-bonds have a key role as part of an extensive H-bond network in the XanQ core domain, which includes interactions between D304 (TMS9), Q90 (TMS3), S88 (TMS3), Q75 (TMS2), H31 (TMS1), N325 (TMS10) and T321 (TMS10) (**Figure 3.4.2.11**). This H-bond network has been proposed as crucial for the binding and transport of substrate, based on the effect of mutations of H31-N325 and Gln75-D304. The additional H-bonds between S27 and T332 and between S27 and H31 might affect the positioning of F322 and subsequently Q324 of the binding site. The rearrangement of the binding site residues could result in restoration of the XanQ-377 incapability to transport guanine. From another point of view, F322 is also interacting with

F94, controlling the fluctuation of the latter, and consequently the F376-F94 hydrophobic interaction which was shown above to regulate specificity.



**Figure 3.4.2.11:** Extensive H-bond network in XanQ (G27S) linking different parts of the core domain. Core residue S27 (in XanQ-G27S) is linked with F322 and Q324 in the binding site through H31 and N325; S27 reinforces the H-bond network by two additional H bonds (a) between its main-chain C=O and H31; (b) between its side-chain -OH and T332, leading to repositioning of Q324 and F322 in the binding pocket. This change might explain the epistatic effect of G27S on S377G (see text for details).

### 3.4.2.7 Conclusion

The overall *in silico* and experimental procedure of this study demonstrate that changes outside the binding region of XanQ can affect specificity drastically, as shown with the S377G mutant. In details, our MD studies on AncXanQ, indicate that guanine (but also hypoxanthine or adenine) is oriented differently in the substrate binding pocket compared to xanthine, even though the the same residues confere the key binding interactions. This difference may explain the result that guanine and xanthine inhibit the transport of one another with one to two orders of magnitude lower apparent affinity than the affinity values in the transport kinetics assays. There is no previous relevant evidence from any structurally solved transporter for the binding mode of guanine, in any transporter family. Crystal structures of *E. coli* enzymes associated with guanine typically show bidentate H-bonds of N1H and C2-NH2 or C2-NH2 and N3 with a Glu or a backbone carboxylate, as in GuaD, XapA and XGPRTase. Nevertheless, guanine

and xanthine binding poses are the same with regard to interactions and orientations, in enzymes that utilize both these substrates, like XapA or XGPRTase, unlike the case with transporters of guanine and xanthine presented in this thesis.

Moreover, all permeases and mutants of the XanQ clade analyzed experimentally in this study, retain the Asn93 residue and, independently of their substrate profile, are high-affinity transporters for xanthine but do not recognize uric acid. Thus, our new data strengthen the evidence presented in the previous section, that Asn93 is important for the specificity for xanthine.

Also, our data highlight S377 (TMS12) as an additional important xanthine-specificity-controlling residue of XanQ transporters. The specificity effect of S377G does not hold in all XanQ backgrounds, but coexistence of S377G with G27S or S312T or both (in mutant XanQ-5M) abolishes the specificity effect and reestablishes the xanthine-specific profile of XanQ. In contrast, S377G phenotype is not affected by the coexistence with G191S. The epistatic effect of G27S (TMS1) might be attributed to the extensive H-bond network that stabilizes the core domain. The additional interactions offered by G27S, result in repositioning of Q324 in the binding pocket, which may lead to restoration of the specificity constraints, like the abolishment of the binding of guanine and guanine-related analogs, in the S377G mutant. Such an epistatic effect is not possible with G191S (TMS6), a gate-domain residue which does not participate in a similar H-bond network.

Further, F376 contributes to substrate coordination in the inward-facing conformation of the transporter, as shown by both the structural data on UapA and UraA, and the MD studies on XanQ. Mutations that retain the aromatic rings, like F94Y, F376Y, in XanQ do not inhibit xanthine binding and transport, but replacement of one of the two phenyl rings leads to loss of 80–90% (F94C, F376C) or 40–60% of activity (F94I, F376L). F94Y, F94I36 and F376L37 show changes in specificity. The specificity towards binding and uptake of thymine in UraA and other related uracil transporters, is enlarged in the case that the equivalent Phe is mutated to Ala. Based on all the above, one can safely assume that the F376-Phe94 deregulation effect is correlated to the enlarged specificity of the S377G mutant. Although it is considered here that S377G has a role in specificity by causing an indirect effect affecting the binding site, we cannot exclude the possibility that S377G could also modify the dynamics of transport by uncoupling the sliding of the core domain upon the gate. A similar theory was recently proposed for its fungal NAT/NCS2 homologs. Such an effect is difficult to demonstrate with regular MD calculations but can clearly be a possibility. This concept is supported by two factors. First, S377 in XanQ is in the middle of TMS12, a helix of the gate domain that forms

part of the interface on which the core/elevator/binding-site slides, upon substrate loading in the binding site. It is reasonable to assume that, by tilting TMS12, S377G affects this elevator-like motion of the core domain and modulates the kinetics of transport. Several specific mutations at the core-gate interface of UapA, a highly selective xanthine and uric acid transporter, (e.g. mutations in TMS12, in positions corresponding to M375, I379, R385 in XanQ), can lead to enlarged specificity allowing recognition and/or transport of most purines and uracil. The effects were proposed to reflect changes in the dynamics of the elevator sliding, specifically the sliding of the core domain which includes the binding site against the gate domain, rather than changes in residues of the binding site per se. S377G in XanQ might also affect specificity via a similar route. Second, S377G does not affect the affinity for xanthine or most xanthine analogs, implying that it does not change the overall binding site topology, but displays high affinity for guanine and other non-physiological ligands, such as hypoxanthine and guanine or hypoxanthine analogs. This leads to the conclusion that a modulation of geometry or interactions in the binding site alone would not be sufficient to change the affinity for guanine and other analogs from not being recognized at the mM level (wild-type) to low-micromolar or sub-micromolar affinity (S377G). On the other hand, the effects observed with ligands other than guanine were only on recognition/binding ( $K_i$  values) and not on transport (S377G does not transport the other non-canonical ligands), implying that a component of the S377G effect is certainly the modulation of the binding affinity and binding-site interactions. Summarizing, it is possible that S377G can affect both the binding of guanine and other non-canonical substrates and the dynamics of transport. As it was proposed also for UapA, non-canonical substrates like guanine might also be loaded to the binding site of XanQ, but due to different orientation and/or weaker interactions might not release the sliding of the elevator/binding-site to the inward conformation. In the S377G mutant, uncoupling of the elevator from substrate loading may allow translocation of guanine because of the loosened sliding.

## References

- (1) Papageorgiou, I.; Gournas, C.; Vlantı, A.; Amillis, S.; Pantazopoulou, A.; Dıallinas, G. Specific Interdomain Synergy in the UapA Transporter Determines Its Unique Specificity for Uric Acid among NAT Carriers. *J. Mol. Biol.* **2008**, *382* (5), 1121–1135. <https://doi.org/10.1016/j.jmb.2008.08.005>.
- (2) Goudela, S.; Karatza, P.; Koukaki, M.; Frillingos, S.; Dıallinas, G. Comparative Substrate Recognition by Bacterial and Fungal Purine Transporters of the NAT/NCS2 Family. *Mol. Membr. Biol.* **2005**, *22* (3), 263–275. <https://doi.org/10.1080/09687860500093016>.
- (3) Gorfinkiel, L.; Dıallinas, G.; Scazzocchio, C. Sequence and Regulation of the UapA Gene Encoding a Uric Acid-Xanthine Permease in the Fungus *Aspergillus Nidulans*. *J. Biol. Chem.* **1993**, *268* (31), 23376–23381. [https://doi.org/10.1016/S0021-9258\(19\)49473-8](https://doi.org/10.1016/S0021-9258(19)49473-8).
- (4) Sonders, M. S.; Quick, M.; Javitch, J. A. How Did the Neurotransmitter Cross the Bilayer? A Closer View. *Curr. Opin. Neurobiol.* **2005**, *15* (3), 296–304. <https://doi.org/10.1016/j.conb.2005.05.009>.
- (5) Wright, E. M.; Turk, E. The Sodium/Glucose Cotransport Family SLC5. *Pflugers Arch.* **2004**, *447* (5), 510–518. <https://doi.org/10.1007/s00424-003-1063-6>.
- (6) Dıallinas, G.; Scazzocchio, C. A Gene Coding for the Uric Acid-Xanthine Permease of *Aspergillus Nidulans*: Inactivational Cloning, Characterization, and Sequence of a Cis-Acting Mutation. *Genetics* **1989**, *122* (2), 341–350.
- (7) Yamamoto, S.; Inoue, K.; Murata, T.; Kamigaso, S.; Yasujima, T.; Maeda, J.; Yoshida, Y.; Ohta, K.; Yuasa, H. Identification and Functional Characterization of the First Nucleobase Transporter in Mammals: IMPLICATION IN THE SPECIES DIFFERENCE IN THE INTESTINAL ABSORPTION MECHANISM OF NUCLEOBASES AND THEIR ANALOGS BETWEEN HIGHER PRIMATES AND OTHER MAMMALS\*. *J. Biol. Chem.* **2010**, *285* (9), 6522–6531. <https://doi.org/10.1074/jbc.M109.032961>.
- (8) Kosti, V.; Papageorgiou, I.; Dıallinas, G. Dynamic Elements at Both Cytoplasmically and Extracellularly Facing Sides of the UapA Transporter Selectively Control the Accessibility of Substrates to Their Translocation Pathway. *J. Mol. Biol.* **2010**, *397* (5), 1132–1143. <https://doi.org/10.1016/j.jmb.2010.02.037>.
- (9) Saier, M. H.; Reddy, V. S.; Tamang, D. G.; Västermark, A. The Transporter Classification Database. *Nucleic Acids Res.* **2014**, *42* (Database issue), D251-258. <https://doi.org/10.1093/nar/gkt1097>.
- (10) Wang, S. C.; Davejan, P.; Hendargo, K. J.; Javadi-Razaz, I.; Chou, A.; Yee, D. C.; Ghazi, F.; Lam, K. J. K.; Conn, A. M.; Madrigal, A.; Medrano-Soto, A.; Saier, M. H. Expansion of the Major Facilitator Superfamily (MFS) to Include Novel Transporters as Well as Transmembrane-Acting Enzymes. *Biochim. Biophys. Acta Biomembr.* **2020**, *1862* (9), 183277. <https://doi.org/10.1016/j.bbamem.2020.183277>.
- (11) Yan, N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu. Rev. Biophys.* **2015**, *44* (1), 257–283. <https://doi.org/10.1146/annurev-biophys-060414-033901>.
- (12) Pao, S. S.; Paulsen, I. T.; Saier, M. H. Major Facilitator Superfamily. *Microbiol. Mol. Biol. Rev. MMBR* **1998**, *62* (1), 1–34. <https://doi.org/10.1128/MMBR.62.1.1-34.1998>.

- (13) Marger, M. D.; Saier, M. H. A Major Superfamily of Transmembrane Facilitators That Catalyze Uniport, Symport and Antiport. *Trends Biochem. Sci.* **1993**, *18* (1), 13–20. [https://doi.org/10.1016/0968-0004\(93\)90081-w](https://doi.org/10.1016/0968-0004(93)90081-w).
- (14) Abramson, J.; Smirnova, I.; Kasho, V.; Verner, G.; Kaback, H. R.; Iwata, S. Structure and Mechanism of the Lactose Permease of Escherichia Coli. *Science* **2003**, *301* (5633), 610–615. <https://doi.org/10.1126/science.1088196>.
- (15) Huang, Y.; Lemieux, M. J.; Song, J.; Auer, M.; Wang, D.-N. Structure and Mechanism of the Glycerol-3-Phosphate Transporter from Escherichia Coli. *Science* **2003**, *301* (5633), 616–620. <https://doi.org/10.1126/science.1087619>.
- (16) Yin, Y.; He, X.; Szewczyk, P.; Nguyen, T.; Chang, G. Structure of the Multidrug Transporter EmrD from Escherichia Coli. *Science* **2006**, *312* (5774), 741–744. <https://doi.org/10.1126/science.1125629>.
- (17) Radestock, S.; Forrest, L. R. The Alternating-Access Mechanism of MFS Transporters Arises from Inverted-Topology Repeats. *J. Mol. Biol.* **2011**, *407* (5), 698–715. <https://doi.org/10.1016/j.jmb.2011.02.008>.
- (18) Sun, L.; Zeng, X.; Yan, C.; Sun, X.; Gong, X.; Rao, Y.; Yan, N. Crystal Structure of a Bacterial Homologue of Glucose Transporters GLUT1-4. *Nature* **2012**, *490* (7420), 361–366. <https://doi.org/10.1038/nature11524>.
- (19) Pedersen, B. P.; Kumar, H.; Waight, A. B.; Risenmay, A. J.; Roe-Zurz, Z.; Chau, B. H.; Schlessinger, A.; Bonomi, M.; Harries, W.; Sali, A.; Johri, A. K.; Stroud, R. M. Crystal Structure of a Eukaryotic Phosphate Transporter. *Nature* **2013**, *496* (7446), 533–536. <https://doi.org/10.1038/nature12042>.
- (20) Sun, J.; Bankston, J. R.; Payandeh, J.; Hinds, T. R.; Zagotta, W. N.; Zheng, N. Crystal Structure of the Plant Dual-Affinity Nitrate Transporter NRT1.1. *Nature* **2014**, *507* (7490), 73–77. <https://doi.org/10.1038/nature13074>.
- (21) Deng, D.; Xu, C.; Sun, P.; Wu, J.; Yan, C.; Hu, M.; Yan, N. Crystal Structure of the Human Glucose Transporter GLUT1. *Nature* **2014**, *510* (7503), 121–125. <https://doi.org/10.1038/nature13306>.
- (22) Niño-González, M.; Novo-Uzal, E.; Richardson, D. N.; Barros, P. M.; Duque, P. More Transporters, More Substrates: The Arabidopsis Major Facilitator Superfamily Revisited. *Mol. Plant* **2019**, *12* (9), 1182–1202. <https://doi.org/10.1016/j.molp.2019.07.003>.
- (23) Vishwakarma, P.; Banerjee, A.; Pasrija, R.; Prasad, R.; Lynn, A. Phylogenetic and Conservation Analyses of MFS Transporters. *3 Biotech* **2018**. <https://doi.org/10.1007/s13205-018-1476-8>.
- (24) Jack, D. L.; Paulsen, I. T.; Saier, M. H. The Amino Acid/Polyamine/Organocation (APC) Superfamily of Transporters Specific for Amino Acids, Polyamines and Organocations. *Microbiol. Read. Engl.* **2000**, *146* ( Pt 8), 1797–1814. <https://doi.org/10.1099/00221287-146-8-1797>.
- (25) Västermark, Å.; Saier, M. H. Evolutionary Relationship between 5+5 and 7+7 Inverted Repeat Folds within the Amino Acid-Polyamine-Organocation Superfamily: Evolutionary Relationship of Inverted Repeat Folds. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (2), 336–346. <https://doi.org/10.1002/prot.24401>.
- (26) Wong, F. H.; Chen, J. S.; Reddy, V.; Day, J. L.; Shlykov, M. A.; Wakabayashi, S. T.; Saier, Jr., M. H. The Amino Acid-Polyamine-Organocation Superfamily. *J. Mol. Microbiol. Biotechnol.* **2012**, *22* (2), 105–113. <https://doi.org/10.1159/000338542>.

- (27) Vastermark, A.; Wollwage, S.; Houle, M. E.; Rio, R.; Saier, M. H. Expansion of the APC Superfamily of Secondary Carriers. *Proteins* **2014**, *82* (10), 2797–2811. <https://doi.org/10.1002/prot.24643>.
- (28) Yamashita, A.; Singh, S. K.; Kawate, T.; Jin, Y.; Gouaux, E. Crystal Structure of a Bacterial Homologue of Na<sup>+</sup>/Cl<sup>-</sup>-Dependent Neurotransmitter Transporters. *Nature* **2005**, *437* (7056), 215–223. <https://doi.org/10.1038/nature03978>.
- (29) Masson, J.; Sagné, C.; Hamon, M.; El Mestikawy, S. Neurotransmitter Transporters in the Central Nervous System. *Pharmacol. Rev.* **1999**, *51* (3), 439–464.
- (30) Nelson, N. The Family of Na<sup>+</sup>/Cl<sup>-</sup> Neurotransmitter Transporters. *J. Neurochem.* **1998**, *71* (5), 1785–1803. <https://doi.org/10.1046/j.1471-4159.1998.71051785.x>.
- (31) Amara, S. G.; Sonders, M. S. Neurotransmitter Transporters as Molecular Targets for Addictive Drugs. *Drug Alcohol Depend.* **1998**, *51* (1–2), 87–96. [https://doi.org/10.1016/s0376-8716\(98\)00068-4](https://doi.org/10.1016/s0376-8716(98)00068-4).
- (32) Hahn, M. K.; Blakely, R. D. Monoamine Transporter Gene Structure and Polymorphisms in Relation to Psychiatric and Other Complex Disorders. *Pharmacogenomics J.* **2002**, *2* (4), 217–235. <https://doi.org/10.1038/sj.tpj.6500106>.
- (33) Krogsgaard-Larsen, P.; Frølund, B.; Frydenvang, K. GABA Uptake Inhibitors. Design, Molecular Pharmacology and Therapeutic Aspects. *Curr. Pharm. Des.* **2000**, *6* (12), 1193–1209. <https://doi.org/10.2174/1381612003399608>.
- (34) Ozaki, N.; Goldman, D.; Kaye, W. H.; Plotnicov, K.; Greenberg, B. D.; Lappalainen, J.; Rudnick, G.; Murphy, D. L. Serotonin Transporter Missense Mutation Associated with a Complex Neuropsychiatric Phenotype. *Mol. Psychiatry* **2003**, *8* (11), 933–936. <https://doi.org/10.1038/sj.mp.4001365>.
- (35) Richerson, G. B.; Wu, Y. Role of the GABA Transporter in Epilepsy. In *Recent Advances in Epilepsy Research*; Binder, D. K., Scharfman, H. E., Eds.; Advances in Experimental Medicine and Biology; Springer US: Boston, MA, 2004; pp 76–91. [https://doi.org/10.1007/978-1-4757-6376-8\\_6](https://doi.org/10.1007/978-1-4757-6376-8_6).
- (36) Mazei-Robinson, M. S.; Blakely, R. D. ADHD and the Dopamine Transporter: Are There Reasons to Pay Attention? *Handb. Exp. Pharmacol.* **2006**, No. 175, 373–415. [https://doi.org/10.1007/3-540-29784-7\\_17](https://doi.org/10.1007/3-540-29784-7_17).
- (37) Singh, S. K.; Pal, A. Biophysical Approaches to the Study of LeuT, a Prokaryotic Homolog of Neurotransmitter Sodium Symporters. *Methods Enzymol.* **2015**, *557*, 167–198. <https://doi.org/10.1016/bs.mie.2015.01.002>.
- (38) Penmatsa, A.; Wang, K. H.; Gouaux, E. X-Ray Structure of Dopamine Transporter Elucidates Antidepressant Mechanism. *Nature* **2013**, *503* (7474), 85–90. <https://doi.org/10.1038/nature12533>.
- (39) Sitte, H. H.; Farhan, H.; Javitch, J. A. Sodium-Dependent Neurotransmitter Transporters: Oligomerization as a Determinant of Transporter Function and Trafficking. *Mol. Interv.* **2004**, *4* (1), 38–47. <https://doi.org/10.1124/mi.4.1.38>.
- (40) Torres, G. E.; Carneiro, A.; Seamans, K.; Fiorentini, C.; Sweeney, A.; Yao, W.-D.; Caron, M. G. Oligomerization and Trafficking of the Human Dopamine Transporter. Mutational Analysis Identifies Critical Domains Important for the Functional Expression of the Transporter. *J. Biol. Chem.* **2003**, *278* (4), 2731–2739. <https://doi.org/10.1074/jbc.M201926200>.
- (41) Hamilton, P. J.; Belovich, A. N.; Khelashvili, G.; Saunders, C.; Erreger, K.; Javitch, J. A.; Sitte, H. H.; Weinstein, H.; Matthies, H. J. G.; Galli, A. PIP2 Regulates Psychostimulant



- Behaviors through Its Interaction with a Membrane Protein. *Nat. Chem. Biol.* **2014**, *10* (7), 582–589. <https://doi.org/10.1038/nchembio.1545>.
- (42) Khelashvili, G.; Stanley, N.; Sahai, M. A.; Medina, J.; LeVine, M. V.; Shi, L.; De Fabritiis, G.; Weinstein, H. Spontaneous Inward Opening of the Dopamine Transporter Is Triggered by PIP2-Regulated Dynamics of the N-Terminus. *ACS Chem. Neurosci.* **2015**, *6* (11), 1825–1837. <https://doi.org/10.1021/acschemneuro.5b00179>.
- (43) Razavi, A. M.; Khelashvili, G.; Weinstein, H. How Structural Elements Evolving from Bacterial to Human SLC6 Transporters Enabled New Functional Properties. *BMC Biol.* **2018**, *16* (1), 31. <https://doi.org/10.1186/s12915-018-0495-6>.
- (44) Coleman, J. A.; Green, E. M.; Gouaux, E. X-Ray Structures and Mechanism of the Human Serotonin Transporter. *Nature* **2016**, *532* (7599), 334–339. <https://doi.org/10.1038/nature17629>.
- (45) Kilic, F.; Rudnick, G. Oligomerization of Serotonin Transporter and Its Functional Consequences. *Proc. Natl. Acad. Sci.* **2000**, *97* (7), 3106–3111. <https://doi.org/10.1073/pnas.97.7.3106>.
- (46) Dhalluin, G.; Gournas, C. The Ubiquitous Nucleobase-Ascorbate Transporter (NAT) Family: Lessons from Model Microbial Genetic Systems. *Channels* **2008**, *2* (5), 363–372. <https://doi.org/10.4161/chan.2.5.6902>.
- (47) Gournas, C.; Papageorgiou, I.; Dhalluin, G. The Nucleobase–Ascorbate Transporter (NAT) Family: Genomics, Evolution, Structure–Function Relationships and Physiological Role. *Mol. Biosyst.* **2008**, *4* (5), 404–416. <https://doi.org/10.1039/B719777B>.
- (48) Huynh, K. W.; Jiang, J.; Abuladze, N.; Tsurulnikov, K.; Kao, L.; Shao, X.; Newman, D.; Azimov, R.; Pushkin, A.; Zhou, Z. H.; Kurtz, I. CryoEM Structure of the Human SLC4A4 Sodium-Coupled Acid-Base Transporter NBCe1. *Nat. Commun.* **2018**, *9* (1), 900. <https://doi.org/10.1038/s41467-018-03271-3>.
- (49) Coudray, N.; Seyler, S.; Lasala, R.; Zhang, Z.; Clark, K. M.; Dumont, M. E.; Rohou, A.; Beckstein, O.; Stokes, D. L. Structure of the SLC4 Transporter Bor1p in an Inward-Facing Conformation. *Protein Sci. Publ. Protein Soc.* **2017**, *26* (1), 130–145. <https://doi.org/10.1002/pro.3061>.
- (50) Thurtle-Schmidt, B. H.; Stroud, R. M. Structure of Bor1 Supports an Elevator Transport Mechanism for SLC4 Anion Exchangers. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (38), 10542–10546. <https://doi.org/10.1073/pnas.1612603113>.
- (51) Arakawa, T.; Kobayashi-Yurugi, T.; Alguet, Y.; Iwanari, H.; Hatae, H.; Iwata, M.; Abe, Y.; Hino, T.; Ikeda-Suno, C.; Kuma, H.; Kang, D.; Murata, T.; Hamakubo, T.; Cameron, A. D.; Kobayashi, T.; Hamasaki, N.; Iwata, S. Crystal Structure of the Anion Exchanger Domain of Human Erythrocyte Band 3. *Science* **2015**, *350* (6261), 680–684. <https://doi.org/10.1126/science.aaa4335>.
- (52) Chi, X.; Jin, X.; Chen, Y.; Lu, X.; Tu, X.; Li, X.; Zhang, Y.; Lei, J.; Huang, J.; Huang, Z.; Zhou, Q.; Pan, X. Structural Insights into the Gating Mechanism of Human SLC26A9 Mediated by Its C-Terminal Sequence. *Cell Discov.* **2020**, *6*, 55. <https://doi.org/10.1038/s41421-020-00193-7>.
- (53) Walter, J. D.; Sawicka, M.; Dutzler, R. Cryo-EM Structures and Functional Characterization of Murine Slc26a9 Reveal Mechanism of Uncoupled Chloride Transport. *eLife* **2019**, *8*, e46986. <https://doi.org/10.7554/eLife.46986>.
- (54) Wang, C.; Sun, B.; Zhang, X.; Huang, X.; Zhang, M.; Guo, H.; Chen, X.; Huang, F.; Chen, T.; Mi, H.; Yu, F.; Liu, L.-N.; Zhang, P. Structural Mechanism of the Active Bicarbonate

- Transporter from Cyanobacteria. *Nat. Plants* **2019**, 5 (11), 1184–1193.  
<https://doi.org/10.1038/s41477-019-0538-1>.
- (55) Geertsma, E. R.; Chang, Y.-N.; Shaik, F. R.; Neldner, Y.; Pardon, E.; Steyaert, J.; Dutzler, R. Structure of a Prokaryotic Fumarate Transporter Reveals the Architecture of the SLC26 Family. *Nat. Struct. Mol. Biol.* **2015**, 22 (10), 803–808.  
<https://doi.org/10.1038/nsmb.3091>.
- (56) Lu, F.; Li, S.; Jiang, Y.; Jiang, J.; Fan, H.; Lu, G.; Deng, D.; Dang, S.; Zhang, X.; Wang, J.; Yan, N. Structure and Mechanism of the Uracil Transporter UraA. *Nature* **2011**, 472 (7342), 243–246. <https://doi.org/10.1038/nature09885>.
- (57) Yu, X.; Yang, G.; Yan, C.; Baylon, J. L.; Jiang, J.; Fan, H.; Lu, G.; Hasegawa, K.; Okumura, H.; Wang, T.; Tajkhorshid, E.; Li, S.; Yan, N. Dimeric Structure of the Uracil:Proton Symporter UraA Provides Mechanistic Insights into the SLC4/23/26 Transporters. *Cell Res.* **2017**, 27 (8), 1020–1033. <https://doi.org/10.1038/cr.2017.83>.
- (58) Alguel, Y.; Amillis, S.; Leung, J.; Lambrinidis, G.; Capaldi, S.; Scull, N. J.; Craven, G.; Iwata, S.; Armstrong, A.; Mikros, E.; Diallinas, G.; Cameron, A. D.; Byrne, B. Structure of Eukaryotic Purine/H<sup>+</sup> Symporter UapA Suggests a Role for Homodimerization in Transport Activity. *Nat. Commun.* **2016**, 7 (1), 11336.  
<https://doi.org/10.1038/ncomms11336>.
- (59) Martzoukou, O.; Karachaliou, M.; Yalelis, V.; Leung, J.; Byrne, B.; Amillis, S.; Diallinas, G. Oligomerization of the UapA Purine Transporter Is Critical for ER-Exit, Plasma Membrane Localization and Turnover. *J. Mol. Biol.* **2015**, 427 (16), 2679–2696.  
<https://doi.org/10.1016/j.jmb.2015.05.021>.
- (60) Shi, Y. Common Folds and Transport Mechanisms of Secondary Active Transporters. *Annu. Rev. Biophys.* **2013**, 42 (1), 51–72. <https://doi.org/10.1146/annurev-biophys-083012-130429>.
- (61) Drew, D.; Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev. Biochem.* **2016**, 85 (1), 543–572. <https://doi.org/10.1146/annurev-biochem-060815-014520>.
- (62) Jardetzky, O. Simple Allosteric Model for Membrane Pumps. *Nature* **1966**, 211 (5052), 969–970. <https://doi.org/10.1038/211969a0>.
- (63) Diallinas, G. Understanding Transporter Specificity and the Discrete Appearance of Channel-like Gating Domains in Transporters. *Front. Pharmacol.* **2014**, 5.  
<https://doi.org/10.3389/fphar.2014.00207>.
- (64) Karpowich, N. K.; Wang, D.-N. Structural Biology. Symmetric Transporters for Asymmetric Transport. *Science* **2008**, 321 (5890), 781–782.  
<https://doi.org/10.1126/science.1161495>.
- (65) Yan, N. Structural Advances for the Major Facilitator Superfamily (MFS) Transporters. *Trends Biochem. Sci.* **2013**, 38 (3), 151–159.  
<https://doi.org/10.1016/j.tibs.2013.01.003>.
- (66) Mitchell, P. A General Theory of Membrane Transport From Studies of Bacteria. *Nature* **1957**, 180 (4577), 134–136. <https://doi.org/10.1038/180134a0>.
- (67) Mitchell, P. Osmochemistry of Solute Translocation. *Res. Microbiol.* **1990**, 141 (3), 286–289. [https://doi.org/10.1016/0923-2508\(90\)90002-8](https://doi.org/10.1016/0923-2508(90)90002-8).
- (68) Lee, Y.; Nishizawa, T.; Yamashita, K.; Ishitani, R.; Nureki, O. Structural Basis for the Facilitative Diffusion Mechanism by SemiSWEET Transporter. *Nat. Commun.* **2015**, 6 (1), 6112. <https://doi.org/10.1038/ncomms7112>.

- (69) Wang, J.; Yan, C.; Li, Y.; Hirata, K.; Yamamoto, M.; Yan, N.; Hu, Q. Crystal Structure of a Bacterial Homologue of SWEET Transporters. *Cell Res.* **2014**, *24* (12), 1486–1489. <https://doi.org/10.1038/cr.2014.144>.
- (70) Xu, Y.; Tao, Y.; Cheung, L. S.; Fan, C.; Chen, L.-Q.; Xu, S.; Perry, K.; Frommer, W. B.; Feng, L. Structures of Bacterial Homologues of SWEET Transporters in Two Distinct Conformations. *Nature* **2014**, *515* (7527), 448–452. <https://doi.org/10.1038/nature13670>.
- (71) Stelzl, L. S.; Fowler, P. W.; Sansom, M. S. P.; Beckstein, O. Flexible Gates Generate Occluded Intermediates in the Transport Cycle of LacY. *J. Mol. Biol.* **2014**, *426* (3), 735–751. <https://doi.org/10.1016/j.jmb.2013.10.024>.
- (72) Fowler, P. W.; Orwick-Rydmark, M.; Radestock, S.; Solcan, N.; Dijkman, P. M.; Lyons, J. A.; Kwok, J.; Caffrey, M.; Watts, A.; Forrest, L. R.; Newstead, S. Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Struct. England1993* **2015**, *23* (2), 290–301. <https://doi.org/10.1016/j.str.2014.12.012>.
- (73) Andersson, M.; Bondar, A.-N.; Freitas, J. A.; Tobias, D. J.; Kaback, H. R.; White, S. H. Proton-Coupled Dynamics in Lactose Permease. *Structure* **2012**, *20* (11), 1893–1904. <https://doi.org/10.1016/j.str.2012.08.021>.
- (74) Law, C. J.; Almqvist, J.; Bernstein, A.; Goetz, R. M.; Huang, Y.; Soudant, C.; Laaksonen, A.; Hovmöller, S.; Wang, D.-N. Salt-Bridge Dynamics Control Substrate-Induced Conformational Change in the Membrane Transporter GlpT. *J. Mol. Biol.* **2008**, *378* (4), 828–839. <https://doi.org/10.1016/j.jmb.2008.03.029>.
- (75) Garaeva, A. A.; Slotboom, D. J. Elevator-Type Mechanisms of Membrane Transport. *Biochem. Soc. Trans.* **2020**, *48* (3), 1227–1241. <https://doi.org/10.1042/BST20200290>.
- (76) Dhalluin, G. Dissection of Transporter Function: From Genetics to Structure. *Trends Genet.* **2016**, *32* (9), 576–590. <https://doi.org/10.1016/j.tig.2016.06.003>.
- (77) Pyle, E.; Kalli, A. C.; Amillis, S.; Hall, Z.; Lau, A. M.; Hanyaloglu, A. C.; Dhalluin, G.; Byrne, B.; Politis, A. Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chem. Biol.* **2018**, *25* (7), 840–848.e4. <https://doi.org/10.1016/j.chembiol.2018.03.011>.
- (78) Kourkoulou, A.; Grevas, P.; Lambrinidis, G.; Pyle, E.; Dionysopoulou, M.; Politis, A.; Mikros, E.; Byrne, B.; Dhalluin, G. Specific Residues in a Purine Transporter Are Critical for Dimerization, ER Exit, and Function. *Genetics* **2019**, *213* (4), 1357–1372. <https://doi.org/10.1534/genetics.119.302566>.
- (79) Byrne, B. It Takes Two to Transport via an Elevator. *Cell Res.* **2017**, *27* (8), 965–966. <https://doi.org/10.1038/cr.2017.89>.
- (80) DeChancie, J.; Shrivastava, I. H.; Bahar, I. The Mechanism of Substrate Release by the Aspartate Transporter GltPh: Insights from Simulations. *Mol. Biosyst.* **2011**, *7* (3), 832–842. <https://doi.org/10.1039/c0mb00175a>.
- (81) Reyes, N.; Ginter, C.; Boudker, O. Transport Mechanism of a Bacterial Homologue of Glutamate Transporters. *Nature* **2009**, *462* (7275), 880–885. <https://doi.org/10.1038/nature08616>.
- (82) Dhalluin, G. Transporter Specificity: A Tale of Loosened Elevator-Sliding. *Trends Biochem. Sci.* **2021**, *46* (9), 708–717. <https://doi.org/10.1016/j.tibs.2021.03.007>.
- (83) Sauer, D. B.; Trebesch, N.; Marden, J. J.; Cocco, N.; Song, J.; Koide, A.; Koide, S.; Tajkhorshid, E.; Wang, D.-N. Structural Basis for the Reaction Cycle of DASS

- Dicarboxylate Transporters. *eLife* **2020**, *9*, e61350.  
<https://doi.org/10.7554/eLife.61350>.
- (84) Duster, A. W.; Lin, H. Riding Elevators into and out of Cells. *eLife* **2020**, *9*, e62925.  
<https://doi.org/10.7554/eLife.62925>.
- (85) Coincon, M.; Uzdavinyas, P.; Nji, E.; Dotson, D. L.; Winkelmann, I.; Abdul-Hussein, S.; Cameron, A. D.; Beckstein, O.; Drew, D. Crystal Structures Reveal the Molecular Basis of Ion Translocation in Sodium/Proton Antiporters. *Nat. Struct. Mol. Biol.* **2016**, *23* (3), 248–255. <https://doi.org/10.1038/nsmb.3164>.
- (86) Forrest, L. R.; Rudnick, G. The Rocking Bundle: A Mechanism for Ion-Coupled Solute Flux by Symmetrical Transporters. *Physiol. Bethesda Md* **2009**, *24*, 377–386.  
<https://doi.org/10.1152/physiol.00030.2009>.
- (87) Weyand, S.; Shimamura, T.; Yajima, S.; Suzuki, S.; Mirza, O.; Krusong, K.; Carpenter, E. P.; Rutherford, N. G.; Hadden, J. M.; O'Reilly, J.; Ma, P.; Saidijam, M.; Patching, S. G.; Hope, R. J.; Norbertczak, H. T.; Roach, P. C. J.; Iwata, S.; Henderson, P. J. F.; Cameron, A. D. Structure and Molecular Mechanism of a Nucleobase-Cation-Symport-1 Family Transporter. *Science* **2008**, *322* (5902), 709–713.  
<https://doi.org/10.1126/science.1164440>.
- (88) Ressler, S.; Terwisscha van Scheltinga, A. C.; Vonrhein, C.; Ott, V.; Ziegler, C. Molecular Basis of Transport and Regulation in the Na(+)/Betaine Symporter BetP. *Nature* **2009**, *458* (7234), 47–52. <https://doi.org/10.1038/nature07819>.
- (89) Faham, S.; Watanabe, A.; Besserer, G. M.; Cascio, D.; Specht, A.; Hirayama, B. A.; Wright, E. M.; Abramson, J. The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na<sup>+</sup>/Sugar Symport. *Science* **2008**, *321* (5890), 810–814. <https://doi.org/10.1126/science.1160406>.
- (90) Shaffer, P. L.; Goehring, A.; Shankaranarayanan, A.; Gouaux, E. Structure and Mechanism of a Na<sup>+</sup>-independent Amino Acid Transporter. *Science* **2009**, *325* (5943), 1010–1014. <https://doi.org/10.1126/science.1176088>.
- (91) Krishnamurthy, H.; Gouaux, E. X-Ray Structures of LeuT in Substrate-Free Outward-Open and Apo Inward-Open States. *Nature* **2012**, *481* (7382), 469–474.  
<https://doi.org/10.1038/nature10737>.
- (92) Shi, L.; Quick, M.; Zhao, Y.; Weinstein, H.; Javitch, J. A. The Mechanism of a Neurotransmitter:Sodium Symporter – Inward Release of Na<sup>+</sup> and Substrate Is Triggered by Substrate in a Second Binding Site. *Mol. Cell* **2008**, *30* (6), 667–677.  
<https://doi.org/10.1016/j.molcel.2008.05.008>.
- (93) Nyola, A.; Karpowich, N. K.; Zhen, J.; Marden, J.; Reith, M. E.; Wang, D.-N. Substrate and Drug Binding Sites in LeuT. *Curr. Opin. Struct. Biol.* **2010**, *20* (4), 415–422.  
<https://doi.org/10.1016/j.sbi.2010.05.007>.
- (94) Papadaki, G. F.; Lambrinidis, G.; Zamanos, A.; Mikros, E.; Diallinas, G. Cytosolic N- and C-Termini of the *Aspergillus Nidulans* FurE Transporter Contain Distinct Elements That Regulate by Long-Range Effects Function and Specificity. *J. Mol. Biol.* **2019**, *431* (19), 3827–3844. <https://doi.org/10.1016/j.jmb.2019.07.013>.
- (95) Kryptou, E.; Evangelidis, T.; Bobonis, J.; Pittis, A. A.; Gabaldón, T.; Scazzocchio, C.; Mikros, E.; Diallinas, G. Origin, Diversification and Substrate Specificity in the Family of NCS1/FUR Transporters: Origin and Specificity of Fur Transporters. *Mol. Microbiol.* **2015**, *96* (5), 927–950. <https://doi.org/10.1111/mmi.12982>.

- (96) Papadaki, G. F.; Amillis, S.; Diallinas, G. Substrate Specificity of the FurE Transporter Is Determined by Cytoplasmic Terminal Domain Interactions. *Genetics* **2017**, *207* (4), 1387–1400. <https://doi.org/10.1534/genetics.117.300327>.
- (97) Birch, J.; Axford, D.; Foadi, J.; Meyer, A.; Eckhardt, A.; Thielmann, Y.; Moraes, I. The Fine Art of Integral Membrane Protein Crystallisation. *Methods San Diego Calif* **2018**, *147*, 150–162. <https://doi.org/10.1016/j.ymeth.2018.05.014>.
- (98) Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* **2001**, *294* (5540), 93–96. <https://doi.org/10.1126/science.1065659>.
- (99) Boyd, C. M.; Bubeck, D. Advances in CryoEM and Its Impact on  $\beta$ -Pore Forming Proteins. *Curr. Opin. Struct. Biol.* **2018**, *52*, 41–49. <https://doi.org/10.1016/j.sbi.2018.07.010>.
- (100) Majumder, P.; Mallela, A. K.; Penmatsa, A. Transporters through the Looking Glass. An Insight into the Mechanisms of Ion-Coupled Transport and Methods That Help Reveal Them. *J. Indian Inst. Sci.* **2018**, *98* (3), 283–300. <https://doi.org/10.1007/s41745-018-0081-5>.
- (101) Dmitriev, O. Y.; Lutsenko, S.; Muyldermans, S. Nanobodies as Probes for Protein Dynamics in Vitro and in Cells. *J. Biol. Chem.* **2016**, *291* (8), 3767–3775. <https://doi.org/10.1074/jbc.R115.679811>.
- (102) Broecker, J.; Eger, B. T.; Ernst, O. P. Crystallography of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. *Struct. Lond. Engl. 1993* **2017**, *25* (2), 384–392. <https://doi.org/10.1016/j.str.2016.12.004>.
- (103) Chen, K.-Y. M.; Sun, J.; Salvo, J. S.; Baker, D.; Barth, P. High-Resolution Modeling of Transmembrane Helical Protein Structures from Distant Homologues. *PLOS Comput. Biol.* **2014**, *10* (5), e1003636. <https://doi.org/10.1371/journal.pcbi.1003636>.
- (104) Zhu, T.; Cao, S.; Su, P.-C.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M. E.; Hevener, K. E. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J. Med. Chem.* **2013**, *56* (17), 6560–6572. <https://doi.org/10.1021/jm301916b>.
- (105) Kaback, H. R. A Chemiosmotic Mechanism of Symport. *Proc. Natl. Acad. Sci.* **2015**, *112* (5), 1259–1264. <https://doi.org/10.1073/pnas.1419325112>.
- (106) Wen, P.-C.; Verhalen, B.; Wilkens, S.; Mchaourab, H. S.; Tajkhorshid, E. On the Origin of Large Flexibility of P-Glycoprotein in the Inward-Facing State. *J. Biol. Chem.* **2013**, *288* (26), 19211–19220. <https://doi.org/10.1074/jbc.M113.450114>.
- (107) Montanari, F.; Ecker, G. F. Prediction of Drug-ABC-Transporter Interaction--Recent Advances and Future Challenges. *Adv. Drug Deliv. Rev.* **2015**, *86*, 17–26. <https://doi.org/10.1016/j.addr.2015.03.001>.
- (108) Rauch, C.; Paine, S. W.; Littlewood, P. Can Long Range Mechanical Interaction between Drugs and Membrane Proteins Define the Notion of Molecular Promiscuity? Application to P-Glycoprotein-Mediated Multidrug Resistance (MDR). *Biochim. Biophys. Acta* **2013**, *1830* (11), 5112–5118. <https://doi.org/10.1016/j.bbagen.2013.06.038>.
- (109) Booth, P. J.; Curnow, P. Folding Scene Investigation: Membrane Proteins. *Curr. Opin. Struct. Biol.* **2009**, *19* (1–3), 8–13. <https://doi.org/10.1016/j.sbi.2008.12.005>.
- (110) Güler, G.; Gärtner, R. M.; Ziegler, C.; Mäntele, W. Lipid-Protein Interactions in the Regulated Betaine Symporter BetP Probed by Infrared Spectroscopy. *J. Biol. Chem.* **2016**, *291* (9), 4295–4307. <https://doi.org/10.1074/jbc.M114.621979>.

- (111) del Alamo, D.; Sala, D.; Mchaourab, H. S.; Meiler, J. Sampling Alternative Conformational States of Transporters and Receptors with AlphaFold2. *eLife* **2022**, *11*, e75751. <https://doi.org/10.7554/eLife.75751>.
- (112) Frillingos, S.; Sahin-Tóth, M.; Wu, J.; Kaback, H. R. Cys-Scanning Mutagenesis: A Novel Approach to Structure Function Relationships in Polytopic Membrane Proteins. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **1998**, *12* (13), 1281–1299. <https://doi.org/10.1096/fasebj.12.13.1281>.
- (113) Frillingos, S. Insights to the Evolution of Nucleobase-Ascorbate Transporters (NAT/NCS2 Family) from the Cys-Scanning Analysis of Xanthine Permease XanQ. *Int. J. Biochem. Mol. Biol.* **2012**, *3* (3), 250–272.
- (114) Sorgen, P. L.; Hu, Y.; Guan, L.; Kaback, H. R.; Girvin, M. E. An Approach to Membrane Protein Structure without Crystals. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (22), 14037–14040. <https://doi.org/10.1073/pnas.182552199>.
- (115) Kaback, H. R.; Smirnova, I.; Kasho, V.; Nie, Y.; Zhou, Y. The Alternating Access Transport Mechanism in LacY. *J. Membr. Biol.* **2011**, *239* (1), 85–93. <https://doi.org/10.1007/s00232-010-9327-5>.
- (116) Sugihara, J.; Smirnova, I.; Kasho, V.; Kaback, H. R. Sugar Recognition by CscB and LacY. *Biochemistry* **2011**, *50* (51), 11009–11014. <https://doi.org/10.1021/bi201592y>.
- (117) Singh, S. K.; Piscitelli, C. L.; Yamashita, A.; Gouaux, E. A Competitive Inhibitor Traps LeuT in an Open-to-out Conformation. *Science* **2008**, *322* (5908), 1655–1661. <https://doi.org/10.1126/science.1166777>.
- (118) Cunningham, P.; Naftalin, R. J. Implications of Aberrant Temperature-Sensitive Glucose Transport via the Glucose Transporter Deficiency Mutant (GLUT1DS) T295M for the Alternate-Access and Fixed-Site Transport Models. *J. Membr. Biol.* **2013**, *246* (6), 495–511. <https://doi.org/10.1007/s00232-013-9564-5>.
- (119) Cunningham, P.; Afzal-Ahmed, I.; Naftalin, R. J. Docking Studies Show That D-Glucose and Quercetin Slide through the Transporter GLUT1 \*. *J. Biol. Chem.* **2006**, *281* (9), 5797–5803. <https://doi.org/10.1074/jbc.M509422200>.
- (120) Simmons, K. J.; Jackson, S. M.; Brueckner, F.; Patching, S. G.; Beckstein, O.; Ivanova, E.; Geng, T.; Weyand, S.; Drew, D.; Lanigan, J.; Sharples, D. J.; Sansom, M. S.; Iwata, S.; Fishwick, C. W.; Johnson, A. P.; Cameron, A. D.; Henderson, P. J. Molecular Mechanism of Ligand Recognition by Membrane Transport Protein, Mhp1. *EMBO J.* **2014**, *33* (16), 1831–1844. <https://doi.org/10.15252/embj.201387557>.
- (121) Smirnova, I.; Kasho, V.; Kaback, H. R. Lactose Permease and the Alternating Access Mechanism. *Biochemistry* **2011**, *50* (45), 9684–9693. <https://doi.org/10.1021/bi2014294>.
- (122) Kumar, H.; Kasho, V.; Smirnova, I.; Finer-Moore, J. S.; Kaback, H. R.; Stroud, R. M. Structure of Sugar-Bound LacY. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (5), 1784–1788. <https://doi.org/10.1073/pnas.1324141111>.
- (123) Wang, K. H.; Penmatsa, A.; Gouaux, E. Neurotransmitter and Psychostimulant Recognition by the Dopamine Transporter. *Nature* **2015**, *521* (7552), 322–327. <https://doi.org/10.1038/nature14431>.
- (124) Gabrielsen, M.; Ravna, A. W.; Kristiansen, K.; Sylte, I. Substrate Binding and Translocation of the Serotonin Transporter Studied by Docking and Molecular Dynamics Simulations. *J. Mol. Model.* **2012**, *18* (3), 1073–1085. <https://doi.org/10.1007/s00894-011-1133-1>.

- (125) Vlanti, A.; Amillis, S.; Koukaki, M.; Diallynas, G. A Novel-Type Substrate-Selectivity Filter and ER-Exit Determinants in the UapA Purine Transporter. *J. Mol. Biol.* **2006**, *357* (3), 808–819. <https://doi.org/10.1016/j.jmb.2005.12.070>.
- (126) Kosti, V.; Lambrinidis, G.; Myrianthopoulos, V.; Diallynas, G.; Mikros, E. Identification of the Substrate Recognition and Transport Pathway in a Eukaryotic Member of the Nucleobase-Ascorbate Transporter (NAT) Family. *PLOS ONE* **2012**, *7* (7), e41939. <https://doi.org/10.1371/journal.pone.0041939>.
- (127) Koukaki, M.; Vlanti, A.; Goudela, S.; Pantazopoulou, A.; Gioule, H.; Tournaviti, S.; Diallynas, G. The Nucleobase-Ascorbate Transporter (NAT) Signature Motif in UapA Defines the Function of the Purine Translocation Pathway. *J. Mol. Biol.* **2005**, *350* (3), 499–513. <https://doi.org/10.1016/j.jmb.2005.04.076>.
- (128) Kourkoulou, A.; Zantza, I.; Foti, K.; Mikros, E.; Diallynas, G. Context-Dependent Cryptic Roles of Specific Residues in Substrate Selectivity of the UapA Purine Transporter. *J. Mol. Biol.* **2021**, 166814. <https://doi.org/10.1016/j.jmb.2021.166814>.
- (129) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>.
- (130) *AMBER parameter database (Bryce Group: Computational Biophysics and Drug Design - University of Manchester)*. <http://amber.manchester.ac.uk/> (accessed 2021-06-09).
- (131) Wang, L.; Berne, B. J.; Friesner, R. A. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding Affinities. *Proc. Natl. Acad. Sci.* **2012**, *109* (6), 1937–1942. <https://doi.org/10.1073/pnas.1114017109>.
- (132) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. <https://doi.org/10.1002/jcc.20035>.
- (133) *CHARMM-GUI: A web-based graphical user interface for CHARMM - Jo - 2008 - Journal of Computational Chemistry - Wiley Online Library*. <https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20945> (accessed 2021-05-20).
- (134) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- (135) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71–73. <https://doi.org/10.1038/nmeth.4067>.
- (136) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Res.* **2012**, *40* (D1), D370–D376. <https://doi.org/10.1093/nar/gkr703>.
- (137) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069–4074. <https://doi.org/10.1063/1.449071>.
- (138) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.

- (139) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (140) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- (141) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- (142) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566. <https://doi.org/10.1073/pnas.202427399>.
- (143) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci.* **2013**, *110* (16), 6358–6363. <https://doi.org/10.1073/pnas.1303186110>.
- (144) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126* (5), 054103. <https://doi.org/10.1063/1.2432340>.
- (145) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* **2020**, *15* (9), 2837–2866. <https://doi.org/10.1038/s41596-020-0342-4>.
- (146) Bussi, G.; Branduardi, D. Free-Energy Calculations with Metadynamics: Theory and Practice. In *Reviews in Computational Chemistry Volume 28*; John Wiley & Sons, Ltd, 2015; pp 1–49. <https://doi.org/10.1002/9781118889886.ch1>.
- (147) Aqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7* (3), 385–391. <https://doi.org/10.1093/protein/7.3.385>.
- (148) Gouda, H.; Kuntz, I. D.; Case, D. A.; Kollman, P. A. Free Energy Calculations for Theophylline Binding to an RNA Aptamer: Comparison of MM-PBSA and Thermodynamic Integration Methods. *Biopolymers* **2003**, *68* (1), 16–34. <https://doi.org/10.1002/bip.10270>.
- (149) Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43* (20), 3786–3791. <https://doi.org/10.1021/jm000241h>.
- (150) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426. <https://doi.org/10.1063/1.1740409>.
- (151) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320* (3), 597–608. [https://doi.org/10.1016/s0022-2836\(02\)00470-9](https://doi.org/10.1016/s0022-2836(02)00470-9).
- (152) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins* **2004**, *55* (2), 351–367. <https://doi.org/10.1002/prot.10613>.
- (153) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>.



- (154) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (155) Zantza, I. Molecular Basis of Uracil/H<sup>+</sup> Symport Mechanism Operated by the FurE/NCS1 Transporter. [], 2022. <https://doi.org/10.1101/2022.03.28.486045>.
- (156) Penmatsa, A.; Gouaux, E. How LeuT Shapes Our Understanding of the Mechanisms of Sodium-Coupled Neurotransmitter Transporters. *J. Physiol.* **2014**, *592* (5), 863–869. <https://doi.org/10.1113/jphysiol.2013.259051>.
- (157) Forrest, L. R.; Zhang, Y.-W.; Jacobs, M. T.; Gesmonde, J.; Xie, L.; Honig, B. H.; Rudnick, G. Mechanism for Alternating Access in Neurotransmitter Transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (30), 10338–10343. <https://doi.org/10.1073/pnas.0804659105>.
- (158) Kazmier, K.; Claxton, D. P.; Mchaourab, H. S. Alternating Access Mechanisms of LeuT-Fold Transporters: Trailblazing towards the Promised Energy Landscapes. *Curr. Opin. Struct. Biol.* **2017**, *45*, 100–108. <https://doi.org/10.1016/j.sbi.2016.12.006>.
- (159) Krishnamurthy, H.; Piscitelli, C. L.; Gouaux, E. Unlocking the Molecular Secrets of Sodium-Coupled Transporters. *Nature* **2009**, *459* (7245), 347–355. <https://doi.org/10.1038/nature08143>.
- (160) Coleman, J. A.; Navratna, V.; Antermite, D.; Yang, D.; Bull, J. A.; Gouaux, E. Chemical and Structural Investigation of the Paroxetine-Human Serotonin Transporter Complex. *eLife* **2020**, *9*, e56427. <https://doi.org/10.7554/eLife.56427>.
- (161) Quick, M.; Abramyan, A. M.; Wiryasermkul, P.; Weinstein, H.; Shi, L.; Javitch, J. A. The LeuT-Fold Neurotransmitter:Sodium Symporter MhsT Has Two Substrate Sites. *Proc. Natl. Acad. Sci.* **2018**, *115* (34), E7924–E7931. <https://doi.org/10.1073/pnas.1717444115>.
- (162) Tang, L.; Bai, L.; Wang, W.; Jiang, T. Crystal Structure of the Carnitine Transporter and Insights into the Antiport Mechanism. *Nat. Struct. Mol. Biol.* **2010**, *17* (4), 492–496. <https://doi.org/10.1038/nsmb.1788>.
- (163) Fang, Y.; Jayaram, H.; Shane, T.; Kolmakova-Partensky, L.; Wu, F.; Williams, C.; Xiong, Y.; Miller, C. Structure of a Prokaryotic Virtual Proton Pump at 3.2 Å Resolution. *Nature* **2009**, *460* (7258), 1040–1043. <https://doi.org/10.1038/nature08201>.
- (164) Weyand, S.; Cameron, A. D.; Shimamura, T.; Beckstein, O.; Sansom, M. S. P.; Iwata, S.; Henderson, P. J. F. *The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3004247/> (accessed 2020-03-26).
- (165) Shimamura, T.; Weyand, S.; Beckstein, O.; Rutherford, N. G.; Hadden, J. M.; Sharples, D.; Sansom, M. S. P.; Iwata, S.; Henderson, P. J. F.; Cameron, A. D. Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1. *Science* **2010**, *328* (5977), 470–473. <https://doi.org/10.1126/science.1186303>.
- (166) Pantazopoulou, A.; Diallinas, G. Fungal Nucleobase Transporters. *FEMS Microbiol. Rev.* **2007**, *31* (6), 657–675. <https://doi.org/10.1111/j.1574-6976.2007.00083.x>.
- (167) Hamari, Z.; Amillis, S.; Drevet, C.; Apostolaki, A.; Vágvölgyi, C.; Diallinas, G.; Scazzocchio, C. Convergent Evolution and Orphan Genes in the Fur4p-like Family and Characterization of a General Nucleoside Transporter in *Aspergillus nidulans*. *Mol. Microbiol.* **2009**, *73* (1), 43–57. <https://doi.org/10.1111/j.1365-2958.2009.06738.x>.

- (168) Kryptou, E.; Kosti, V.; Amillis, S.; Myriantopoulos, V.; Mikros, E.; Diallinas, G. Modeling, Substrate Docking, and Mutational Analysis Identify Residues Essential for the Function and Specificity of a Eukaryotic Purine-Cytosine NCS1 Transporter \*. *J. Biol. Chem.* **2012**, *287* (44), 36792–36803. <https://doi.org/10.1074/jbc.M112.400382>.
- (169) Sioupouli, G.; Lambrinidis, G.; Mikros, E.; Amillis, S.; Diallinas, G. Cryptic Purine Transporters in *Aspergillus Nidulans* Reveal the Role of Specific Residues in the Evolution of Specificity in the NCS1 Family. *Mol. Microbiol.* **2017**, *103* (2), 319–332. <https://doi.org/10.1111/mmi.13559>.
- (170) Girke, C.; Daumann, M.; Niopek-Witz, S.; Mählmann, T. Nucleobase and Nucleoside Transport and Integration into Plant Metabolism. *Front. Plant Sci.* **2014**, *5*. <https://doi.org/10.3389/fpls.2014.00443>.
- (171) Sucic, S.; Dallinger, S.; Zdrzil, B.; Weissensteiner, R.; Jørgensen, T. N.; Holy, M.; Kudlacek, O.; Seidel, S.; Cha, J. H.; Gether, U.; Newman, A. H.; Ecker, G. F.; Freissmuth, M.; Sitte, H. H. The N Terminus of Monoamine Transporters Is a Lever Required for the Action of Amphetamines \*. *J. Biol. Chem.* **2010**, *285* (14), 10924–10938. <https://doi.org/10.1074/jbc.M109.083154>.
- (172) Gedeon, P. C.; Indarte, M.; Surratt, C. K.; Madura, J. D. Molecular Dynamics of Leucine and Dopamine Transporter Proteins in a Model Cell Membrane Lipid Bilayer. *Proteins* **2010**, *78* (4), 797–811. <https://doi.org/10.1002/prot.22601>.
- (173) Jiang, T.; Wen, P.-C.; Trebesch, N.; Zhao, Z.; Pant, S.; Kapoor, K.; Shekhar, M.; Tajkhorshid, E. Computational Dissection of Membrane Transport at a Microscopic Level. *Trends Biochem. Sci.* **2020**, *45* (3), 202–216. <https://doi.org/10.1016/j.tibs.2019.09.001>.
- (174) Kazmier, K.; Sharma, S.; Islam, S. M.; Roux, B.; Mchaourab, H. S. Conformational Cycle and Ion-Coupling Mechanism of the Na<sup>+</sup>/Hydantoin Transporter Mhp1. *Proc. Natl. Acad. Sci.* **2014**, *111* (41), 14752–14757. <https://doi.org/10.1073/pnas.1410431111>.
- (175) Claxton, D. P.; Quick, M.; Shi, L.; de Carvalho, F. D.; Weinstein, H.; Javitch, J. A.; Mchaourab, H. S. Ion/Substrate-Dependent Conformational Dynamics of a Bacterial Homolog of Neurotransmitter:Sodium Symporters. *Nat. Struct. Mol. Biol.* **2010**, *17* (7), 822–829. <https://doi.org/10.1038/nsmb.1854>.
- (176) Kazmier, K.; Sharma, S.; Quick, M.; Islam, S. M.; Roux, B.; Weinstein, H.; Javitch, J. A.; Mchaourab, H. S. Conformational Dynamics of Ligand-Dependent Alternating Access in LeuT. *Nat. Struct. Mol. Biol.* **2014**, *21* (5), 472–479. <https://doi.org/10.1038/nsmb.2816>.
- (177) Borre, L.; Andreassen, T. F.; Shi, L.; Weinstein, H.; Gether, U. The Second Sodium Site in the Dopamine Transporter Controls Cation Permeation and Is Regulated by Chloride. *J. Biol. Chem.* **2014**, *289* (37), 25764–25773. <https://doi.org/10.1074/jbc.M114.574269>.
- (178) Razavi, A. M.; Khelashvili, G.; Weinstein, H. A Markov State-Based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter. *Sci. Rep.* **2017**, *7*. <https://doi.org/10.1038/srep40076>.
- (179) Malinauskaite, L.; Quick, M.; Reinhard, L.; Lyons, J. A.; Yano, H.; Javitch, J. A.; Nissen, P. A Mechanism for Intracellular Release of Na<sup>+</sup> by Neurotransmitter: Sodium Symporters. *Nat. Struct. Mol. Biol.* **2014**, *21* (11), 1006–1012. <https://doi.org/10.1038/nsmb.2894>.

- (180) Zhao, C.; Noskov, S. Y. The Role of Local Hydration and Hydrogen-Bonding Dynamics in Ion and Solute Release from Ion-Coupled Secondary Transporters. *Biochemistry* **2011**, *50* (11), 1848–1856. <https://doi.org/10.1021/bi101454f>.
- (181) Kniazeff, J.; Shi, L.; Loland, C. J.; Javitch, J. A.; Weinstein, H.; Gether, U. An Intracellular Interaction Network Regulates Conformational Transitions in the Dopamine Transporter. *J. Biol. Chem.* **2008**, *283* (25), 17691–17701. <https://doi.org/10.1074/jbc.M800475200>.
- (182) Khelashvili, G.; Schmidt, S. G.; Shi, L.; Javitch, J. A.; Gether, U.; Loland, C. J.; Weinstein, H. Conformational Dynamics on the Extracellular Side of LeuT Controlled by Na<sup>+</sup> and K<sup>+</sup> Ions and the Protonation State of Glu290. *J. Biol. Chem.* **2016**, *291* (38), 19786–19799. <https://doi.org/10.1074/jbc.M116.731455>.
- (183) Torres, G.; Gainetdinov, R.; Caron, M. G. Plasma Membrane Monoamine Transporters: Structure, Regulation and Function. *Nat. Rev. Neurosci.* **2003**, *4*, 13–25. <https://doi.org/10.1038/nrn1008>.
- (184) Karatza, P.; Panos, P.; Georgopoulou, E.; Frillingos, S. Cysteine-Scanning Analysis of the Nucleobase-Ascorbate Transporter Signature Motif in YgfO Permease of Escherichia Coli: Gln-324 and Asn-325 Are Essential, and Ile-329-Val-339 Form an Alpha-Helix. *J. Biol. Chem.* **2006**, *281* (52), 39881–39890. <https://doi.org/10.1074/jbc.M605748200>.
- (185) Amillis, S.; Koukaki, M.; Diallinas, G. Substitution F569S Converts UapA, a Specific Uric Acid-Xanthine Transporter, into a Broad Specificity Transporter for Purine-Related Solutes. *J. Mol. Biol.* **2001**, *313* (4), 765–774. <https://doi.org/10.1006/jmbi.2001.5087>.
- (186) Amillis, S.; Hamari, Z.; Hamari, Z.; Roumelioti, K.; Scazzocchio, C.; Diallinas, G. Regulation of Expression and Kinetic Modeling of Substrate Interactions of a Uracil Transporter in Aspergillus Nidulans. *Mol. Membr. Biol.* **2007**, *24* (3), 206–214. <https://doi.org/10.1080/09687860601070806>.
- (187) Moran, Y.; Barzilai, M. G.; Liebeskind, B. J.; Zakon, H. H. Evolution of Voltage-Gated Ion Channels at the Emergence of Metazoa. *J. Exp. Biol.* **2015**, *218* (Pt 4), 515–525. <https://doi.org/10.1242/jeb.110270>.
- (188) Karena, E.; Tatsaki, E.; Lambrinidis, G.; Mikros, E.; Frillingos, S. Analysis of Conserved NCS2 Motifs in the Escherichia Coli Xanthine Permease XanQ. *Mol. Microbiol.* **2015**, *98* (3), 502–517. <https://doi.org/10.1111/mmi.13138>.
- (189) Papakostas, K.; Frillingos, S. Substrate Selectivity of YgfU, a Uric Acid Transporter from Escherichia Coli. *J. Biol. Chem.* **2012**, *287* (19), 15684–15695. <https://doi.org/10.1074/jbc.M112.355818>.
- (190) Papakostas, K.; Botou, M.; Frillingos, S. Functional Identification of the Hypoxanthine/Guanine Transporters YjcD and YgfQ and the Adenine Transporters PurP and YicO of Escherichia Coli K-12. *J. Biol. Chem.* **2013**, *288* (52), 36827–36840. <https://doi.org/10.1074/jbc.M113.523340>.
- (191) Karatza, P.; Frillingos, S. Cloning and Functional Characterization of Two Bacterial Members of the NAT/NCS2 Family in Escherichia Coli. *Mol. Membr. Biol.* **2005**, *22* (3), 251–261. <https://doi.org/10.1080/09687860500092927>.
- (192) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (Web Server issue), W665–667. <https://doi.org/10.1093/nar/gkh381>.

- (193) Karena, E.; Frillingos, S. Role of Intramembrane Polar Residues in the YgfO Xanthine Permease: HIS-31 and ASN-93 Are Crucial for Affinity and Specificity, and ASP-304 and GLU-272 Are Irreplaceable. *J. Biol. Chem.* **2009**, *284* (36), 24257–24268. <https://doi.org/10.1074/jbc.M109.030734>.
- (194) Mermelekas, G.; Georgopoulou, E.; Kallis, A.; Botou, M.; Vlantos, V.; Frillingos, S. Cysteine-Scanning Analysis of Helices TM8, TM9a, and TM9b and Intervening Loops in the YgfO Xanthine Permease: A Carboxyl Group Is Essential at ASP-276. *J. Biol. Chem.* **2010**, *285* (45), 35011–35020. <https://doi.org/10.1074/jbc.M110.170415>.
- (195) Chaliotis, A.; Vlastaridis, P.; Ntountoumi, C.; Botou, M.; Yalelis, V.; Lazou, P.; Tatsaki, E.; Mossialos, D.; Frillingos, S.; Amoutzias, G. D. NAT/NCS2-Hound: A Webserver for the Detection and Evolutionary Classification of Prokaryotic and Eukaryotic Nucleobase-Cation Symporters of the NAT/NCS2 Family. *GigaScience* **2018**, *7* (12), giy133. <https://doi.org/10.1093/gigascience/giy133>.
- (196) Pantazopoulou, A.; Diallynas, G. The First Transmembrane Segment (TMS1) of UapA Contains Determinants Necessary for Expression in the Plasma Membrane and Purine Transport. *Mol. Membr. Biol.* **2006**, *23* (4), 337–348. <https://doi.org/10.1080/09687860600738239>.
- (197) Karena, E.; Frillingos, S. The Role of Transmembrane Segment TM3 in the Xanthine Permease XanQ of Escherichia Coli. *J. Biol. Chem.* **2011**, *286* (45), 39595–39605. <https://doi.org/10.1074/jbc.M111.299164>.
- (198) Amillis, S.; Kosti, V.; Pantazopoulou, A.; Mikros, E.; Diallynas, G. Mutational Analysis and Modeling Reveal Functionally Critical Residues in Transmembrane Segments 1 and 3 of the UapA Transporter. *J. Mol. Biol.* **2011**, *411* (3), 567–580. <https://doi.org/10.1016/j.jmb.2011.06.024>.
- (199) Georgopoulou, E.; Mermelekas, G.; Karena, E.; Frillingos, S. Purine Substrate Recognition by the Nucleobase-Ascorbate Transporter Signature Motif in the YgfO Xanthine Permease: ASN-325 Binds and ALA-323 Senses Substrate. *J. Biol. Chem.* **2010**, *285* (25), 19422–19433. <https://doi.org/10.1074/jbc.M110.120543>.
- (200) Kaback, H. R.; Dunten, R.; Frillingos, S.; Venkatesan, P.; Kwaw, I.; Zhang, W.; Ermolova, N. Site-Directed Alkylation and the Alternating Access Model for LacY. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (2), 491–494. <https://doi.org/10.1073/pnas.0609968104>.
- (201) Papakostas, K.; Georgopoulou, E.; Frillingos, S. Cysteine-Scanning Analysis of Putative Helix XII in the YgfO Xanthine Permease: ILE-432 and ASN-430 Are Important. *J. Biol. Chem.* **2008**, *283* (20), 13666–13678. <https://doi.org/10.1074/jbc.M800261200>.
- (202) Kourkoulou, A.; Pittis, A. A.; Diallynas, G. Evolution of Substrate Specificity in the Nucleobase-Ascorbate Transporter (NAT) Protein Family. *Microb. Cell* **2018**, *5* (6), 280–292. <https://doi.org/10.15698/mic2018.06.636>.
- (203) Bürzle, M.; Suzuki, Y.; Ackermann, D.; Miyazaki, H.; Maeda, N.; Clémenton, B.; Burrier, R.; Hediger, M. A. The Sodium-Dependent Ascorbic Acid Transporter Family SLC23. *Mol. Aspects Med.* **2013**, *34* (2), 436–454. <https://doi.org/10.1016/j.mam.2012.12.002>.
- (204) Niopek-Witz, S.; Deppe, J.; Lemieux, M. J.; Möhlmann, T. Biochemical Characterization and Structure–Function Relationship of Two Plant NCS2 Proteins, the Nucleobase Transporters NAT3 and NAT12 from Arabidopsis Thaliana. *Biochim. Biophys. Acta BBA - Biomembr.* **2014**, *1838* (12), 3025–3035. <https://doi.org/10.1016/j.bbamem.2014.08.013>.

- (205) Mansfield, T. A.; Schultes, N. P.; Mourad, G. S. AtAzc1 and AtAzc2 Comprise a Novel Family of Purine Transporters in Arabidopsis. *FEBS Lett.* **2009**, *583* (2), 481–486. <https://doi.org/10.1016/j.febslet.2008.12.048>.
- (206) Botou, M.; Lazou, P.; Papakostas, K.; Lambrinidis, G.; Evangelidis, T.; Mikros, E.; Frillingos, S. Insight on Specificity of Uracil Permeases of the NAT/NCS2 Family from Analysis of the Transporter Encoded in the Pyrimidine Utilization Operon of *Escherichia Coli*: Functional Profile of *E. Coli* Permease RutG. *Mol. Microbiol.* **2018**, *108* (2), 204–219. <https://doi.org/10.1111/mmi.13931>.
- (207) Botou, M.; Yalelis, V.; Lazou, P.; Zantza, I.; Papakostas, K.; Charalambous, V.; Mikros, E.; Flemetakis, E.; Frillingos, S. Specificity Profile of NAT/NCS2 Purine Transporters in *Sinorhizobium* (Ensifer) *Meliloti*. *Mol. Microbiol.* **2020**, *n/a* (n/a). <https://doi.org/10.1111/mmi.14503>.
- (208) Thornton, J. W. Resurrecting Ancient Genes: Experimental Analysis of Extinct Molecules. *Nat. Rev. Genet.* **2004**, *5* (5), 366–375. <https://doi.org/10.1038/nrg1324>.
- (209) Chang, B. S. W.; Jönsson, K.; Kazmi, M. A.; Donoghue, M. J.; Sakmar, T. P. Recreating a Functional Ancestral Archosaur Visual Pigment. *Mol. Biol. Evol.* **2002**, *19* (9), 1483–1489. <https://doi.org/10.1093/oxfordjournals.molbev.a004211>.
- (210) Seeger, M. A. Membrane Transporter Research in Times of Countless Structures. *Biochim. Biophys. Acta BBA - Biomembr.* **2018**, *1860* (4), 804–808. <https://doi.org/10.1016/j.bbamem.2017.08.009>.
- (211) Yan, N. A Glimpse of Membrane Transport through Structures—Advances in the Structural Biology of the GLUT Glucose Transporters. *J. Mol. Biol.* **2017**, *429* (17), 2710–2725. <https://doi.org/10.1016/j.jmb.2017.07.009>.
- (212) Yao, X.; Fan, X.; Yan, N. Cryo-EM Analysis of a Membrane Protein Embedded in the Liposome. *Proc. Natl. Acad. Sci.* **2020**, *117* (31), 18497–18503. <https://doi.org/10.1073/pnas.2009385117>.

## Appendix: Publications and Pre-prints



# Hydroxytyrosol (HT) Analogs Act as Potent Antifungals by Direct Disruption of the Fungal Cell Membrane

George Diallinas<sup>1\*</sup>, Nausica Rafailidou<sup>1</sup>, Ioanna Kalpaktsi<sup>2</sup>, Aikaterini Christina Komianou<sup>2</sup>, Vivian Tsouvali<sup>1</sup>, Iliana Zantza<sup>2</sup>, Emmanuel Mikros<sup>2</sup>, Alexios Leandros Skaltsounis<sup>2</sup> and Ioannis K. Kostakis<sup>2\*</sup>

<sup>1</sup> Department of Biology, National and Kapodistrian University of Athens, Athens, Greece, <sup>2</sup> Department of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece

## OPEN ACCESS

### Edited by:

Raffaella Balestrini,  
Consiglio Nazionale delle Ricerche  
(CNR), Italy

### Reviewed by:

Giorgio Gnani,  
Università degli studi di Torino, Italy  
Walter Chitarra,  
Consiglio per la ricerca in agricoltura e  
l'analisi dell'economia agraria (CREA),  
Italy

### \*Correspondence:

George Diallinas  
diallina@biol.uoa.gr  
Ioannis K. Kostakis  
ikkostakis@pharm.uoa.gr

### Specialty section:

This article was submitted to  
Fungi and Their Interactions,  
a section of the journal  
Frontiers in Microbiology

**Received:** 25 June 2018

**Accepted:** 15 October 2018

**Published:** 06 November 2018

### Citation:

Diallinas G, Rafailidou N, Kalpaktsi I, Komianou AC, Tsouvali V, Zantza I, Mikros E, Skaltsounis AL and Kostakis IK (2018) Hydroxytyrosol (HT) Analogs Act as Potent Antifungals by Direct Disruption of the Fungal Cell Membrane. *Front. Microbiol.* 9:2624. doi: 10.3389/fmicb.2018.02624

Fungal infections constitute an emerging threat and a prevalent health problem due to increasing number of immunocompromised people and pharmacological or other treatments aiming at viral infections, cancer or allergies. Currently used antifungals suffer from inefficiency, toxic side effects and developing drug-resistance. Additionally, over the last two decades no new classes of antifungals have been approved, emphasizing the urgent need for developing a novel generation of antifungals. Here, we investigate the antifungal activity of a series of chemically synthesized Hydroxytyrosol (HT) analogs. HT is one of the major phenolic compounds in olive oil, shown to possess radical-scavenging antioxidant, antiproliferative, proapoptotic and anti-inflammatory activities. No previous report has studied HT analogs as antifungals. We show that specific analogs have broad and strong antifungal activity, significantly stronger than the parent compound HT. Using *Aspergillus nidulans* as an *in vivo* cellular model system, we show that antifungal HT analogs have an unprecedented efficiency in fungal plasma membrane destruction. Importantly, antifungal HT analogs did not show toxicity in a mammalian cell line, whereas no resistance to HT analogs was obtained by standard mutagenesis. Our results open the way for the development of a novel, efficient and safer class of antifungals.

**Keywords:** fungal pathogens, *Aspergillus nidulans*, plasma membrane, antimicrobial, resistance

## INTRODUCTION

In recent years, systemic fungal infections have emerged as an increasingly prevalent health problem (McCarthy et al., 2017). Infections are rising among immunocompromised patients, including individuals suffering from HIV/AIDS or diabetes mellitus, or following organ transplantations and immunosuppressive chemotherapy during cancer treatment (Low and Rotstein, 2011). The most clinically significant invasive opportunistic fungal pathogens belong to one of the four groups: *Aspergillus*, *Candida*, *Cryptococcus* and *Pneumocystis*, with the first two being the most important of all fungal pathogens (Köhler et al., 2017). Currently used antifungals include three major classes of drugs with different mechanisms of action: polyenes (disrupt fungal

membranes), azoles (inhibit ergosterol biosynthesis), and echinocandins (inhibit synthesis of cell wall  $\beta$ -glucan) (Odds et al., 2003). However, all current antifungals suffer from inefficiency, toxic side effects, drug-drug interactions and developing drug-resistance (Arendrup, 2014; Cuenca-Estrella, 2014; Fairlamb et al., 2016). In addition, since 2001, no new classes of antifungals have been approved. This emphasizes the urgent and critical need for developing a novel generation of antifungals.

Pharmacological therapies for various bacterial or viral infections based on natural, mostly herbal, agents are among the most current therapeutic trends in medicine (Pan et al., 2013; Li and Weng, 2017). The antifungal activity of a large number of natural products is well documented (Di Santo, 2010). The most well-known structures exhibiting fungicide activity are polyenes, oligopeptides, terpenoids and macrolides, while different other natural products like flavonoids, alkaloids and phenolic acids have been also reported (Tasleem et al., 2009; Teodoro et al., 2015). Olive (*Olea europaea*), extracts has been reported for their antimicrobial and bacteriostatic activity since 1970 (Walter et al., 1973). Mild antifungal activity has been also described for olive leaf extracts (Korukluoglu et al., 2008) and the major olive phenolic compound oleuropein (Zorić et al., 2016). Recently, plant extracts and chemically synthesized related analogs from olive have also shown an antiprotozoan activity (Belmonte-Reche et al., 2016; Koolen et al., 2017). Hydroxytyrosol (HT), one of the major phenolic compounds in olive oil, has recently received particular attention because of its radical-scavenging, antiproliferative, proapoptotic and anti-inflammatory activities, which seem to have a counteractive effect on carcinogenesis and other cellular malfunctions in animal trials and *in vitro*. Additionally, recent evidence has shown that HT and its analogs might be promising antibacterial, antiviral or antiprotozoan agents (Manna et al., 2005; Fernandez-Bolanos et al., 2008, 2012; Chillemi et al., 2010; Koolen et al., 2017; Thielmann et al., 2017; Robles-Almazan et al., 2018). However, no systematic effort has been invested in search of novel antifungals based on HT, except from some reports concerning yeast species (Pereira et al., 2007; Zoric et al., 2013), or HT analogs.

Based on some preliminary tests of our lab that showed a moderate antifungal effect of HT and some HT analogs on *Aspergillus nidulans*, here we systematically synthesize and test the antifungal activities of an extended series of novel HT analogs. The rationale of their synthesis was based mostly on the possible effect of the length and saturation of the fatty acid chain, and the substitution of the  $\alpha$ -carbon of the HT side chain. We show that several of the synthesized HT analogs have a very potent and broad antifungal activity against major fungal pathogens, such as *A. fumigatus*, *A. flavus*, *Fusarium oxysporum*, *A. nidulans* and *Candida albicans*. Importantly, we reveal that the antifungal activity of HT analogs is directly related to its rapid destructive effect on fungal plasma membranes, which in turn justifies why resistance mutants could not be isolated. Our work is expected to open the way for developing a new class of highly potent novel antifungals.

## RESULTS

### Chemical Synthesis of HT Analogs

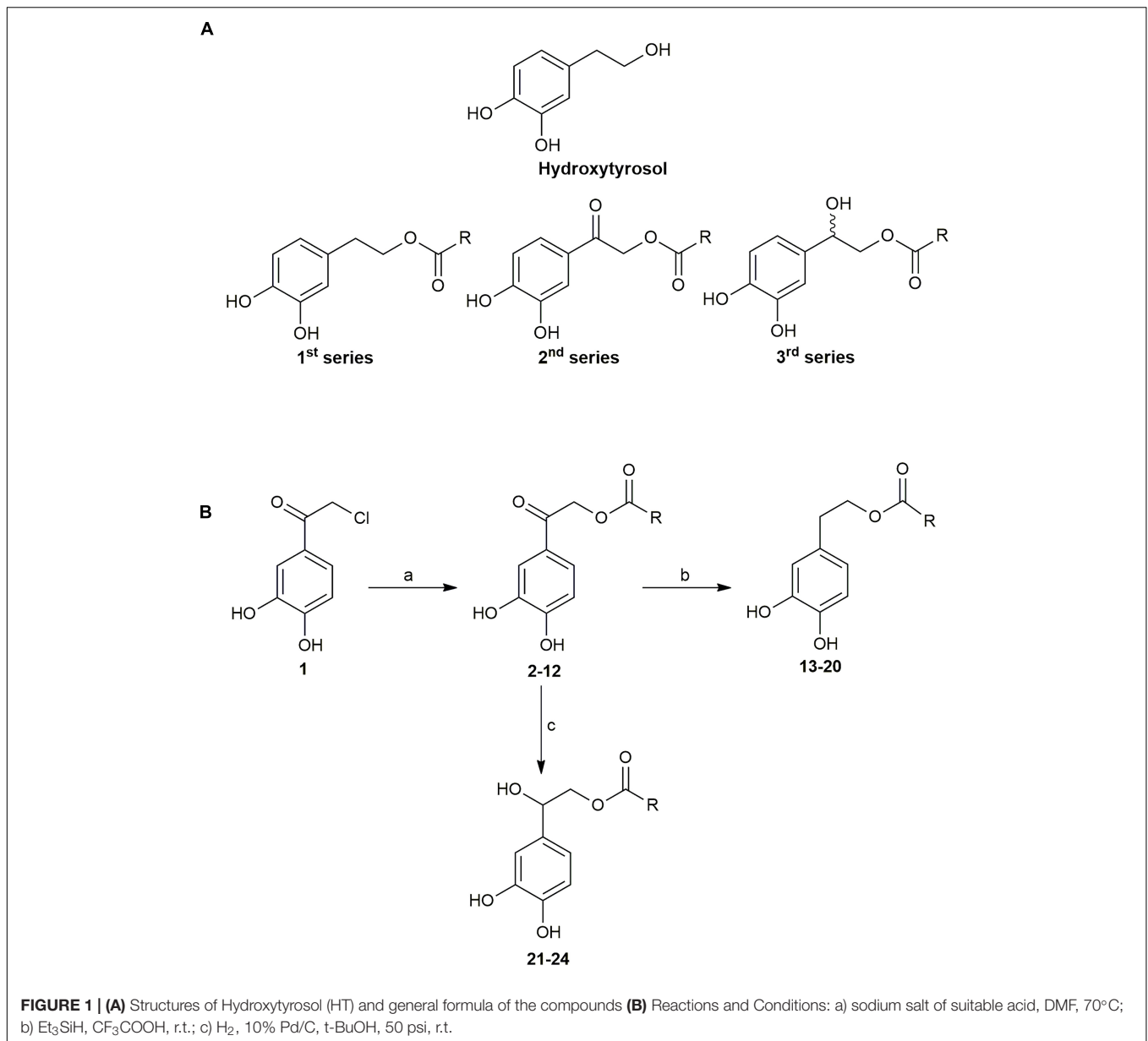
Twenty three analogs of HT were synthesized as described in detail in the section “Materials and Methods” and in **Supplementary Material (Supplementary Figure S1)**. The new compounds are lipophilic esters of HT bearing modifications on the  $\alpha$ -carbon of the catechol side chain. More precisely, the new compounds are categorized in 3 different series (**Figure 1A**). The compounds of the second series are simple esters of HT in the aliphatic hydroxyl group, whereas the compounds of the first series possess a carbonyl group on the  $\alpha$ -carbon of the catechol side chain. The compounds of the third series possess a hydroxyl group on the  $\alpha$ -carbon of the catechol side chain. In brief, chloride **1** was treated with the sodium salt of the appropriate acid to afford the required keto esters **2-12 (Figure 1B)**. The majority of the acids are commercially available, though the acids for the preparation of ester **9** was synthesized through a Wittig reaction of methyl 4-methylene cyclohexane carboxylate. Treatment of the previous compounds with triethylsilane in trifluoroacetic acid, provided the desired, fully reduced, lipophilic esters **13-20**. It is noteworthy that this method was successful in yields up to 80%, with high reproducibility and scale up to grams. Finally, partial reduction with hydrogenation over Pd/C, provided the hydroxyl derivatives **21-24**, as racemic mixtures.

### HT Analogs Show High Potential as Antifungals

The 23 chemically synthesized analogs of HT were tested for their antifungal potential against *A. nidulans*, *A. fumigatus*, *A. flavus*, *C. albicans* and *Fusarium oxysporum*, which are important fungal pathogens of animals and/or plants. The rationale for choosing *A. nidulans* was also based on its unique amenability as a model system for genetic and functional analysis, rather than its pathogenic profile, which would allow the investigation of the molecular mechanisms underlying of the antifungal action of HT analogs. All synthesized HT analogs were tested as described in Materials and methods. We tested all analogs on solid minimal media at physiological and optimal pH ranges (5.5–6.8) and temperatures (25–37°C). The effect on *C. albicans* was also recorded in liquid fresh cultures at their logarithmic phase of growth. Importantly, similar results were obtained at the different pH tested and on complete media or minimal media. Notably, recorded apparent antifungal activities were significantly higher at 37°C compared to 25°C (see also later).

**Figure 2** and **Table 1** highlight our findings and reflects the outcome obtained through at least three independent growth tests that showed practically identical results. Nine HT analogs (**2, 4, 5, 10, 11, 15, 16, 18** and **19**), shown in **Figure 3** had strong antifungal activity against *A. nidulans* at 37°C, mostly evident at 400  $\mu$ M, and six of them (**2, 5, 11, 15, 16** and **19**) were also very active at 25°C. Most of the same nine analogs also had strong antifungal activity against other fungi tested, at 25°C and 37°C (**Figure 2A, Table 1** and test not shown). In particular,



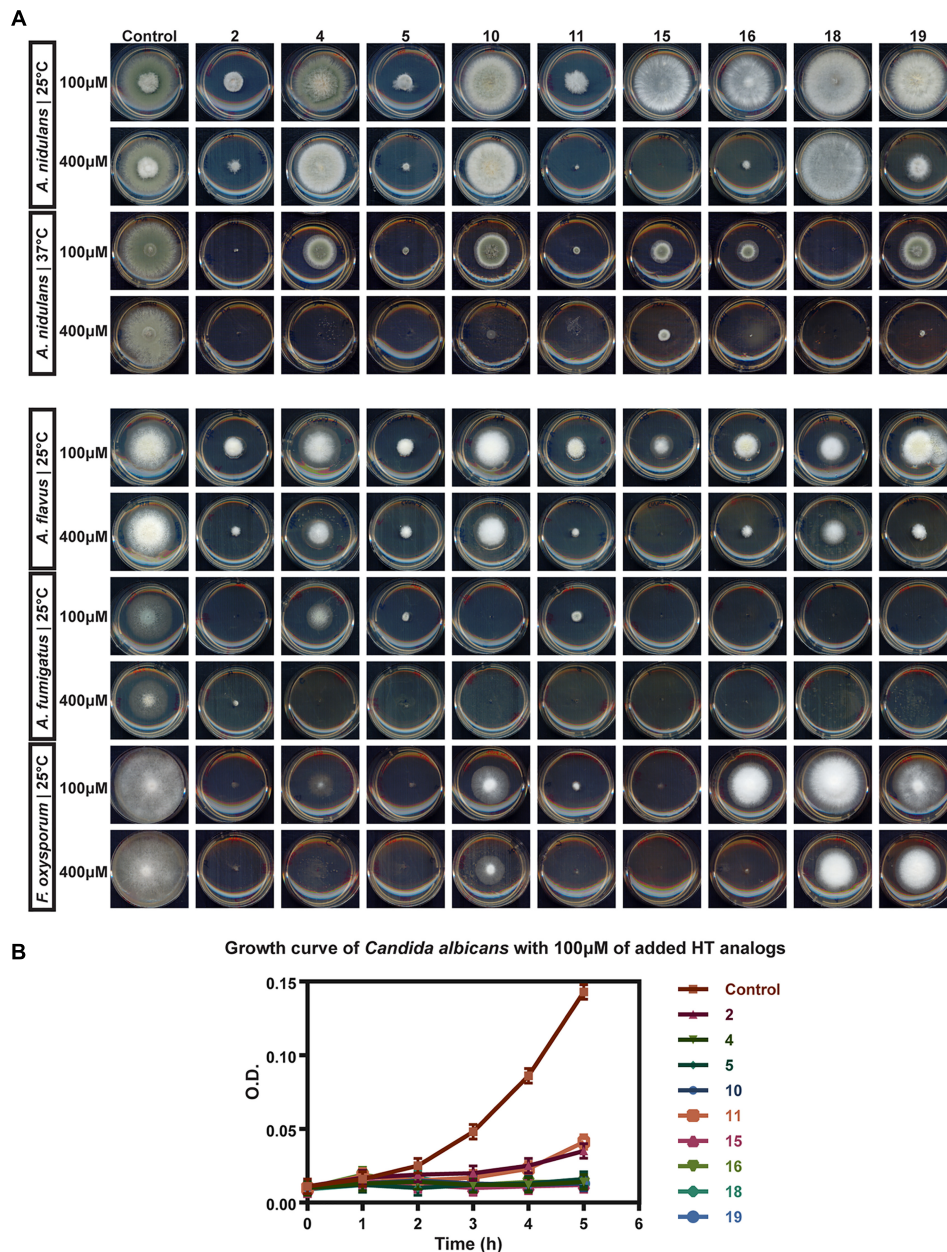


all nine analogs were extremely toxic to *A. fumigatus*, leading to total or extremely severe inhibition of growth at 100  $\mu$ M. *A. flavus* proved to be the most resistant fungus among those tested, but still several analogs were highly inhibitory for its growth (2, 5, 11, 15, 16 and 19). Best antifungal agents against *F. oxysporum* proved to be analogs 2, 5, 11 and 15. All analogs, at 100  $\mu$ M, severely inhibited *C. albicans* growth in liquid cultures (Figure 2B) or on solid minimal media (not shown) at 37°C. When liquid cultures were left to grow for more than 24 h after the initial addition of HT analogs, growth *C. albicans* resumed in several cases, but not in the presence of analogs 4 or 15 (Supplementary Figure S2). This indicates that these analogs had the strongest cytotoxic effect or that these compounds were the most stable under the relative experimental conditions. Overall, several synthesized analogs of HT have a dramatically

increased antifungal activity, compared to the “mother” natural compound (HT), and importantly, the most of active of them are comparable to Amphotericin B (Supplementary Figure S3). The rationale of this latter comparison becomes apparent later.

### Antifungal HT Analogs Show Variable Antibacterial Activities

The HT analogs with the highest antifungal activity were also tested, at a range of 100–200  $\mu$ M, in fresh exponentially growing bacterial cultures in order to gain further insight on their mode of action and as a first step for investigating their potential use as broad range antimicrobials. Figure 4 shows the results obtained with *E. coli* and *B. subtilis*,



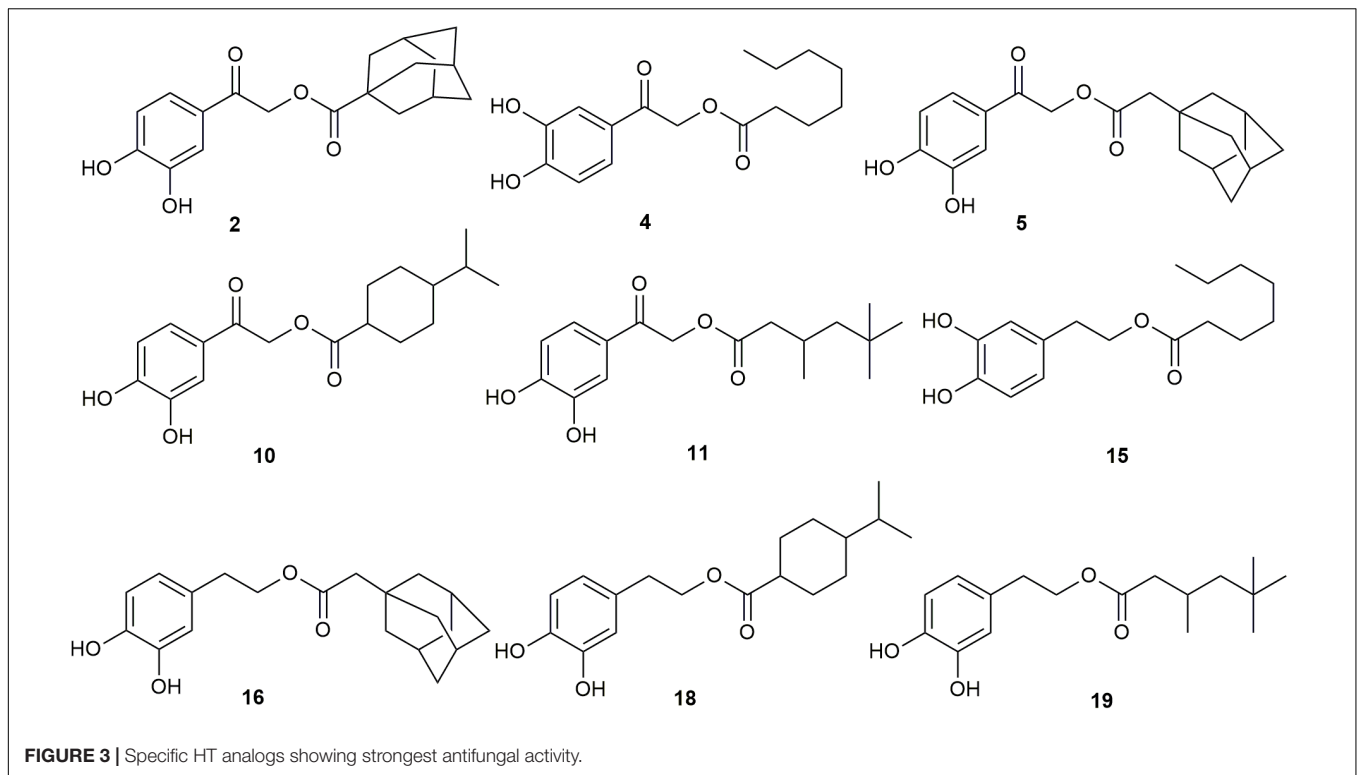
**FIGURE 2 |** *In vivo* evaluation of HT analogs as antifungals **(A)** Growth tests showing the antifungal activity of certain HT analogs against pathogens *A. flavus*, *A. fumigatus*, *F. oxysporum* and *A. nidulans*. Growth on two concentrations of HT analogs for each microorganism is shown. **(B)** Growth curve of *C. albicans*. O.D stands of Optical Density at 600 nm of liquid cultures recorded hourly. Analogs were added to the cultures at 200 μM. Control stands for cultures were only DMSO solvent was added in the cultures, at the same concentrations as the analogs.

as typical representatives of G<sup>-</sup> and G<sup>+</sup> bacteria. All analogs were highly toxic to *B. subtilis* at 100 μM, but not at all to *E. coli*. *Pseudomonas* species showed differential growth behavior, with *P. aeruginosa* being fully resistant, but *P. fluorescens* highly sensitive at 200 μM (Supplementary Figure S4). Additional bacterial species were also tested (*Klebsiella*, *Enterococcus*, *Staphylococcus*, *Streptococcus* and other G<sup>-</sup> enterobacteria) showing varying degrees of sensitivity to HT analogs, but this will be reported elsewhere as the

present work is directed toward the discovery of novel antifungals.

### Antifungal HT Analogs Disrupt the Structure and Function of *A. nidulans* Plasma Membrane

The non-dependence of the antifungal action of the HT analogs on whether minimal or complete media are used, or the pH



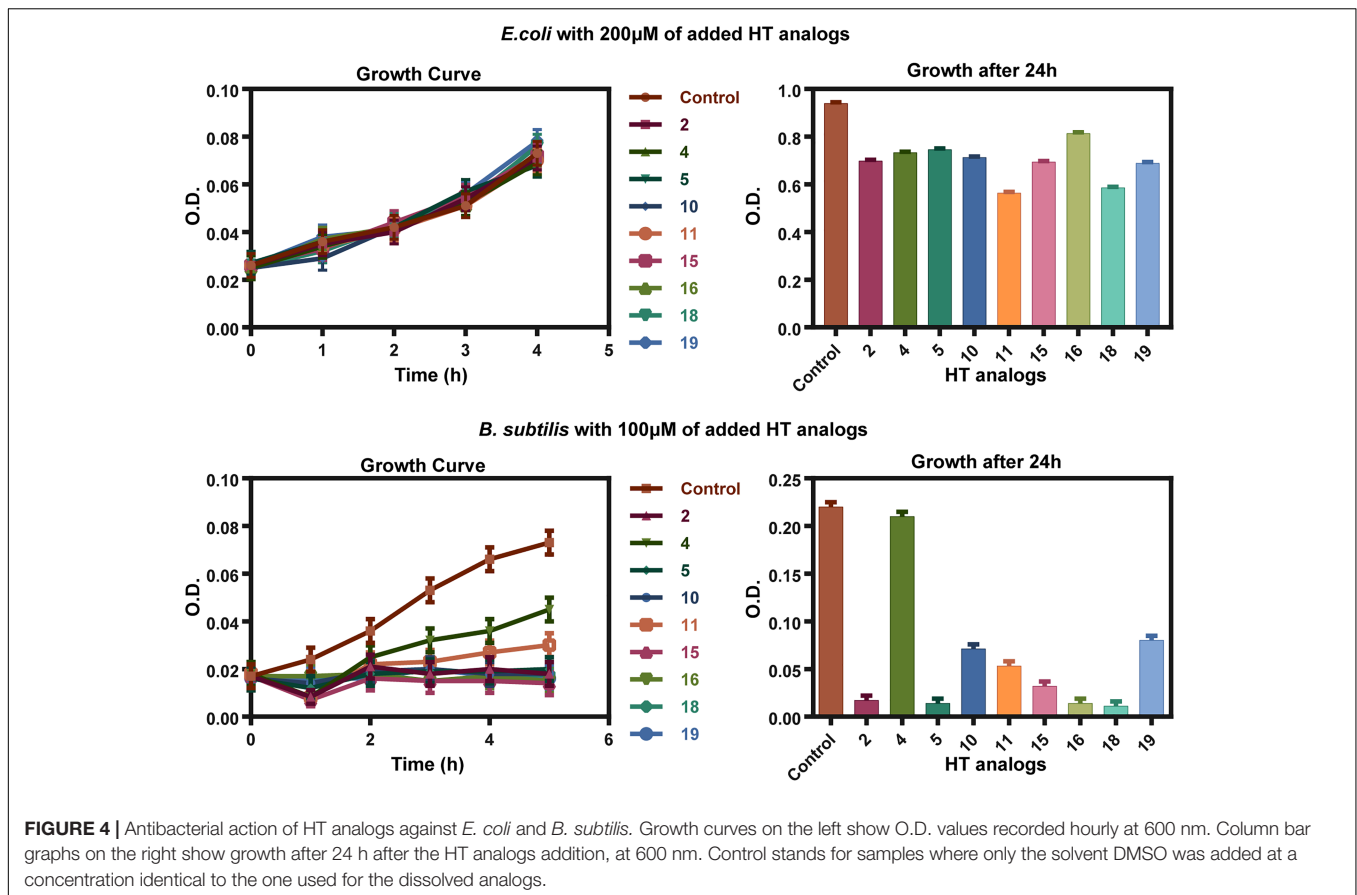
**TABLE 1** | Approximate concentrations of HT analogs that lead to 50% reduction of fungal growth ( $IC_{50}$ ).

HT	2	4	5	10	11	15	16	18	19	
<i>A. nidulans</i> 37°C	>400	~50	<100	~50	~100	~50	<100	<100	~50	~100
<i>A. nidulans</i> 25°C	>400	<100	<200	<100	<200	~100	<200	<200	<200	<200
<i>A. fumigatus</i>	>400	~50	<100	~50	~50	<100	<50	<100	~50	~50
<i>A. flavus</i>	>400	<100	>100	<100	>100	<100	<100	~100	~100	>100
<i>F. oxysporum</i>	>400	<50	<50	<50	~100	<50	<50	<200	<200	<200
<i>C. albicans</i>	>400	~100	~100	~100	~100	~100	~100	~100	~100	~100

$IC_{50}$  values shown correspond to the approximate  $\mu\text{M}$  concentration of the compounds that reduce the diameter of colony growth by 50% after 4–6 days at 37°C, on standard agar minimal media.  $IC_{50}$  are estimated by testing growth in the presence of a range of  $\mu\text{M}$  concentration (i.e., 0, 50, 100, 200, 400, 500, and 1000  $\mu\text{M}$ ). The values shown were estimated by at least three independent experiments, which showed no significant variation ( $SD = 20 \pm 5\%$ ). For one of the most active analogs, **15**, additional experiments performed at a lower concentration range (5, 10, 20, 50, and 100  $\mu\text{M}$ ), led to a more precise estimation of  $IC_{50}$ , which corresponds to 17  $\mu\text{M}$ . Most analogs led to total inhibition of growth at 100–400  $\mu\text{M}$  against all fungi tested (see **Figure 2**).

range, when this was kept within limits proper for fungal growth, suggested that these compounds might be either taken up by non-facilitated diffusion, or exert their activity directly without the need to enter the cell (i.e., on the cell wall or the plasma membrane). The relatively increased activity observed at 37°C vs. 25°C in the case of *A. nidulans* (see **Figure 2**) does not distinguish between the two possibilities, as increased membrane fluidity at a higher temperature would favor diffusion, as well as, binding of hydrophobic HT analogs in specific lipids of the fungal membrane. Thus, to investigate this issue directly we followed the effect of all antifungal HT analogs on the microscopic morphology and the plasma membrane (PM) of *A. nidulans* hyphae, using Brightfield and Epifluorescence microscopy, respectively. For investigating the effect on the plasma membrane we used strains expressing two GFP-tagged plasma membrane transporters, namely the UapA

uric acid-xanthine transporter (Pantazopoulou et al., 2007) or the FurA allantoin transporter (Kryptou et al., 2015). **Figure 5** summarizes our results. Most HT analogs, when added at final concentration as low as 37.5  $\mu\text{M}$ , for 0–30 min, had a rapid and prominent effect on plasma membrane integrity, reflected in dramatic reduction of transporter-associated peripheral GFP fluorescence signal and concomitant appearance of static, non-cortical cytosolic membrane fluorescent aggregates (most evident with analogs **2**, **5**, **10**, **11**, **16** or **18**). The great majority of these cytosolic fluorescent aggregates did correspond to vacuoles or other known endomembrane compartments, as evidenced by CMAC or FM4-64 staining (not shown). Under brightfield light we did not notice any dramatic modification that would be compatible with disruption of the cell wall or overall hyphae morphology (see also **Supplementary Figure S5**), despite a visible increase in vacuole number and size. These



effects were practically immediate, becoming evident in 1–5 min, which also somehow excludes the idea that HT analogs act primarily by metabolic inhibition of an enzyme. In general, the relative strength of the detrimental effect of the different HT analogs on the PM was variable, with some analogs leading to total apparent disintegration, while others led to significant but not total, of the PM in 10 min. The concentration used to test the analogs was 37.5  $\mu$ M in order to avoid any effect of DMSO (the solvent used) on the stable localization of transporters in the PM, as we have noticed that DMSO concentrations > 50  $\mu$ M elicit a degree of endocytic turnover for most transporters studied in our lab (unpublished observations).

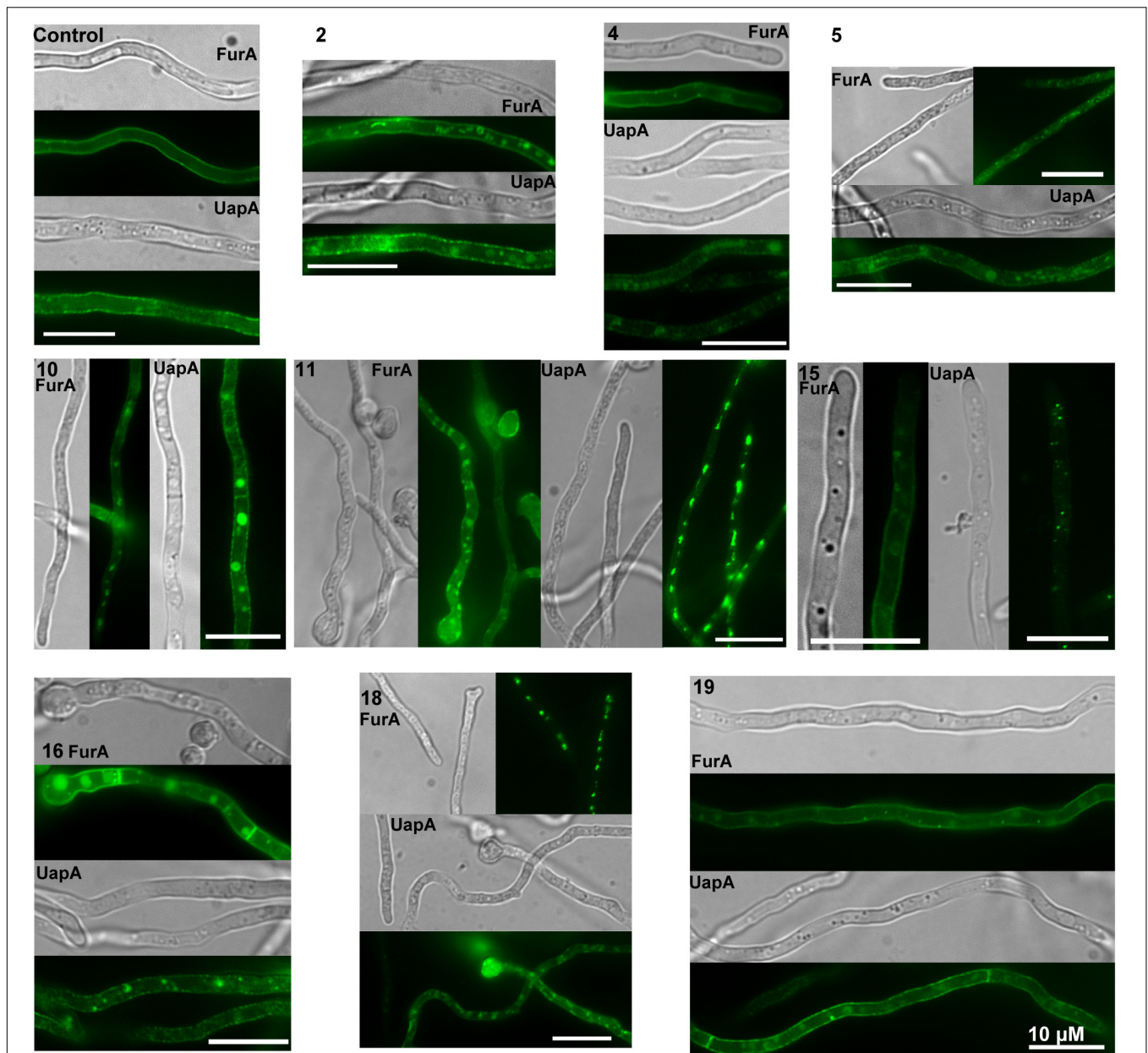
To further examine the nature of the effect of HT analogs on the PM, we also performed direct transport assays of radiolabeled metabolites, imported by specific transporters, in the presence of excess analogs. Transport assays used measure initial uptake rates (at 60 s) of radiolabeled substrates in germinated conidiospores (Kryptou and Diallinas, 2014). In particular, we tested the uptake of radiolabeled xanthine, which is specifically transported by two uptake systems, the UapA (~70%) and UapC (~30%) transporters (Pantazopoulou and Diallinas, 2007; Kryptou and Diallinas, 2014). Results are summarized in **Figure 6**. HT analogs 2, 4, 5, 10, 11, 15 and 16 reduced xanthine uptake to ~20–40% of the control sample (only DMSO added), whereas the rest (18 and 19) showed less inhibitory effect. Overall, these

results strongly suggested that most HT analogs tested had a rapid negative effect on fungal transport systems. Given that the relevant xanthine transporters, UapA and UapC, as most fungal transporters, are H<sup>+</sup> symporters, our results can in principle be explained by two scenarios; either the analogs lead to rapid disorganization of the PM, or they led to a rapid depolarization of the membrane, acting as direct H<sup>+</sup> gradient uncouplers. However, the latter case seems unlikely, or secondary, under the light of the microscopic analysis shown in **Figure 5**, which directly confirmed the dramatic effect of all HT analogs tested on the PM within some minutes after their addition to the fungal cultures. Additionally, the strength of inhibition of transporter-mediated xanthine uptake by the different HT analogs tested was in good agreement with the results obtained following their effect on PM integrity (**Figure 5**) and their *in vivo* antifungal activity (**Figure 2**). We thus conclude that the direct target of HT analogs is disruption of the fungal PM integrity and function.

### Lack of Resistance to HT Analog Antifungal Action

Based on our findings that showed that HT analogs act directly and rapidly on the PM, we presumed that resistance to this novel type of antifungals will be infrequent if any, similar to the case of other antimicrobials that target directly the periphery





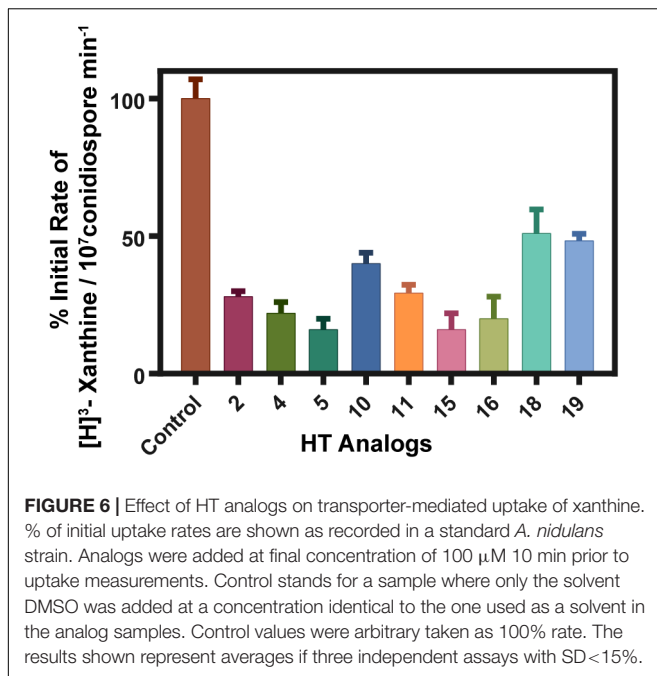
**FIGURE 5 |** Effect of HT analogs on *A. nidulans* plasma membrane and transporter-mediated uptake of xanthine. Epifluorescence *in vivo* microscopy showing the effect of HT analogs (37.5  $\mu$ M) on the plasma membrane of *A. nidulans*. The picture shows young hyphae of strains expressing functional GFP-tagged FurA or UapA transporters as PM molecular markers. FurA expression is stronger than that of UapA, due to transcription via a strong constitutive promoter (*gpdA<sub>p</sub>*), while UapA is transcribed by its native, relatively weak, promoter. Notice that upon addition of most analogs cortical fluorescent labeling is reduced with concurrent appearance of cytosolic fluorescent foci, which represent membrane aggregates. Scale bars shown are 10  $\mu$ M.

of microbes, such as amphotericin B (plasma membrane) or Echinocandins (cell wall) (Cuenca-Estrella, 2014). To test this, we performed several experiments of standard u.v. or transposon-driven mutagenesis as described in Materials and methods of an appropriate *A. nidulans* strain (i.e., one carrying the Minos transposable element (Evangelinos et al., 2015)), and tried to select mutant colonies resistant to 400  $\mu$ M of analog 15. We did not manage to isolate any resistance colony, even in cases where lethality of u.v. mutagenesis was 93.7%. This apparently negative

result is fully compatible with HT analogs targeting the PM of *A. nidulans*.

## DISCUSSION

Here, we showed that specific chemically synthesized HT analogs possess broad and strong antifungal activity due to an immediate primary effect on the fungal plasma membrane.



We do not, however, exclude that the HT analogs also have secondary cytotoxic targets (e.g., cell wall or enzymes), mostly in other microorganisms, which we did not test directly here. An additional finding classifies this specific set of compounds as very promising novel antifungals; lack of detectable fungal (*A. nidulans*) resistance due to mutations. The lack of obtaining spontaneous or mutation-induced resistance to an antifungal is a strong indication that the primary site of action of this antifungal is a molecule in the periphery of fungal cells, most probably a specific lipid of the PM. This is in full agreement with the effect detected herein of the HT analogs on the PM. Moreover, the lack of resistance resembles cases of other antifungals targeting the plasma membrane, such as amphotericin B, where only infrequent resistance is obtained due to mutations affecting the level of ergosterol and phospholipids in the membrane (Cuenca-Estrella, 2014; Fairlamb et al., 2016). Finally, an other aspect concerning the promising future of present HT analogs as novel antifungals is the observation that in preliminary tests analogs 15 and 16, which showed high antifungal activity, showed no toxicity against a mammalian cell line (see Materials and methods). Overall, our results open the way for the rigorous development of a novel, efficient and maybe safer class of antifungals. Additionally, the antifungal HT analogs were also shown to also possess antimicrobial activity against specific G+ bacteria. This is an interesting, albeit rather preliminary, observation that we want to exploit in the future as it might give us information on the exact mechanism of action of interaction of these compounds with specific membrane lipids.

In several previous studies, HT has been shown to have variable antibacterial activity at relatively high concentrations, in the range of 1–4 mM (Capasso et al., 1995; Bisignano et al., 1999; Medina et al., 2006, 2007; Tafesh et al., 2011; Medina-Martínez et al., 2016). However, bacterial growth was

never fully inhibited and a general conclusion on the effects of HT on Gram-positive or Gram-negative species could not be made (Medina-Martínez et al., 2016). To our knowledge, the only investigations regarding the antifungal potential of HT are those against medically important *Candida* spp., which showed minimal growth inhibitory values ranging between 633  $\mu$ M and 40 mM (Pereira et al., 2007; Zoric et al., 2013). Notably, fluorescent dye-exclusion based studies with *C. albicans* revealed a membrane associated antifungal mechanism at sub-inhibitory concentrations (Zoric et al., 2013), in line with results presented herein. In the past, HT-derived compounds have been synthesized aiming at analogs with a better hydrophilic/lipophilic balance (HLB) to increase their cellular uptake and thus enhance their antioxidant or other unknown activities (Grasso et al., 2007; Bernini et al., 2012, 2015). However, no previous studies evaluating synthesized HT analogs as antifungal agents have been reported. HT derivatives have only recently been shown to have a very promising antitrypanosomal and antileishmanial activity (Belmonte-Reche et al., 2016). IC<sub>50</sub> values against *Trypanosoma brucei* for HT decanoate ester and HT dodecanoate ester were 0.6 and 0.36  $\mu$ M, respectively. This represented a significant 79–132 fold improvement in activity compared to HT. Focusing on structure-activity relationships, the authors found via rational design several more HT analogs to have targeted cytotoxic activity against *T. brucei* with IC<sub>50</sub> values in the low micromolar range. They concluded that the di-ortho phenolic ring and medium size alkyl chain are essential for activity, whereas the nature of the chemical bond among them seems less important. Importantly, antiprotozoan HT analogs displayed a high selectivity index against MRC-5, a non-tumoral human cell line, which is in line with the targeted antimicrobial specificity we also found for the HT antifungal analogs reported here. Thus, HT is indeed a highly promising mother compound to develop novel broad range antimicrobial agents.

Several of the new compounds described here showed varying degrees of antifungal activity against the tested pathogens, nevertheless compounds 2, 5, 11, 15, 16 and 19 possessed the strongest activity, clearly higher than the mother compound HT. All the active compounds belong to the first and second series, while compounds of the third series (21, 22, 23, 24; see **Supplementary Figure S1**) were non-active. The carbonyl substituted compounds of the first series, and especially the alkyl-substituted analogs 2, 5 and 11, were among the most potent compounds, suggesting that the substitution on the  $\alpha$ -carbon is crucial for the activity. It seems that the presence of the hydroxyl group, is incompatible with activity, whereas the carbonyl group increases the antifungal capacity, probably by increasing the acidity of the catechol moiety. Notably also, all compounds in this study with an aromatic substitution (3, 6, 14 and 22; see **Supplementary Figure S1**) showed no activity against the tested pathogens, while the alkyl substituted compounds of first and second series were shown to be potent antifungals. The role of the double bond of the alkyl side chain is detrimental for the activity (see **Supplementary Figure S1**). Additionally, the role of the branching is not clear since the branched alkyl chained compounds 2, 5, 10, 11, 16, 18 and 19 showed potent antifungal activity, whereas compounds 9,

12, 13 and 20 seem to be inactive. Nevertheless, our results demonstrate that only the alkyl substituted compounds of first and second series are active against the tested pathogens, with the most effective being those with an alkyl chain of 6 to 10 carbons. These results are also in agreement with the ones against *Trypanosoma brucei* for HT decanoate ester and HT dodecanoate ester (Belmonte-Reche et al., 2016).

According to recent data, alkyl gallates and other amphiphilic phenols, compounds possessing similar structure to the HT analogs presented here, have significant antifungal activity (Kozubek et al., 2001; Kubo et al., 2001b). It is suggested that their potency may result from the interaction with cell wall components (phospholipids, etc.); therefore, their activity is due partly to the ability of these molecules to function as nonionic surfactants (surface-active compounds), disrupting the fungi membrane. The mechanism by which surfactants exert their activity is not entirely understood, it is, however, hypothesized that, due to their amphiphilic character, they are inserted into the fungi lipid bilayer (possibly in other microbial membranes also), where they elicit an immediate destruction effect (Kubo et al., 2001a; Koziróg et al., 2018). In that aspect, the amphiphilic character of the HT analogs presented in this work and the experimental evidence for rapid destruction of the fungus membrane allow us to reasonably assume that they act primarily as nonionic surfactants, though other additional molecular interaction could account for the final effect. However, specific structural characteristics are needed for optimal antifungal activity strongly depended from a balance between the polar hydrophilic and the non-polar hydrophobic part of the molecules. The aliphatic ester side chain (hydrophobic tail) is assumed to interfere with the cell membrane, while the acidity of the catechol group (hydrophilic head) seems to be essential for the antifungal activity. The initial contact with the lipid bilayer should involve the electrostatic interaction of the catechol system with the negatively charged phosphate groups of the membrane; thus the increased acidity of the hydroxyl groups due to the carbonyl substitution might be essential for activity. The divergence of the activity in relation to the structure of the alkyl tail indicates its vital role in the action of the compounds. The length (6 to 10 carbons) and increased volume because of alkyl substitutions are fundamental for optimum activity, whereas the substitution of the  $\alpha$ -carbon is critical for the electrostatic interaction of the hydrophilic head. The rationale for the role of the side alkyl chain is not entirely understood, and further research is needed in order to design more effective antifungal agents. Nevertheless, the fact that the HT analogs, very likely, targets the extra- cytoplasmic region could explain, as mentioned before, the lack of resistance of the described molecules.

Although recent technological advances made accessible unprecedented tools for drug discovery the development of new molecular scaffolds with demonstrated pharmacological properties is becoming slower and more expensive. Our study shows once more that the natural product chemical space provides the basis of successful design of new small-molecular-weight molecules with excellent pharmacological properties. As

HT has been shown to undergo rapid oxidation the antifungal HT analogs described here will, however, need to be tested *in vivo* in mouse models and evaluated in respect to their stability or synergistic or antagonistic effects with other antifungals.

## MATERIALS AND METHODS

### Synthesis of HT Analogs

Chemical synthesis of HT analogs is described in detail in **Supplementary Material**. HT analogs were prepared in DMSO and aliquots and kept at  $-20^{\circ}\text{C}$ . For experiments, the final concentration of DMSO in the medium was  $<0.1\%$  (v/v) and the controls received DMSO only.

### Fungal and Bacterial Strains and Growth Media

Standard “wild-type” fungal strains of *A. nidulans* (FGSC A4), *A. fumigatus* (Af293), *A. flavus* (NRRL3357), *F. oxysporum* (F3; (Christakopoulos et al., 1996)) and *C. albicans* clinical isolate from the Mycetothea ATHUM collection of Athens University<sup>1</sup> were used. The *A. nidulans* used contains a mutant allele (*veA1*) at the *velvet* locus, and a standard vitamin auxotrophy (*pabA1* for para-aminobenzoic acid requirement). All other Aspergilli used correspond to standard strains used for genome sequencing<sup>2</sup>. For *in vivo* epifluorescence microscopy, strains expressing GFP-tagged functional UapA (Pantazopoulou et al., 2007) or FurA (Fernandez-Bolanos et al., 2012) transporters were used. Standard Aspergillus Minimal and Complete Media (MM and CM) were used for growth of all fungi<sup>3</sup>. The nitrogen source used in MM was 10 mM  $\text{NaNO}_3$ . Standard bacterial strains, coming from an in-house stock, of *E. coli* (DH5a), *P. aeruginosa*, *P. fluorescens*, *Klebsiella sp.*, *B. subtilis* and *S. aureus* were used. Luria-Bertani medium was used for growth of bacterial strains.

### In vivo Evaluation of HT Analogs as Antifungals

Fungal strains of *A. nidulans*, *A. fumigatus*, *A. flavus*, *F. oxysporum* and *C. albicans* were tested for their sensitivity/resistance to different concentrations. Fungal spores of each strain were used to centrally inoculate a series of 35mm petri dishes containing Minimal Media (MM) with  $\text{NaNO}_3$  as nitrogen source, in the presence or absence of various HT analog concentrations dissolved in DMSO. Controls without HT analogs contained solely DMSO. The final concentration range tested for toxicity of analogs in this work was 0–1000  $\mu\text{M}$ . Figures shown highlight results using concentrations of 100 or 400  $\mu\text{M}$ . Growth was followed after for 4–6 days, at 25 or  $37^{\circ}\text{C}$  and different pH. Approximate concentrations leading to 50% reduction in the radial diameter of the growing colonies, ( $IC_{50}$ ), after 2–4 days (depending on the fungus) were recorded for each

<sup>1</sup><http://mycetothea.biol.uoa.gr/>

<sup>2</sup><http://genome.jgi.doe.gov/programs/fungi/index.jsf>

<sup>3</sup><http://www.fgsc.net>



analog. Approximate concentrations leading to no evident fungal growth were also recorded. For the non-filamentous fungus *C. albicans* we performed sensitivity tests in both liquid and solid cultures. For solid cultures testing, cells from a fresh liquid culture (O.D.<sub>600nm</sub> = 0.5 at 10 nm) were streaked on standard CM containing 100–400  $\mu$ M of HT analogs, and incubated for 2 days at 37°C. For liquid culture testing, HT analogs were added, at 100–400  $\mu$ M, at the start of the exponential phase (O.D.<sub>600nm</sub> = 0.5 at 10 nm) of a *C. albicans* culture and O.D.<sub>600nm</sub> measurements were recorded hourly for the next 6 h, and after 24 h. All final values shown in this article are averages of at least three independent experiments with no significant SD variation (e.g., <20%).

### **In vivo Evaluation of HT Analogs as Antibacterial Agents**

Standard bacterial strains of *E. coli* (DH5a), *P. aeruginosa*, *P. fluorescens*, *Klebsiella* sp., *B. subtilis* and *S. aureus* were tested for their sensitivity/resistance to different concentrations of compounds, which proved to act as antifungals. These tests were carried out either in solid or liquid LB media, in the presence or absence of 100–500  $\mu$ M of HT analogs. Tests on solid media were performed by recording single colony growth after bacterial streaking. Tests in liquid media were performed by recording O.D.<sub>600nm</sub> values (10 nm) after addition of 100–200  $\mu$ M of compounds in fresh exponentially growing bacterial cultures (O.D.<sub>600nm</sub> = 0.2–0.4), and comparing these values to control cultures with no analogs. All final values shown in this article are averages of at least three independent experiments (SD variation <15%).

### **In vivo Epifluorescence Microscopy**

Samples for standard inverted epifluorescence microscopy of *A. nidulans* strains were prepared as previously described (Martzoukou et al., 2017). In brief, germlings were incubated in sterile 35 mm  $\mu$ -dishes, high glass bottom (*ibidi*, Germany) in 2 ml liquid MM with NaNO<sub>3</sub> as nitrogen source and the necessary vitamin supplements for 20 h at 25°C. DMSO (0.1%) or HT analogs (final concentrations tested 37.5 or 100  $\mu$ M) dissolved in DMSO were added in samples under the microscope. Control samples where only DMSO was added (0.03–0.01%) were also evaluated. Images were taken before and immediately after the addition of the analogs (or DMSO) and for a period of up to 30 min. The strains used express the UapA or the FurA transporter as protein fluorescent markers specific for the plasma membrane (Pantazopoulou et al., 2007; Kryptou et al., 2015). Calcuofluor white staining of cell wall was as described in Martzoukou et al. (2017). Images were obtained with an AxioCa m HR R3 camera using the Zen lite 2012 software. Contrast adjustment, area selection and color combining were made using the Zen 2012 software. Images exported as tiffs were annotated and further processed in Adobe Photoshop CS4 Extended version 11.0.2 software for brightness adjustment, rotation and alignment.

### **Transport Measurements**

Kinetic analysis-<sup>3</sup>H]-xanthine (21.6 Ci/mmol, Moravek Biochemicals, CA, United States) uptake in MM was assayed in germinating conidiospores of *A. nidulans* concentrated at 10<sup>7</sup> conidiospores/100  $\mu$ L, at 37°C, pH 6.8, as previously described (Kryptou and Diallinas, 2014). Initial velocities were measured at 1 min of incubation with concentrations of 0.2–2.0  $\mu$ M of [<sup>3</sup>H]-xanthine at the polarity maintenance stage (3–4 h, 130 rpm).

### **Mutagenesis**

Two U.V. mutagenesis experiments, with lethality rates of 84.1 and 93.7%, were performed at a standard distance of 20 cm from an Osram HNS30 UV-B/C lamp. 10<sup>9–10</sup> conidiospores of a standard wild-type *A. nidulans* strain or a strain possessing a dual transposition system based on the *Minos* element (Evangelinos et al., 2015) were irradiated for 4 min and subsequently plated on MM plus nitrate medium containing 400  $\mu$ M of the HT analog 15. No colonies appeared after 1-week incubation at 25°C.

### **Toxicity in a Mammalian Cell Line**

N2A mouse neuroblastoma cells were used, in a standard MTT assay (Berridge et al., 2005), to test the whether antifungal HT analogs elicit cytotoxicity in a standard mammalian cell line. Analogs 15 or 16 were tested at two concentrations (100 or 400  $\mu$ M) as described in materials and methods. The N2A mouse neuroblastoma cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) that contained 10% fetal bovine serum, 1% of penicillin and streptomycin in 96-well plates at a density of 15,000 cells/well. A standard MTT assay was used to assess cell metabolic activity in the absence (addition of solely DMSO) and presence of HT analogs (Berridge et al., 2005). The cultures were grown for 6 days at 37°C with 5% CO<sub>2</sub>. Then the medium was changed to one containing 100 or 400  $\mu$ M of 15 or 16 and incubated for 24 h at 37°C with 5% CO<sub>2</sub>. In all cases the final concentration of DMSO was  $\leq$ 0.1%. 20  $\mu$ l of the dye MTT (2.5 mg/ml MTT in PBS) was added to each well and incubated for 4 h. The resulting formazan dye was extracted with 100  $\mu$ l isopropanol/HCl (100 ml isopropanol + 833  $\mu$ l HCl) and the absorbance was measured spectrophotometrically at a wavelength of 545 nm. Statistical analysis: All experiments were repeated three times. One-way ANOVA with Bonferroni's Multiple Comparison Test was used to evaluate the statistical significance of the differences. Statistical significance was defined as  $p < 0.05$ .

### **AUTHOR CONTRIBUTIONS**

GD performed and designed the biological experiments, analyzed the results and wrote the manuscript. NR performed the biological experiments and made relevant figures. VT and IZ performed the biological experiments. IK and AK performed the chemical experiments. EM and AS analyzed the results. IKK designed the chemical synthesis, analyzed the results and wrote the manuscript.



## FUNDING

This work was partly supported by a “Stavros Niarchos Foundation” research grant and by “Fondation Santé.”

## ACKNOWLEDGMENTS

This work has been released as a pre-print at BioRxiv (Diallinas et al., 2018). We thank Dr. Joseph Meletiadis, Assistant Professor

## REFERENCES

- Arendrup, M. C. (2014). Update on antifungal resistance in *Aspergillus* and *Candida*. *Clin. Microbiol. Infect.* 20(Suppl. 6), 42–48. doi: 10.1111/1469-0691.12513
- Belmonte-Reche, E., Martínez-García, M., Peñalver, P., Gómez-Pérez, V., Lucas, R., Gamarro, F., et al. (2016). Tyrosol and hydroxytyrosol derivatives as antityrosinase and antileishmanial agents. *Eur. J. Med. Chem.* 119, 132–140. doi: 10.1016/j.ejmech.2016.04.047
- Bernini, R., Crisante, F., Barontini, M., Tofani, D., Balducci, V., and Gambacorta, A. (2012). Synthesis and structure/antioxidant activity relationship of novel catecholic antioxidant structural analogues to hydroxytyrosol and its lipophilic esters. *J. Agric. Food Chem.* 60, 7408–7416. doi: 10.1021/jf301131a
- Bernini, R., Gilardini Montani, M. S., Merendino, N., Romani, A., and Velotti, F. (2015). Hydroxytyrosol-derived compounds: a basis for the creation of new pharmacological agents for cancer prevention and therapy. *J. Med. Chem.* 58, 9089–9107. doi: 10.1021/acs.jmedchem.5b00669
- Berridge, M. V., Herst, P. M., and Tan, A. S. (2005). Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol. Annu. Rev.* 11, 127–152. doi: 10.1016/S1387-2656(05)11004-7
- Bisignano, G., Tomaino, A., Lo Cascio, R., Crisafi, G., Uccella, N., and Saija, A. (1999). On the in-vitro antimicrobial activity of oleuropein and hydroxytyrosol. *J. Pharm. Pharmacol.* 51, 971–974. doi: 10.1211/0022357991773258
- Capasso, R., Evidente, A., Schivo, L., Orru, G., Marcialis, M. A., and Cristinzio, G. (1995). Antibacterial polyphenols from olive oil mill waste waters. *J. Appl. Bacteriol.* 79, 393–398. doi: 10.1111/j.1365-2672.1995.tb03153.x
- Chillemi, R., Sciuto, S., Spatafora, C., and Tringali, C. (2010). “Hydroxytyrosol lipophilic analogues: synthesis, radical scavenging activity and human cell oxidative damage protection,” in *Olives and Olive Oil in Health and Disease Prevention*, eds V. Preedy and R. Watson (New York, NY: Academic Press), 1233–1243. doi: 10.1016/B978-0-12-374420-3.00135-2
- Christakopoulos, P., Nerinckx, W., Kekos, D., Macris, B., and Claeysens, M. (1996). Purification and characterization of two low molecular mass alkaline xylanases from *Fusarium oxysporum* F3. *J. Biotechnol.* 51, 181–189. doi: 10.1016/0168-1656(96)01619-7
- Cuenca-Estrella, M. (2014). Antifungal drug resistance mechanisms in pathogenic fungi: from bench to bedside. *Clin. Microbiol. Infect.* 20(Suppl. 6), 54–59. doi: 10.1111/1469-0691.12495
- Di Santo, R. (2010). Natural products as antifungal agents against clinically relevant pathogens. *Nat. Prod. Rep.* 27, 1084–1098. doi: 10.1039/b914961a
- Diallinas, G., Rafailidou, N., Kalpaktsi, I., Komianou, A. C., Tsovali, V., Zantza, I., et al. (2018). Hydroxytyrosol (HT) analogues act as potent antifungals by direct disruption of the fungal cell membrane. *bioRxiv* [Preprint]. doi: 10.1101/350025
- Evangelinos, M., Anagnostopoulos, G., Karvela-Kalogeraki, I., Stathopoulou, P. M., Scazzocchio, C., and Diallinas, G. (2015). Minos as a novel Tc1/mariner-type transposable element for functional genomic analysis in *Aspergillus nidulans*. *Fungal Genet. Biol.* 81, 1–11. doi: 10.1016/j.fgb.2015.05.007
- Fairlamb, A. H., Gow, N. A., Matthews, K. R., and Waters, A. P. (2016). Drug resistance in eukaryotic microorganisms. *Nat. Microbiol.* 1:16092. doi: 10.1038/nmicrobiol.2016.92
- Fernandez-Bolanos, J. G., Lopez, O., Fernandez-Bolanos, J., and Rodriguez-Gutierrez, G. (2008). Hydroxytyrosol and derivatives: isolation, synthesis

of Microbiology in the Medical School of University of Athens, for Amphotericin B and Michalis Papadourakis for help in preliminary uptake experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02624/full#supplementary-material>

- and biological properties. *Curr. Org. Chem.* 12, 442–463. doi: 10.2174/138527208784083888
- Fernandez-Bolanos, J. G., Lopez, O., Lopez-Garcia, M. A., and Marsset, A. (2012). “Biological properties of hydroxytyrosol and its derivatives,” in *Olive Oil Constituents, Quality, Health Properties and Bioconversions*, ed. D. Boskou (Rijeka: InTech Europe), 375–396.
- Grasso, S., Siracusa, L., Spatafora, C., Renis, M., and Tringali, C. (2007). Hydroxytyrosol lipophilic analogues: enzymatic synthesis, radical scavenging activity and DNA oxidative damage protection. *Bioorg. Chem.* 35, 137–152. doi: 10.1016/j.bioorg.2006.09.003
- Köhler, J. R., Hube, B., Puccia, R., Casadevall, A., and Perfect, J. R. (2017). Fungi that infect humans. *Microbiol. Spectr.* 5:FUNK-0014-2016. doi: 10.1128/microbiolspec.FUNK-0014-2016
- Koolen, H. H., Pral, E. M., Alfieri, S. C., Marinho, J. V., Serain, A. F., Hernández-Tasco, A. J., et al. (2017). Antiprotozoal and antioxidant alkaloids from *Alternanthera littoralis*. *Phytochemistry* 134, 106–113. doi: 10.1016/j.phytochem.2016.11.008
- Korukluoglu, M., Sahan, Y., and Yigit, A. (2008). Antifungal properties of olive leaf extracts and their phenolic compounds. *J. Food Saf.* 28, 76–87. doi: 10.1111/j.1745-4565.2007.00096.x
- Koziróg, A., Brycki, B., and Pielech-Przybylska, K. (2018). Impact of cationic and neutral gemini surfactants on conidia and hyphal forms of *Aspergillus brasiliensis*. *Int. J. Mol. Sci.* 19:E873. doi: 10.3390/ijms19030873
- Kozubek, A., Zarnowski, R., Stasiuk, M., and Gubernator, J. (2001). Natural amphiphilic phenols as bioactive compounds. *Cell. Mol. Biol. Lett.* 6, 351–355.
- Kryptou, E., and Diallinas, G. (2014). Transport assays in filamentous fungi: kinetic characterization of the UapC purine transporter of *Aspergillus nidulans*. *Fungal Genet. Biol.* 63, 1–8. doi: 10.1016/j.fgb.2013.12.004
- Kryptou, E., Evangelidis, T., Bobonis, J., Pittis, A. A., Gabaldón, T., Scazzocchio, C., et al. (2015). Origin, diversification and substrate specificity in the family of NCS1/FUR transporters. *Mol. Microbiol.* 96, 927–950. doi: 10.1111/mmi.12982
- Kubo, I., Fujita, K., and Lee, S. H. (2001a). Antifungal mechanism of polygodial. *J. Agric. Food Chem.* 49, 1607–1611. doi: 10.1021/jf000136g
- Kubo, I., Xiao, P., and Fujita, K. (2001b). Antifungal activity of octyl gallate: structural criteria and mode of action. *Bioorg. Med. Chem. Lett.* 11, 347–350. doi: 10.1016/S0960-894X(00)00656-9
- Li, F. S., and Weng, J. K. (2017). Demystifying traditional herbal medicine with modern approach. *Nat. Plants* 3:17109. doi: 10.1038/nplants.2017.109
- Low, C. Y., and Rotstein, C. (2011). Emerging fungal infections in immunocompromised patients. *F1000 Med. Rep.* 3:14. doi: 10.3410/M3-14
- Manna, C., Migliardi, V., Sannino, F., De Martino, A., and Capasso, R. (2005). Protective effects of synthetic hydroxytyrosol acetyl derivatives against oxidative stress in human cells. *J. Agric. Food Chem.* 53, 9602–9607. doi: 10.1021/jf058110i
- Martzoukou, O., Amillis, S., Zervakou, A., Christoforidis, S., and Diallinas, G. (2017). The AP-2 complex has a specialized clathrin-independent role in apical endocytosis and polar growth in fungi. *eLife* 6:e20083. doi: 10.7554/eLife.20083
- McCarthy, M. W., Kontoyiannis, D. P., Cornely, O. A., Perfect, J. R., and Walsh, T. J. (2017). Novel agents and drug targets to meet the challenges of resistant fungi. *J. Infect. Dis.* 216(Suppl. 3), S474–S483. doi: 10.1093/infdis/jix130

- Medina, E., Brenes, M., Romero, C., García, A., and de Castro, A. (2007). Main antimicrobial compounds in table olives. *J. Agric. Food Chem.* 55, 9817–9823. doi: 10.1021/jf0719757
- Medina, E., de Castro, A., Romero, C., and Brenes, M. (2006). Comparison of the concentrations of phenolic compounds in olive oils and other plant oils: correlation with antimicrobial activity. *J. Agric. Food Chem.* 54, 4954–4961. doi: 10.1021/jf0602267
- Medina-Martínez, M. S., Truchado, P., Castro-Ibáñez, I., and Allende, A. (2016). Antimicrobial activity of hydroxytyrosol: a current controversy. *Biosci. Biotechnol. Biochem.* 80, 801–810. doi: 10.1080/09168451.2015.1116924
- Odds, F. C., Brown, A. J., and Gow, N. A. (2003). Antifungal agents: mechanisms of action. *Trends Microbiol.* 11, 272–279. doi: 10.1016/S0966-842X(03)00117-3
- Pan, S. Y., Zhou, S. F., Gao, S. H., Yu, Z. L., Zhang, S. F., Tang, M. K., et al. (2013). New perspectives on how to discover drugs from herbal medicines: CAM's outstanding contribution to modern therapeutics. *Evid. Based Complement. Altern. Med.* 2013:627375. doi: 10.1155/2013/627375
- Pantazopoulou, A., and Diallinas, G. (2007). Fungal nucleobase transporters. *FEMS Microbiol. Rev.* 31, 657–675. doi: 10.1111/j.1574-6976.2007.00083.x
- Pantazopoulou, A., Lemuh, N. D., Hatzinikolaou, D. G., Drevet, C., Cecchetto, G., Scazzocchio, C., et al. (2007). Differential physiological and developmental expression of the UapA and AzgA purine transporters in *Aspergillus nidulans*. *Fungal Genet. Biol.* 44, 627–640. doi: 10.1016/j.fgb.2006.10.003
- Pereira, A. P., Ferreira, I. C., Marcelino, F., Valentão, P., Andrade, P. B., Seabra, R., et al. (2007). Phenolic compounds and antimicrobial activity of olive (*Olea europaea* L. Cv. Cobrançosa) leaves. *Molecules* 12, 1153–1162. doi: 10.3390/12051153
- Robles-Almazan, M., Pulido-Moran, M., Moreno-Fernandez, J., Ramirez-Tortosa, C., Rodriguez-Garcia, C., Quiles, J. L., et al. (2018). Hydroxytyrosol: bioavailability, toxicity, and clinical applications. *Food Res. Int.* 105, 654–667. doi: 10.1016/j.foodres.2017.11.053
- Tafesh, A., Najami, N., Jadoun, J., Halahlih, F., Riepl, H., and Azaizeh, H. (2011). Synergistic antibacterial effects of polyphenolic compounds from olive mill wastewater. *Evid. Based Complement. Altern. Med.* 2011:431021. doi: 10.1155/2011/431021
- Tasleem, A., Bhosale, J. D., Naresh, K., Mandal, T. K., Bendre, R. S., Lavekar, G. S., et al. (2009). Natural products-antifungal agents derived from plants. *J. Asian Nat. Prod. Res.* 11, 621–638. doi: 10.1080/10286020902942350
- Teodoro, G., Ellepola, K., Senevirante, C. J., and Koga-Ito, C. Y. (2015). Potential use of phenolic acids as anti-candida agents: a review. *Front. Microbiol.* 6:1420. doi: 10.3389/fmicb.2015.01420
- Thielmann, J., Kohnen, S., and Hauser, C. (2017). Antimicrobial activity of *Olea europaea* Linné extracts and their applicability as natural food preservative agents. *Int. J. Food Microbiol.* 251, 48–66. doi: 10.1016/j.ijfoodmicro.2017.03.019
- Walter, W. M. Jr., Flemming, H. P., and Etchells, J. L. (1973). Preparation of antimicrobial compounds by hydrolysis of oleuropein from green olives. *Appl. Microbiol.* 26, 773–776.
- Zoric, N., Horvat, I., Kopjar, N., Vucemilovic, A., Kremer, D., Tomic, S., et al. (2013). Hydroxytyrosol expresses antifungal activity in vitro. *Curr. Drug Targets* 14, 992–998. doi: 10.2174/13894501113149990167
- Zoric, N., Kopjar, N., Bobnjarić, I., Horvat, I., Tomić, S., and Kosalec, I. (2016). Antifungal activity of oleuropein against *Candida albicans*—the in vitro study. *Molecules* 21:1631. doi: 10.3390/molecules21121631

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Diallinas, Rafailidou, Kalpaktsi, Komianou, Tsouvali, Zantza, Mikros, Skaltsounis and Kostakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*

Maria Botou <sup>1</sup> | Vassilis Yaelis <sup>1</sup> | Panayiota Lazou<sup>1</sup> | Iliana Zantza<sup>2</sup> | Konstantinos Papakostas<sup>1</sup> | Vassiliki Charalambous<sup>1</sup> | Emmanuel Mikros<sup>2</sup> | Emmanouil Fletmetakis<sup>3</sup> | Stathis Frillingos <sup>1</sup>

<sup>1</sup>Laboratory of Biological Chemistry, Department of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece

<sup>2</sup>Division of Pharmaceutical Chemistry, Department of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, Athens, Greece

<sup>3</sup>Laboratory of Molecular Biology, Department of Biotechnology, Agricultural University of Athens, Athens, Greece

## Correspondence

Stathis Frillingos, Laboratory of Biological Chemistry, Department of Medicine, School of Health Sciences, University of Ioannina, Ioannina 45110, Greece.  
Email: efriligo@uoi.gr

## Funding information

Stavros Niarchos Foundation; European Social Fund

## Abstract

*Sinorhizobium (Ensifer) meliloti* is a model example of a soil alpha-proteobacterium which induces the formation of nitrogen-fixing symbiotic nodules on the legume roots. In contrast to all other rhizobacterial species, *S. meliloti* contains multiple homologs of nucleobase transporter genes that belong to NAT/NCS2 family (Nucleobase-Ascorbate Transporter/Nucleobase-Cation Symporter-2). We analyzed functionally all (six) relevant homologs of *S. meliloti* 1,021 using *Escherichia coli* K-12 as a host and found that five of them are high-affinity transporters for xanthine (SmLL9), uric acid (SmLL8, SmLL9, SmX28), adenine (SmVC3, SmYE1), guanine (SmVC3), or hypoxanthine (SmVC3). Detailed analysis of substrate profiles showed that two of these transporters display enlarged specificity (SmLL9, SmVC3). SmLL9 is closely related in sequence with the xanthine-specific XanQ of *E. coli*. We subjected SmLL9 to rationally designed site-directed mutagenesis and found that the role of key binding-site residues of XanQ is conserved in SmLL9, whereas a single amino-acid change (S93N) converts the xanthine/uric-acid transporter SmLL9 to a xanthine-preferring variant, due to disruption of an essential hydrogen bond with the C8 oxygen of uric acid. The results highlight the presence of several different purine nucleobase transporters in *S. meliloti* and imply that the purine transport might be important in the nodule symbiosis involving *S. meliloti*.

## KEYWORDS

active transport, molecular dynamics, purines, *Sinorhizobium*, site-directed mutagenesis

## 1 | INTRODUCTION

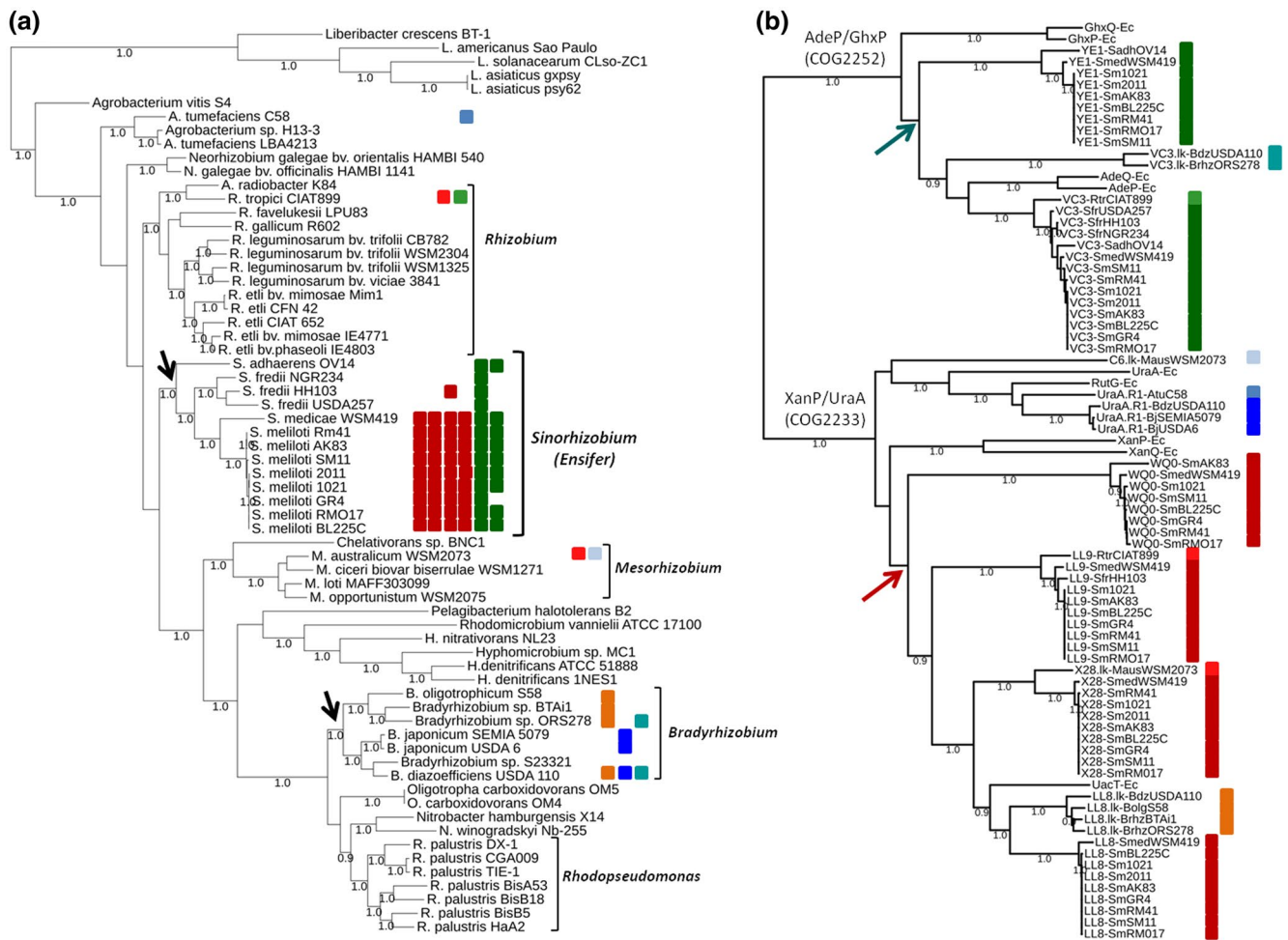
Soil bacteria that can form nitrogen-fixing symbiotic nodules on legumes, collectively called rhizobacteria or rhizobia, have a global impact on the sustainable agriculture and the cycling of nitrogen in the biosphere. In the symbiotic nodules, rhizobia are differentiated to organelle-like bacteroids, which populate the infected plant cells and specialize in reducing the atmospheric nitrogen to ammonium. The efficiency of this process depends on many factors, a central one being the proper exchange of metabolites and ions between the

two symbionts to allow integration of metabolism in the symbiotic nodule. This exchange is mediated by the membrane transporters of both rhizobial and plantal origin. A few well-known transporters are used in all types of symbiotic nitrogen fixation to transport dicarboxylic acids from the plant to rhizobium (to feed the TCA cycle) and reduced (fixed) nitrogen, mostly as ammonia, from the rhizobium to the plant. Apart from these unequivocally needed transporters, distinct groups of rhizobia and legume symbionts can use additional important metabolite transporters most of which remain unstudied to date (Trepolilli *et al.*, 2012; Udvardi and Poole, 2013; Clarke *et al.*,

2014; Liu *et al.*, 2018). Such metabolite uptake systems might be needed for optimal nitrogen fixation in certain plant-rhizobia symbioses but are absent in others. Understanding the function of such transporters would reinforce efforts to design and optimize synthetic plant-microbe interactions (Mus *et al.*, 2016) for increasing the efficiency of nitrogen fixation in the root nodules of legumes or even extending it to non-leguminous crops (Rogers and Oldroyd, 2014; Mus *et al.*, 2016; Bueno Batista and Dixon, 2019).

Nucleobase/nucleoside transport is a rather neglected theme in symbiotic nitrogen fixation systems. Nucleobases imported by cytoplasmic membrane transporters can be used in purine-pyrimidine salvage pathways to feed the intracellular nucleotide pool or in the catabolic pathways to provide the cell with nitrogen sources

but also with carbon sources (Girke *et al.*, 2014; Izaguirre-Mayoral *et al.*, 2018). Nucleobase/nucleoside transport and metabolism have not been reported to play a central role in the symbiotic nitrogen fixation, but it cannot be excluded that some types of rhizobia-legume symbiosis benefit energetically from such processes. Our interest on this theme was prompted by the observation that one particular rhizobial species, *Sinorhizobium (Ensifer) meliloti*, harbors multiple homologs of nucleobase transporter genes that are absent in other rhizobia (see later, Figure 1). These homologs belong to NAT/NCS2 family (Nucleobase-Ascorbate Transporter/ Nucleobase-Cation Symporter-2), the largest and evolutionarily broader nucleobase transporter family that contains ion gradient-driven transporters with diverse substrate preferences, ranging



**FIGURE 1** Phylogenetic analysis of the NAT/NCS2 homologs of *Sinorhizobium (Ensifer) meliloti*. (a) Distribution of NAT/NCS2 genes in Rhizobiales. Fully sequenced genomes were used for tree construction. The phylogenetic tree was based on concatenated sequences from four housekeeping genes, *atpD* (720 bp), *recA* (876bp), *rpoB* (1164 bp) partial sequences and 16S rRNA (~1,470 bp), with an overall length ~4,230 bp, from 67 genomes of Rhizobiales retrieved from JGI IMG/M. The analysis was performed with the Maximum likelihood (ML) method using the Kimura's two parameter model, as implemented in MEGA7. NAT/NCS2 homologs are indicated with *small boxes* color-coded according to their phylogenetic clustering (*red* and *orange*, Cluster 1; *blue*, Cluster 2 (subgroup R1); *light blue*, Cluster 6; *green*, Subfamily 2) (see text and Table S1 for further details). (b) Phylogenetic analysis of the NAT/NCS2 homologs of the organisms shown in (a) and the related functionally known homologs of *Escherichia coli* (GhxP, GhxQ, AdeP, AdeQ, UraA, RutG, XanP, XanQ, UacT). The analysis involved the amino acid sequences and was performed with ML method using the Jones-Taylor-Thornton matrix-based model as implemented in MEGA7. The percentage (given as a decimal) of trees in which the associated taxa clustered together is shown next to the branches. Branches with a bootstrap value < 0.7 have been deleted and bootstrap values < 0.9 are not shown in these presentations [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



from purine or pyrimidine permeases in bacteria, fungi, plants, and animals to Na<sup>+</sup>-dependent L-ascorbic acid transporters in human and other mammals (Gournas *et al.*, 2008; Yamamoto *et al.*, 2010; Frillingos, 2012; Bürzle *et al.*, 2013; Girke *et al.*, 2014; Botou *et al.*, 2018; Chaliotis *et al.*, 2018). Phylogenetically, all homologs of *S. meliloti* cluster together with known purine nucleobase transporters from *Escherichia coli* or other proteobacteria (Frillingos, 2012; Papakostas and Frillingos, 2012; Papakostas *et al.*, 2013; Karena *et al.*, 2015). Most other rhizobia are devoid of NAT/NCS2 homologs or contain only homologs of the related transporter family NCS1 that do not share significant similarity to the known purine transporters. Thus, it is conceivable that the purine nucleobase transporters may be needed specifically in certain types of nodule symbiosis involving *S. meliloti*.

In this work, we analyze functionally all (six) NAT/NCS2 transporter homologs of *S. meliloti* strain 1,021 through the heterologous expression in *E. coli* K-12 and show that five of them are high-affinity purine transporters with diverse specificity and kinetics, representing a spectrum of different and complementary purine substrate profiles. Two of these transporters display broader specificity than their *E. coli* homologs, similar to fungal homologs. We subjected one of them (SmLL9) to rationally designed site-directed mutagenesis and delineated key amino acids of the binding-site region that are essential for its function or contribute to its broader substrate profile. The putative involvement of the above purine nucleobase transporters of *S. meliloti* in nitrogen-fixing symbiosis is discussed.

## 2 | RESULTS

### 2.1 | *S. meliloti* contains putative purine-transporter genes of the NAT/NCS2 family which are absent in other rhizobia

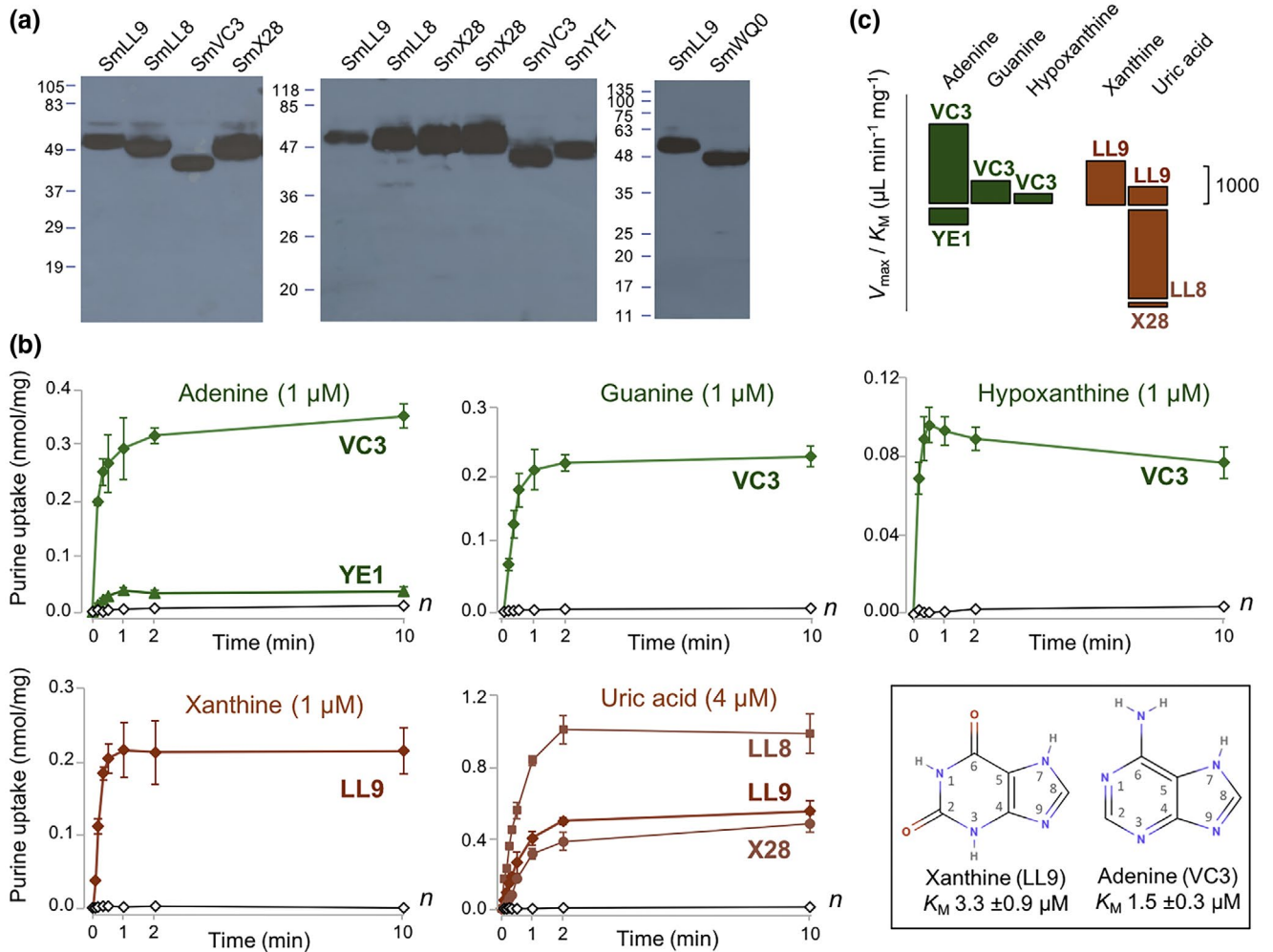
Rhizobia are a polyphyletic group of Bacteria. More than 110 rhizobial species belonging to 14 genera are known, of which the majority are alpha-proteobacteria classified in the order Rhizobiales and some are beta-proteobacteria in the order Burkholderiales (Berrada and Fikri-Benbrahim, 2014; Andrews and Andrews, 2017). About 80% of these species belong to four genera, *Rhizobium*, *Sinorhizobium* (*Ensifer*), *Mesorhizobium*, or *Bradyrhizobium*, of Rhizobiales. We performed a phylogenetic analysis of 61 genomes including 39 well-known strains of these rhizobia and 22 related non-rhizobial strains of Rhizobiales based on the multilocus sequence analysis (MLSA) of four housekeeping genes (as in Mousavi *et al.*, 2015) and investigated the distribution of putative nucleobase transporter homologs in these genomes (Figure 1a). The results show lack of NAT/NCS2 homologs in most of the Rhizobiales except in the genus *Sinorhizobium* (*Ensifer*), in particular all strains of *S. meliloti* or *S. medicae* which contain five or six NAT/NCS2 homologs per strain, and sporadic presence in *Bradyrhizobium* (containing 0–3 homologs per strain), *Rhizobium* (*R. tropici*), and *Mesorhizobium* (*M. australicum*). A phylogenetic tree of these rhizobial NAT/NCS2 homologs reveals

that four *Sinorhizobium* homologs classify in NAT/COG2233 (in Cluster C1\_Xanthine-Uric Acid; see Chaliotis *et al.*, 2018) and two in AzgA-like/COG2252, whereas the *Bradyrhizobium* homologs classify in different clusters of these subfamilies (Figure 1b). Of the six *Sinorhizobium* homologs, only SmVC3 (AzgA-like/COG2252) is consistently present throughout the genus, whereas SmX28, SmLL8, SmLL9, SmWQ0 (NAT/COG2233), and SmYE1 (AzgA-like/COG2252) are confined in *S. meliloti* and *S. medicae*. Based on the known substrate preferences of the related *E. coli* K-12 homologs (Karatzas and Frillingos, 2005; Papakostas and Frillingos, 2012; Papakostas *et al.*, 2013; Karena *et al.*, 2015; Botou *et al.*, 2018), it appears that all *Sinorhizobium* homologs are phylogenetically relevant to the purine nucleobase transporters which prefer uric acid (NAT/COG2233) or adenine (AzgA-like/COG2252) as substrates (Figure 1b).

A comprehensive analysis of the distribution of the above transporter genes in Rhizobiales (Figure S1; Table S1) (Botou *et al.*, 2020) shows that the four NAT/COG2233 homologs are consistently present only in *Sinorhizobium* (almost invariably in *S. meliloti*) with SmLL9 appearing also in *Methylobacterium* and *Ochrobactrum* and SmLL8-related sequences in few strains of *Bradyrhizobium*, whereas a different set of NAT/COG2233 homologs (belonging to cluster R1 which includes the pyrimidine transporter RutG; Botou *et al.*, 2018) appears in *Bradyrhizobium*, *Agrobacterium*, *Methylobacterium*, *Ochrobactrum*, and *Brucella*. Of the AzgA-like/COG2252 homologs, SmVC3 is present consistently in *Sinorhizobium*, *Methylobacterium*, *Ochrobactrum*, and *Brucella*, whereas SmYE1 is present only in *Sinorhizobium* (Figure S1; Table S1).

### 2.2 | The NAT/NCS2 genes of *S. meliloti* 1,021 encode purine nucleobase transporters with distinct and complementary substrate profiles

The NAT/NCS2 genes of *S. meliloti* strain 1,021 were PCR-mobilized from the genome, tagged with a biotin-acceptor domain at C-terminus and transferred to vector pT7-5 for expression in *E. coli* K-12, followed by immunoblot analysis of the corresponding protein levels in the cytoplasmic membrane and analysis of their transport activities for several putative substrates (adenine, guanine, hypoxanthine, xanthine, uric acid, uracil, thymine, and cytosine). The results are presented in Figure 2 and Table 1. All of the proteins are expressed in the *E. coli* membrane (Figure 2a). None of the encoded proteins was found to transport any pyrimidine nucleobase. Of the purine nucleobases, homologs that belong to NAT/COG2233 transport with high-affinity xanthine and/or uric acid (except SmWQ0 which does not seem to transport any nucleobase), and homologs that belong to AzgA-like/COG2252 transport with high-affinity adenine and/or guanine/hypoxanthine. SmLL9 is a high-affinity transporter for xanthine ( $K_M = 3.3 \mu\text{M}$ ) but also transports uric acid with lower affinity ( $K_M = 27.3 \mu\text{M}$ ); SmLL8 and SmX28 are uric acid transporters ( $K_M = 3.5$  and  $21.4 \mu\text{M}$ , respectively); SmVC3 transports with high-affinity adenine ( $K_M = 1.5 \mu\text{M}$ ), guanine ( $K_M = 2.7 \mu\text{M}$ ), and hypoxanthine ( $K_M = 4.5 \mu\text{M}$ ); SmYE1 is specific for adenine ( $K_M = 1.6 \mu\text{M}$ ).



**FIGURE 2** Substrate preferences of the NAT/NCS2 transporters of *Sinorhizobium (Ensifer) meliloti*. *Escherichia coli* K-12 strains expressing the corresponding NAT/NCS2 genes of *S. meliloti* 1,021 through pT7-5/-BAD were assayed for the transporter protein levels in the membrane (a) and for active transport of various substrates (b, c). (a) Membrane fractions prepared from *E. coli* T184 expressing the indicated *S. meliloti* homologs were subjected to SDS-PAGE (12%) and western blotting using HRP-conjugated streptavidin. Representative blots are shown. Each lane contains 50  $\mu$ g of total membrane protein. Also shown are the migration positions of molecular mass standards run in parallel. (b) *E. coli* JW3692, JW4025 or T184 expressing the indicated *S. meliloti* homologs were assayed for uptake of [ $^3$ H]-adenine and [ $^3$ H]-hypoxanthine, [ $^3$ H]-guanine, or [ $^3$ H]-xanthine and [ $^{14}$ C]-uric acid, respectively, at the indicated concentration, at 25°C. The levels achieved in each case with the remaining transporters (indistinguishable from cells harboring pT7-5/-BAD alone) are shown in the time curves labeled with an *n*. Each data point represents the mean and SD from triplicate measurements. The major substrates of SmLL9 (xanthine) and SmVC3 (adenine) are shown as an inset. (c) Comparison of the transport efficiencies ( $V_{max}/K_M$ ) of SmVC3, SmYE1, SmLL9, SmLL8 and SmX28 for each of the five purine nucleobases (see Table 1 for further details) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Considering both  $K_M$  and  $V_{max}$ , the results show that SmLL8 is the major transporter for uric acid (5- and 20-fold higher  $V_{max}/K_M$  relative to SmLL9 and SmX28, respectively) and SmVC3 is the major transporter for adenine (4-fold higher  $V_{max}/K_M$  relative to SmYE1) (Figure 2b; Table 1). Of the two homologs that transport multiple nucleobase substrates, SmLL9 prefers xanthine (3-fold higher  $V_{max}/K_M$  relative to uric acid) and SmVC3 prefers adenine (4- and 9-fold higher  $V_{max}/K_M$  relative to guanine and hypoxanthine, respectively) (Figure 2b,c; Table 1). The xanthine uptake activity of SmLL9 and the adenine uptake activity of SmVC3 are abolished by the protonophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) implying that they are dependent on  $H^+$  symport (Table S2).

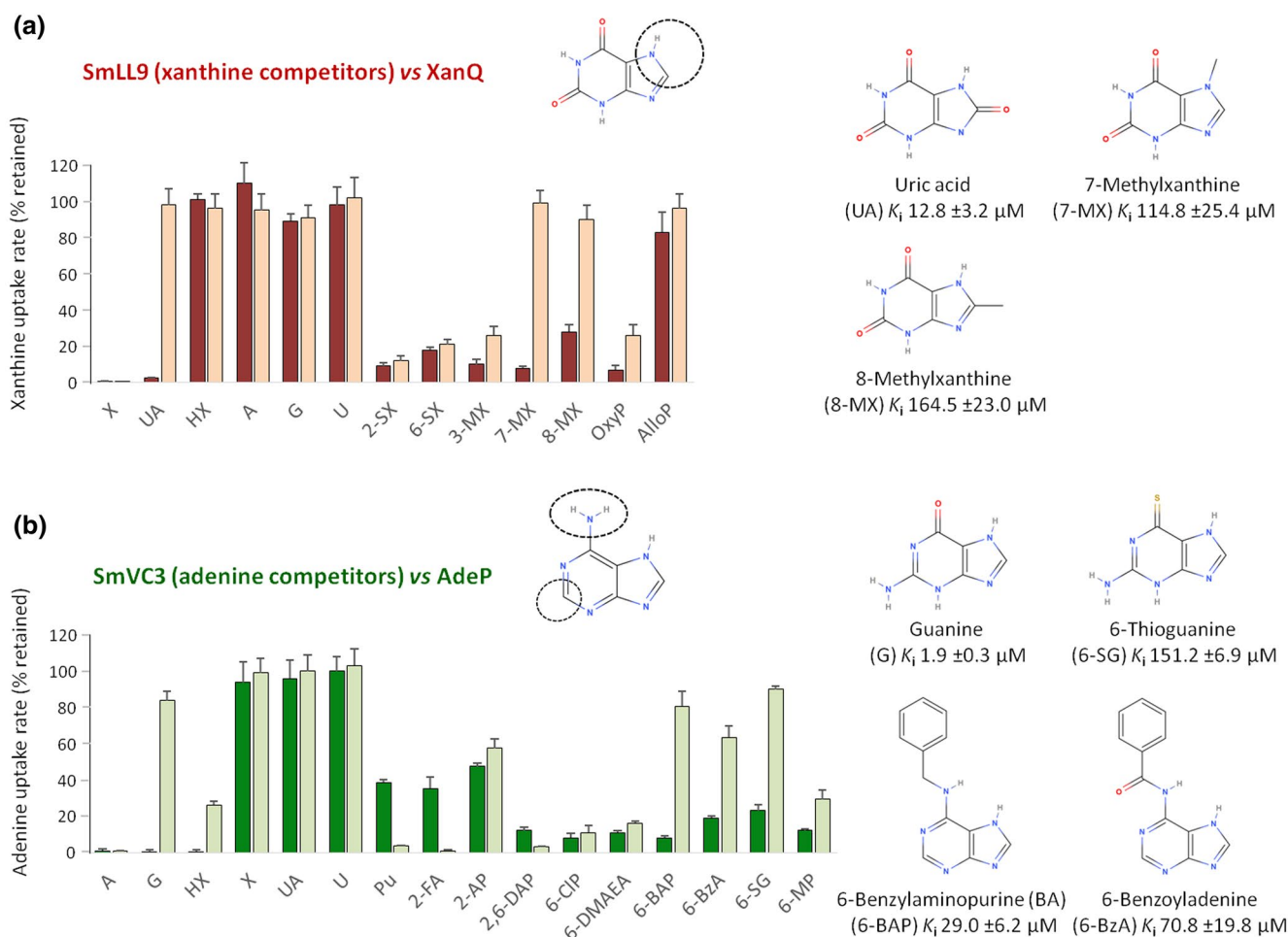
### 2.3 | SmVC3 and SmLL9 are enlarged-specificity purine nucleobase transporters

The two homologs that transport multiple purine substrates were subjected to transport competition assays using a range of purine analogs as putative competitors. Known *E. coli* transporters were used as controls in each case, as appropriate. Thus, SmLL9 was compared with the xanthine-specific XanQ (Karatzas and Frillingos, 2005) for competition against xanthine transport and SmVC3 was compared with the adenine-specific AdeP (also known as PurP) (Papakostas *et al.*, 2013) for competition against adenine transport. The results are presented in Figure 3 and Table 2. As shown, SmLL9 recognizes

**TABLE 1**  $K_M$  and  $V_{max}$  values of the *Sinorhizobium (Ensifer) meliloti* NAT/NCS2 transporters

Transporter	Substrate	$K_M$ ( $\mu\text{M}$ )	$V_{max}$ ( $\text{nmol min}^{-1} \text{mg}^{-1} \text{protein}$ )	$V_{max}/K_M$ ( $\mu\text{l min}^{-1} \text{mg}^{-1}$ )
SmLL9	Xanthine	3.3 ( $\pm 0.9$ )	3.9 ( $\pm 0.3$ )	1,182
	Uric acid	27.3 ( $\pm 6.9$ )	13.3 ( $\pm 0.9$ )	487
SmLL8	Uric acid	3.5 ( $\pm 1.3$ )	8.3 ( $\pm 0.6$ )	2,370
SmX28	Uric acid	21.4 ( $\pm 6.8$ )	2.3 ( $\pm 0.2$ )	107
SmVC3	Adenine	1.5 ( $\pm 0.3$ )	2.6 ( $\pm 0.2$ )	1,733
	Guanine	2.7 ( $\pm 0.4$ )	1.3 ( $\pm 0.1$ )	482
	Hypoxanthine	4.5 ( $\pm 0.8$ )	0.9 ( $\pm 0.1$ )	200
SmYE1	Adenine	1.6 ( $\pm 0.4$ )	0.7 ( $\pm 0.2$ )	437

Note: *Escherichia coli* T184 expressing SmLL9, SmLL8, or SmX28 was assayed for initial rates of [ $^3\text{H}$ ]-xanthine (0.1–100  $\mu\text{M}$ ) or [ $^{14}\text{C}$ ]-uric acid (1–400  $\mu\text{M}$ ) uptake at 25°C. *E. coli* JW3692 expressing SmVC3 or SmYE1 was assayed for initial rates of [ $^3\text{H}$ ]-adenine (0.1–40  $\mu\text{M}$ ) or [ $^3\text{H}$ ]-hypoxanthine (0.1–40  $\mu\text{M}$ ) uptake at 25°C. *E. coli* JW4025 expressing SmVC3 was assayed for initial rates of [ $^3\text{H}$ ]-guanine (0.1–40  $\mu\text{M}$ ) uptake at 25°C. Kinetic parameters were determined from non-linear regression fitting to the Michaelis–Menten equation. The values shown are the means of at least three determinations with SD shown in parentheses.



**FIGURE 3** Ligand specificity profiles of SmLL9 and SmVC3. (a) *Escherichia coli* T184 expressing SmLL9 (red) or XanQ (orange) were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) uptake in the absence or presence of the indicated xanthine analogs (1 mM). Each bar represents the means of three determinations with SD shown. SmLL9 ligands that are not recognized by XanQ are shown on the right (see Table 2 for further details). (b) *E. coli* JW3692 expressing SmVC3 (green) or AdeP (light green) were assayed for initial rates of [ $^3\text{H}$ ]-adenine (0.1  $\mu\text{M}$ ) uptake in the absence or presence of the indicated adenine/guanine analogs (1 mM). Each bar represents the mean and SD of three determinations. SmVC3 ligands that are not recognized by AdeP are shown on the right (see Table 2 for further details). 2,6-DAP, 2,6-diaminopurine; 2-AP, 2-aminopurine; 2-FA, 2-fluoroadenine; 2-SX, 2-thioxanthine; 3-MX, 3-methylxanthine; 6-BAP, 6-benzylaminopurine; 6-BzA, 6-benzoyladenine; 6-CIP, 6-chloropurine; 6-DMAEA, 6-(dimethylamino)ethyladenine; 6-MP, 6-mercaptopurine; 6-SG, 6-thioguanine; 6-SX, 6-thioxanthine; 7-MX, 7-methylxanthine; 8-MX, 8-methylxanthine; A, adenine; AlloP, allopurinol; G, guanine; HX, hypoxanthine; OxyP, oxypurinol; Pu, purine; : U, uracil; UA, uric acid; X, xanthine [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 2** Substrate/ligand profiles of SmVC3 and SmLL9

$K_i$ ( $\mu\text{M}$ ) for inhibition of [ $^3\text{H}$ ]-xanthine (X) uptake by other nucleobases/analogs			$K_i$ ( $\mu\text{M}$ ) for inhibition of [ $^3\text{H}$ ]-adenine (A) uptake by other nucleobases/analogs		
Nucleobase or analog as competitor	SmLL9	XanQ	Nucleobase or analog as competitor	SmVC3	AdeP
Xanthine ( $K_M$ )	3.3 ( $\pm 0.9$ )	4.0 ( $\pm 0.3$ )	Adenine ( $K_M$ )	1.5 ( $\pm 0.2$ )	1.2 ( $\pm 0.1$ )
Uric acid	12.8 ( $\pm 3.2$ )	ND	Guanine	1.9 ( $\pm 0.3$ )	ND
Hypoxanthine	ND	ND	Hypoxanthine	1.4 ( $\pm 0.2$ )	150.0 ( $\pm 13.5$ )
Adenine	ND	ND	Xanthine	ND	ND
Guanine	ND	ND	Uric acid	ND	ND
Uracil	ND	ND	Uracil	ND	ND
2-thioX	77.8 ( $\pm 15.3$ )	91.0 ( $\pm 11.1$ )	Purine (P)	126.5 ( $\pm 36.8$ )	3.4 ( $\pm 0.4$ )
6-thioX	219.8 ( $\pm 9.2$ )	41.5 ( $\pm 5.3$ )	2-FA	72.5 ( $\pm 20.4$ )	38.3 ( $\pm 3.9$ )
3-methylX	22.3 ( $\pm 4.4$ )	70.9 ( $\pm 9.5$ )	2-aminoP	ND	ND
7-methylX	114.8 ( $\pm 25.4$ )	ND	2,6-diaminoP	48.3 ( $\pm 9.4$ )	8.8 ( $\pm 3.2$ )
8-methylX	164.5 ( $\pm 23.0$ )	ND	6-chloroP	188.7 ( $\pm 21.6$ )	55.0 ( $\pm 10.2$ )
Oxypurinol	34.1 ( $\pm 3.2$ )	37.0 ( $\pm 6.8$ )	6-DMAEA	37.6 ( $\pm 14.3$ )	70.4 ( $\pm 8.6$ )
Allopurinol	ND	ND	6-BAP (BA)	29.0 ( $\pm 6.2$ )	ND
			6-benzoylA	70.8 ( $\pm 19.8$ )	ND
			6-thioguanine	151.2 ( $\pm 6.9$ )	ND
			6-mercaptoP	20.0 ( $\pm 4.8$ )	188.0 ( $\pm 25.9$ )

Note: *Escherichia coli* T184 expressing SmLL9 or XanQ or *E. coli* JW3692 expressing SmVC3 or AdeP was assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) or [ $^3\text{H}$ ]-adenine (0.1  $\mu\text{M}$ ) uptake, respectively, in the absence or presence of a concentration range (0.01–1,000  $\mu\text{M}$ ) of the indicated competitors. The data were analyzed with Prism7 to obtain  $\text{IC}_{50}$  and thereof  $K_i$  values. Also given are the  $K_M$  values of the permeases for xanthine and adenine. The values shown are the means of at least three determinations with SD shown in parentheses. ND, not detectable inhibition of activity by more than 50% at any of the concentrations tested.

Abbreviations: 2-FA, 2-fluoroadenine; 6-BAP, 6-benzylaminopurine (BA); 6-DMAEA, 6-(dimethylamino)ethyladenine; A, adenine; P, purine; X, xanthine.

with high-affinity uric acid, 8-methylxanthine, and 7-methylxanthine which are not ligands for XanQ, highlighting that it displays broader specificity toward analogs at the imidazole moiety of purine substrate. SmVC3 recognizes with high-affinity guanine, 6-thioguanine, 6-benzylaminopurine (BA), or 6-benzoyladenine which are not ligands for AdeP as well as hypoxanthine and 6-mercaptapurine which are recognized by AdeP with 10- to 100-fold lower affinity; purine or 2-fluoroadenine (2-FA) which are high-affinity ligands of AdeP are recognized by SmVC3 with lower affinities (Table 2). Thus, SmVC3 displays broader specificity than AdeP in recognizing substitutions at positions C2 or C6 of the pyrimidine moiety of the purine substrate (Figure 3); it also displays broader specificity than the *E. coli* guanine-hypoxanthine transporter GhxP which does not recognize adenine or any of the analogs retaining an amino group at C6 (Papakostas *et al.*, 2013).

## 2.4 | Homology modeling of SmLL9 and its differences from XanQ

The xanthine/uric acid transporter SmLL9 falls in the NAT/COG2233 Cluster 1 (C1\_Xanthine-Uric Acid) (Chalotiotis *et al.*, 2018) which also includes two very well studied homologs, the xanthine-specific

XanQ of *E. coli* (see Figure 1b) (Frillingos, 2012; Karena *et al.*, 2015) and the structurally known xanthine/uric transporter UapA from *Aspergillus nidulans* (Alguel *et al.*, 2016; Diallinas, 2016). It also includes several other functionally known homologs (Figure S2). Of the above, SmLL9 is the first bacterial homolog shown to transport with high affinity both xanthine and uric acid, since the previously well studied homologs from *E. coli* are specific for xanthine (XanQ) or uric acid (UacT). In addition, SmLL9 classifies in a different clade relative to the one including XanQ and UapA (Karena *et al.*, 2015). SmLL9 and XanQ share a rather limited sequence homology overall (29.8% identity based on MUSCLE alignment; 26.2% identity, 43.9% similarity, 18.3% gaps based on the EMBOSS Needle tool), but most residues found to be crucial for function in XanQ (based on Cys-scanning analysis; Karena *et al.*, 2015) are retained in both transporters (Table S3). Prompted by these observations, we set to examine structure-function relationships of SmLL9 in relation to XanQ using molecular dynamics and site-directed mutagenesis at key amino acid positions.

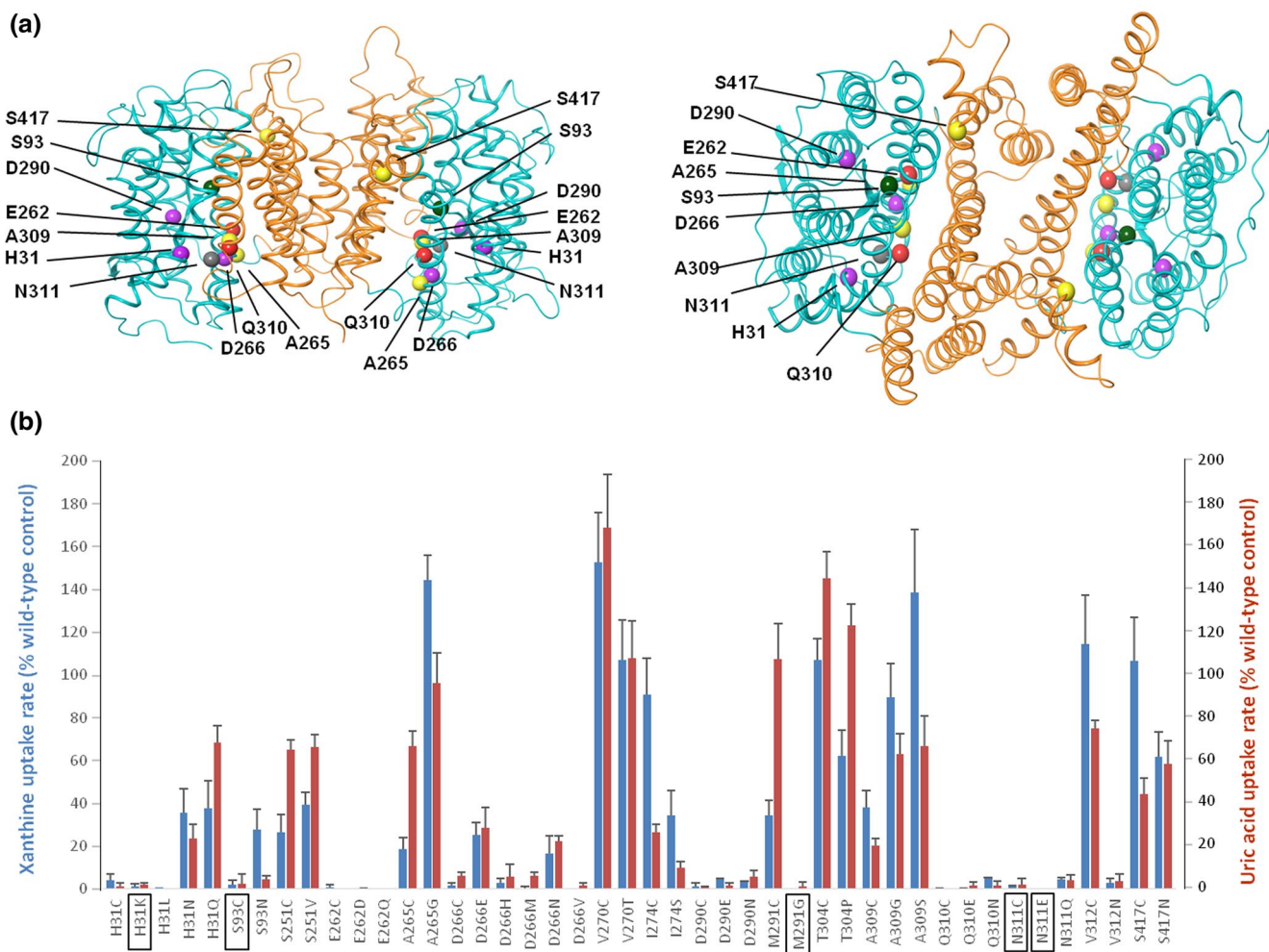
Application of the HHpred search tool for homology detection and structure prediction for SmLL9 ranks UapA (Alguel *et al.*, 2016) (PDB ID: 5I6C) first with 100% probability ( $p = 2.5e-40$ ), 24% identity, and 0.38 similarity and the same holds for XanQ with 100% probability ( $p = 6e-43$ ), 30% identity, and 0.54 similarity. Both



transporters also exhibit homology to UraA (Yu *et al.*, 2017) (PDB ID: 5XLS) with 100% probability ( $p = 8.9e-40$ ), 24% identity and 0.42 similarity for SmLL9, and 100% probability ( $p = 2e-37$ ), 24% identity and 0.38 similarity for XanQ. Structural models of both SmLL9 and XanQ were constructed and specific protein-substrate interactions were simulated accordingly. The structural models were generated as monomers by homology to the UapA crystal structure of *Aspergillus nidulans* (Alguel *et al.*, 2016) corresponding to cytoplasm-facing conformers. The construction of the dimeric form of both models was performed by 3D superposition on the UapA transporter dimer (Figure 4a). Further, the structures were subjected to 40 ns molecular dynamics simulation in the *apo* form embedded in DPPC lipid bilayer and solvating the membrane by explicit water using GROMACS (Abraham *et al.*, 2018) software and charmm36 (Huang *et al.*, 2017)

force field. The RMSD of the C $\alpha$ -carbons of all helices between starting and ending structures did not exceed 1 Å, indicating the stability of the theoretical model.

According to the UapA template the 3D models of SmLL9 and XanQ are dimeric consisting of 14 transmembrane segments (TMs) for each unit (Figure 4a). Each transporter unit structure is divided in a core domain (TMs 1–4 and 8–11) and a gate domain (TMs 5–7 and 12–14). The transmembrane helices have the same length except TM9 which is shorter for SmLL9 by one turn. The loops connecting the TMs are also of the same length on average with few exceptions: the loop between TM3–4 in XanQ is longer by 10 amino acids, and the loop between TM4–5 is shorter by three amino acids. Compared to UapA, three exoplasmic loops of both bacterial homologs are shorter, by at least 10 residues (TM3–4), 15 residues (TM5–6), and 12



**FIGURE 4** Homology modeling and site-directed mutagenesis of SmLL9. (a) Dimeric structural model of SmLL9 as viewed parallel to the membrane plane (*left*) and from the cytoplasmic side (*right*). Different colors highlight the core and the gate domain of each protomer. Residues found to be important with site-directed mutagenesis are shown as spheres in red (irreplaceable, involved in substrate binding), gray (irreplaceable, involved in H-bond interactions in the vicinity of the binding site), purple (H-bond interactions in the core domain), green (crucial for specificity), and yellow (sensitive to inactivation by site-directed alkylation). (b) Site-directed mutagenesis of SmLL9. *Escherichia coli* T184 expressing the corresponding mutants from pT7-5/SmLL9-BAD were assayed for uptake of [ $^3$ H]-xanthine (1  $\mu$ M) (*blue bars*) or [ $^{14}$ C]-uric acid (4  $\mu$ M) (*red bars*) and the deduced transport rates were expressed as percentages of the rate obtained with wild-type SmLL9. Values obtained with cells harboring pT7-5/BAD alone had been subtracted from the sample measurements in all cases. Each bar represents the mean of at least three determinations with SD shown. Mutants shown in boxes attain negligible protein levels in the *E. coli* membrane (see Figure S6) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

residues (TM13–14). Notably, the loop between TM3–4 in the bacterial homologs does not contain a pair of cysteines shown to form a disulfide bond that is needed for correct intracellular folding in UapA and possibly other fungal homologs (Alguel *et al.*, 2016) and also lacks the 10-residue segment between these cysteines. Of the cytoplasmic loops, the loop between TM4–5 in SmLL9 is shorter by one residue relative to UapA distorting the small helix (H1) that forms between TM4 and TM5 in the UapA crystal structure (Alguel *et al.*, 2016).

In both SmLL9 and XanQ modeled structures the ionized residues are distributed reasonably on the protein surface, most of them at the cytoplasmic and periplasmic sides or along the pore in the protein interior. Positive charges are mostly located in the cytoplasm-facing loops and negative charges in the periplasm-facing loops. Both models were subjected to the calculations of pKa for carboxylates in the binding cavity using PDB2PQR version 2.0.0 (Dolinsky *et al.*, 2004) server. D266 in SmLL9 and the corresponding D276 in XanQ were found to have pKa values higher than 8 and they were considered as protonated in the docking calculations. In the case of XanQ the theoretical pKa of E272 was 7.5 and thus docking calculations were performed twice using the protonated and charged carboxylate form, whereas E262 in SmLL9 was found to have a pKa of 7 and was considered as negatively charged in all calculations.

According to the 3D alignment of the constructed models and the UapA and UraA crystal structures, residues crucial for the binding and transport of substrate, specificity, kinetics, and transporter function in UapA and UraA that are located in the vicinity of substrate (4.5 Å) in TMs 1, 3, 8, and 10 exhibit a high similarity between the four transporters SmLL9, XanQ, UapA, and UraA (Table S4). In the core domain, a hydrogen bond network is formed in SmLL9 between the side chains of Q90, Q75, Q305, D290, and T263 connecting TMs 2, 3, 8, 9, and 10, similar to the one described for XanQ (Karena *et al.*, 2015). The network consists of two parts, the one between Q90–Q75–Q305 (Figure S3a) and the other between D290, T263, and the backbone of Q305. During the MD simulations the D290 side chain reorients toward the center of the core domain and forms strong H-bonds with Q75 and Q90 while keeping the interaction with T263 (Figure S4). In UapA (Alguel *et al.*, 2016), the residues corresponding to Q90 and Q305 are M151 and M403 which seem to stabilize through the hydrophobic interactions as the two S-CH<sub>3</sub> moieties are almost parallel in a distance of about 4.6 Å and, in addition, there is a weak H-bond interaction between Q130 (corresponding to Q75) and T405 (A307 in SmLL9) which is absent in SmLL9 (Figure S3b). Thus, the H-bond network of the core domain of the bacterial homologs is replaced in UapA with a looser network that seems to include different types of interactions. MD simulations showed also that the Asn311 side chain in SmLL9 exhibits a relative flexibility as it is re-oriented during the simulated period forming interactions with H31 from the one side and D266 on the other side (Figure S5).

Induced fit docking (IFD) calculations showed that SmLL9 and XanQ bind xanthine in a similar manner, whereas uric acid is bound by XanQ in a reverse orientation and less optimally relative to SmLL9. The results from the IFD calculations are presented in the final part of the Section 2 (see later, Figure 7).

## 2.5 | Conservation of the functional role of key binding-site residues between SmLL9 and XanQ

In an attempt to address the role of key amino acid residues of SmLL9 and understand the major requirements for its functionality and broader specificity than XanQ, we subjected SmLL9 to site-directed mutagenesis at positions corresponding to important residues of XanQ (Table S3) that are either involved in substrate binding or participate in H-bond interactions at the periphery of the binding site or belong to TMs 8 and 10 that are central in the coordination of substrate binding (Figure 4a). Residues of SmLL9 that are conserved in XanQ were replaced by amino acids with similar side-chain properties and additionally, if the similar side-chain mutants retained significant activity, they were subjected to more extensive mutagenesis; residues differing from XanQ were replaced by the amino acid occupying the corresponding position in XanQ. In addition, all the residues examined were mutagenized to Cys, taking advantage of the fact that SmLL9 does not contain any native Cys in its sequence.

Our mutagenesis data reveal that most SmLL9 mutants display remarkably similar properties with the corresponding XanQ mutants studied previously (see Table S3). The substrate-binding residues Glu262 (TM8) and Gln310 (TM10) are strictly irreplaceable for function, since mutants E262C, E262D, E262Q, Q310C, Q310E, Q310N are inactive for either xanthine or uric acid transport (Figure 4b); the same has been shown for the corresponding XanQ mutants at Glu272 (Karena and Frillingos, 2009) and Gln324 (Karatza *et al.*, 2006), with the difference that Q324N and Q324E retain some activity but are grossly impaired in the affinity for xanthine or xanthine analogs (Karatza *et al.*, 2006). Also irreplaceable are the residues Asp290 (TM9), which is central in a H-bond network stabilizing the core domain in XanQ (Karena *et al.*, 2015), and Asn311 (TM10), which is involved in interactions with His31 and Asp266 (Figure S4) at the periphery of the binding site, since mutants D290C, D290E, D290N, N311C, N311E, N311Q are inactive (Figure 4b); the same has been shown for the corresponding XanQ mutants at Asp304 (Karena and Frillingos, 2009) and Asn325 (Karatza *et al.*, 2006). Interestingly, N311C and N311E are not detected in the *E. coli* membrane (Figure S6), pointing to a more austere role of Asn311 in affecting the protein stability/folding relative to Asn325 in XanQ. His31 (TM1) is replaceable by the H-bonding residues Q or N but irreplaceable by C, L, or K (Figure 4b) and, additionally, mutant H31K is not detected in the membrane (Figure S6); strikingly similar data have been obtained with the corresponding XanQ mutants (Karena and Frillingos, 2009), the only difference being that H31C and H31L in XanQ display low transport activity but impaired affinity for xanthine or xanthine analogs (Karena and Frillingos, 2009). Asp266 (TM8) is replaceable only by E or N and irreplaceable by C, H, M, or V (Figure 4b), in close resemblance with the properties of Asp276 in XanQ (Mermelekas *et al.*, 2010); a difference is that D276E in XanQ displays high activity, in contrast to D266E in SmLL9 which transports at a rate of 20% relative to wild type (Figure 4b) pointing to a more crucial role of this carboxyl group in SmLL9. In the uric acid transporter UacT which also belongs to the NAT/COG2233

Cluster C1\_Xanthine-Uric Acid (Chalotiotis *et al.*, 2018) and is closely related with SmLL8 (Figure 1b), all residues corresponding to the above six crucial residues of SmLL9 were found to be irreplaceable (Papakostas and Frillingos, 2012). Of the corresponding residues in the fungal UapA, the two substrate-binding residues Glu356 (TM8) (Papageorgiou *et al.*, 2008) and Gln408 (TM10) (Koukaki *et al.*, 2005) are also essential and Asp388 (TM9) is also functionally irreplaceable (Papageorgiou *et al.*, 2008), but the other residues are less stringently related with activity: Asn409 (TM10) is replaceable by Ala, Ser or Gln with retention of wild-type properties, and only N409D results in loss of function (Koukaki *et al.*, 2005), His86 (TM1) is linked with defects in folding and targeting to the plasma membrane as shown with the low-activity mutants H86A, H86K, or H86D, while H86N is active and indistinguishable in function from wild type (Pantazopoulou and Diallinas, 2006), and Asp360 (TM8) is replaceable by Ala retaining wild-type properties (Kosti *et al.*, 2012).

Ser93 (TM3) is irreplaceable by C and S93C is not detected in the membrane (Figure S6), whereas S93N retains the significant

xanthine uptake activity but marginal transport activity for uric acid (Figure 4b). Kinetic analysis (Table 3) shows that S93N transports uric acid with low affinity (3-fold higher  $K_M$  relative to wild type) and low efficiency (10-fold lower  $V_{max}/K_M$  for uric acid than for xanthine, representing a 4-fold increase of this ratio over wild type). In the xanthine-specific XanQ, the corresponding Asn93 is also irreplaceable by C with low levels of the N93C mutant in the membrane (Karena and Frillingos, 2011), whereas N93S and N93A display high transport activity for xanthine and marginally detectable transport activity for uric acid as well (Karena and Frillingos, 2011). In the related uric acid transporter UacT, replacement of the corresponding Thr100 by Ala results in significant ability to transport xanthine in addition to uric acid (Papakostas and Frillingos, 2012). In the fungal UapA, replacements of the corresponding Ser154 with A or N lead to defects in transport activity and kinetic changes relative to wild type indicative of a role of Ser154 in specificity. S154N shows 5-fold improved affinity for xanthine without affecting affinity for uric acid and S154A shows 7-fold decreased affinity for uric acid without

**TABLE 3**  $K_M$  and  $V_{max}$  values of SmLL9 and mutants

Permease	Xanthine (X) uptake			Uric acid (UA) uptake			Ratio
	$K_M$ ( $\mu\text{M}$ )	$V_{max}$ ( $\text{nmol min}^{-1} \text{mg}^{-1}$ )	$V_{max}/K_M$ ( $\mu\text{l min}^{-1} \text{mg}^{-1}$ )	$K_M$ ( $\mu\text{M}$ )	$V_{max}$ ( $\text{nmol min}^{-1} \text{mg}^{-1}$ )	$V_{max}/K_M$ ( $\mu\text{l min}^{-1} \text{mg}^{-1}$ )	$V_{max}/K_M$ X/UA
SmLL9(wt)	3.3 ( $\pm 0.9$ )	3.9 ( $\pm 0.3$ )	1,182	27.3 ( $\pm 6.9$ )	13.3 ( $\pm 0.9$ )	487	2.43
H31N	9.7 ( $\pm 2.4$ )	2.1 ( $\pm 0.2$ )	216	37.5 ( $\pm 13.1$ )	9.6 ( $\pm 1.4$ )	256	0.85
H31Q	17.3 ( $\pm 1.5$ )	2.2 ( $\pm 0.1$ )	127	82.2 ( $\pm 14.7$ )	17.7 ( $\pm 1.7$ )	215	0.59
S93N	12.9 ( $\pm 1.0$ )	1.3 ( $\pm 0.1$ )	101	76.6 ( $\pm 17.4$ )	0.8 ( $\pm 0.1$ )	10	10.10
S251C	4.0 ( $\pm 0.4$ )	0.4 ( $\pm 0.0$ )	100	69.2 ( $\pm 13.8$ )	7.3 ( $\pm 0.7$ )	105	0.95
S251V	3.1 ( $\pm 0.9$ )	0.7 ( $\pm 0.1$ )	226	51.8 ( $\pm 16.8$ )	9.1 ( $\pm 1.4$ )	176	1.28
A265C	7.3 ( $\pm 1.4$ )	1.1 ( $\pm 0.1$ )	151	39.0 ( $\pm 13.1$ )	9.6 ( $\pm 1.2$ )	246	0.61
A265G	2.3 ( $\pm 0.8$ )	3.0 ( $\pm 0.3$ )	1,304	39.6 ( $\pm 6.4$ )	8.3 ( $\pm 0.6$ )	210	6.21
D266E	5.0 ( $\pm 0.7$ )	0.6 ( $\pm 0.0$ )	120	239.6 ( $\pm 31.4$ )	9.3 ( $\pm 0.9$ )	39	3.08
D266N	0.5 ( $\pm 0.1$ )	0.04 ( $\pm 0.01$ )	80	98.0 ( $\pm 28.8$ )	3.1 ( $\pm 0.6$ )	32	2.40
V270C	2.0 ( $\pm 0.6$ )	2.9 ( $\pm 0.2$ )	1,450	62.5 ( $\pm 10.8$ )	20.2 ( $\pm 1.7$ )	323	4.20
V270T	2.5 ( $\pm 0.5$ )	1.2 ( $\pm 0.1$ )	480	26.1 ( $\pm 4.3$ )	4.0 ( $\pm 0.3$ )	153	3.14
I274C	1.8 ( $\pm 0.4$ )	0.6 ( $\pm 0.1$ )	334	21.0 ( $\pm 6.7$ )	1.4 ( $\pm 0.2$ )	67	4.98
I274S	1.3 ( $\pm 0.2$ )	0.1 ( $\pm 0.0$ )	77	21.7 ( $\pm 9.1$ )	0.4 ( $\pm 0.1$ )	18	4.28
M291C	35.8 ( $\pm 6.0$ )	14.5 ( $\pm 1.3$ )	405	93.5 ( $\pm 6.9$ )	26.2 ( $\pm 0.9$ )	280	1.45
T304C	1.9 ( $\pm 0.7$ )	1.1 ( $\pm 0.1$ )	579	49.0 ( $\pm 8.8$ )	19.6 ( $\pm 1.5$ )	400	1.45
T304P	35.4 ( $\pm 4.2$ )	14.6 ( $\pm 1.0$ )	412	70.6 ( $\pm 5.2$ )	23.3 ( $\pm 0.9$ )	330	1.25
A309C	3.0 ( $\pm 0.9$ )	0.3 ( $\pm 0.0$ )	100	27.6 ( $\pm 5.5$ )	1.6 ( $\pm 0.1$ )	58	1.72
A309G	6.7 ( $\pm 1.0$ )	5.8 ( $\pm 0.3$ )	866	74.6 ( $\pm 6.2$ )	10.1 ( $\pm 0.4$ )	135	6.41
A309S	2.0 ( $\pm 0.7$ )	2.3 ( $\pm 0.2$ )	1,150	49.2 ( $\pm 7.4$ )	7.6 ( $\pm 0.6$ )	154	7.47
V312C	2.9 ( $\pm 1.0$ )	3.7 ( $\pm 0.4$ )	1,276	87.3 ( $\pm 3.9$ )	21.9 ( $\pm 0.5$ )	251	5.08
S417C	2.7 ( $\pm 0.9$ )	0.4 ( $\pm 0.1$ )	148	43.4 ( $\pm 5.8$ )	2.4 ( $\pm 0.2$ )	55	2.69
S417N	2.9 ( $\pm 1.3$ )	0.9 ( $\pm 0.2$ )	310	42.8 ( $\pm 7.4$ )	5.0 ( $\pm 0.4$ )	117	2.65

Note: *Escherichia coli* T184 expressing the corresponding constructs were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (0.1–250  $\mu\text{M}$ ) or [ $^{14}\text{C}$ ]-uric acid (1–400  $\mu\text{M}$ ) uptake at 25°C. Kinetic parameters were determined from non-linear regression fitting to the Michaelis–Menten equation. The values shown are the means of three determinations with SD shown in parentheses. Xanthine or UA uptake was not detectable with mutants H31C, H31K, H31L, S93C, E262C, E262D, E262Q, D266C, D266H, D266M, D266V, D290C, D290E, D290N, M291G, Q310C, Q310E, Q310N, N311C, N311E, N311Q, or V312N, in any of the concentrations tested.

affecting affinity for xanthine. These mutants are defective in transport of both substrates but their relative transport efficiencies for xanthine and uric acid have not been tested (Amillis *et al.*, 2011).

Of the other residues, most mutants display high transport activities for both uric acid and xanthine (Figure 4b) and near wild-type kinetics (Table 3), with few exceptions mentioned below. Met291 (TM9) is irreplaceable by G and M291G is not detected in the membrane (Figure S6), but M291C is highly active (Figure 4b), albeit with significantly lower affinities (higher  $K_M$ ) relative to wild type for both uric acid and xanthine (Table 3); in XanQ, Gly305 at the corresponding position has been annotated as important for folding/stability and mutant G305C is not detectable in the membrane (Mermelekas *et al.*, 2010). Thr304 (TM10) is replaceable by C or P with retention of high transport activity for both xanthine and uric acid (Figure 4b), although T304P displays significantly lower affinities (higher  $K_M$ ) relative to wild type (Table 3); in XanQ, Pro318 at the corresponding position is crucial for stability and detection of the protein in the membrane (Karatzas *et al.*, 2006). Val312 (TM10) is irreplaceable by N but replaceable by C with high activity (Figure 4b) and small differences from wild type in kinetics (Table 3); the corresponding Asn326 in XanQ is replaceable by C but sensitive to inactivation by alkylation with *N*-ethylmaleimide (Karatzas *et al.*, 2006; Georgopoulou *et al.*, 2010), and Val320 in the related uric acid transporting homolog UacT is irreplaceable by N (Papakostas and Frillingos, 2012).

The Cys-replacement mutants in the natively C-less SmLL9 background offer us the possibility to examine the effects of alkylation

of these mutants by sulfhydryl-specific reagents on the transport activity. Inhibition of an active Cys-mutant by alkylation by the small and relatively hydrophobic *N*-ethylmaleimide would indicate involvement of the corresponding position in the substrate translocation pathway and modification of this inhibitory effect in the presence of substrate would indicate a position that is sensitive to the conformational alternations of the transporter or involved directly in substrate binding (Frillingos *et al.*, 1998; Kaback *et al.*, 2007). Of the nine active Cys-mutants of SmLL9, V270C, I274C, M291C, and T304C are not inhibited by *N*-ethylmaleimide (Table 4), S251C is inhibited with  $IC_{50} = 169 \mu\text{M}$ , V312C with  $IC_{50} = 47 \mu\text{M}$ , and three mutants are inhibited with  $IC_{50} = 14\text{--}20 \mu\text{M}$  which remains unchanged (A309C, S417C) or increases 3-fold (A265C) in the presence of 1 mM substrate (xanthine) (Table 4). The corresponding alkylation-sensitive mutants of XanQ display very similar responses to *N*-ethylmaleimide; V261C is the least sensitive ( $IC_{50} = 97 \mu\text{M}$ ) (Mermelekas *et al.*, 2010), N326C is inhibited with  $IC_{50} = 50 \mu\text{M}$  (Georgopoulou *et al.*, 2010), A323C with  $IC_{50} = 34 \mu\text{M}$  (Georgopoulou *et al.*, 2010), N430C with  $IC_{50} = 10 \mu\text{M}$  (Papakostas *et al.*, 2008), G275C with  $IC_{50} = 15 \mu\text{M}$  (Mermelekas *et al.*, 2010). A rigorous comparison of the *N*-ethylmaleimide effects on the xanthine transport activity of mutants A265C (TM8) and A309C (TM10) with the corresponding effects in XanQ illustrates that the sensitivity to *N*-ethylmaleimide is in the same  $IC_{50}$  range for all mutants but the SmLL9 mutants respond differently to xanthine binding (the sensitivity of A265C decreases, whereas the sensitivity of G275C in

**TABLE 4** Sensitivity of single-Cys mutants of SmLL9 to *N*-ethylmaleimide (NEM)

Permease	Xanthine uptake activity (% wt)	Activity retained with NEM (2 mM) (% untreated)	$IC_{50}$ of NEM ( $\mu\text{M}$ ) in the absence of substrate	$IC_{50}$ of NEM ( $\mu\text{M}$ ) in the presence of substrate (1 mM xanthine)
SmLL9(wt)	100	112.0 ( $\pm 1.8$ )	—	—
H31C	ND	—	—	—
S93C	ND	—	—	—
S251C	26.5 ( $\pm 8.1$ )	30.6 ( $\pm 6.3$ )	169.2 ( $\pm 54.2$ )	NA
E262C	ND	—	—	—
A265C	18.7 ( $\pm 5.5$ )	0.4 ( $\pm 0.5$ )	20.1 ( $\pm 2.0$ )	69.1 ( $\pm 5.4$ )
D266C	ND	—	—	—
V270C	152.7 ( $\pm 23.0$ )	88.5 ( $\pm 5.1$ )	—	—
I274C	90.8 ( $\pm 16.9$ )	103.9 ( $\pm 14.0$ )	—	—
D290C	ND	—	—	—
M291C	34.6 ( $\pm 7.1$ )	99.6 ( $\pm 2.1$ )	—	—
T304C	107.0 ( $\pm 9.8$ )	116.5 ( $\pm 5.0$ )	—	—
A309C	38.3 ( $\pm 7.6$ )	0.0 ( $\pm 0.0$ )	19.2 ( $\pm 2.4$ )	21.2 ( $\pm 3.2$ )
Q310C	ND	—	—	—
N310C	ND	—	—	—
V312C	114.6 ( $\pm 22.8$ )	0.2 ( $\pm 0.2$ )	47.2 ( $\pm 2.0$ )	NA
S417C	106.6 ( $\pm 20.5$ )	23.5 ( $\pm 6.2$ )	13.9 ( $\pm 2.3$ )	15.5 ( $\pm 1.4$ )

Note: *Escherichia coli* T184 expressing the corresponding constructs were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) uptake at 25°C, after a 10 min preincubation with NEM ranging from 0.4  $\mu\text{M}$  to 2 mM, in the absence of presence of xanthine (1 mM), as indicated, and compared with control cells preincubated in the absence of NEM and xanthine. The experimental details are described in Section 4. The data referring to the dose-dependence of the NEM effect were analyzed with Prism7 to obtain  $IC_{50}$  values. The values shown are the means of three to six determinations with SD shown in the parentheses. ND indicates not detectable activity. NA indicates that assays were not performed in the corresponding cases.



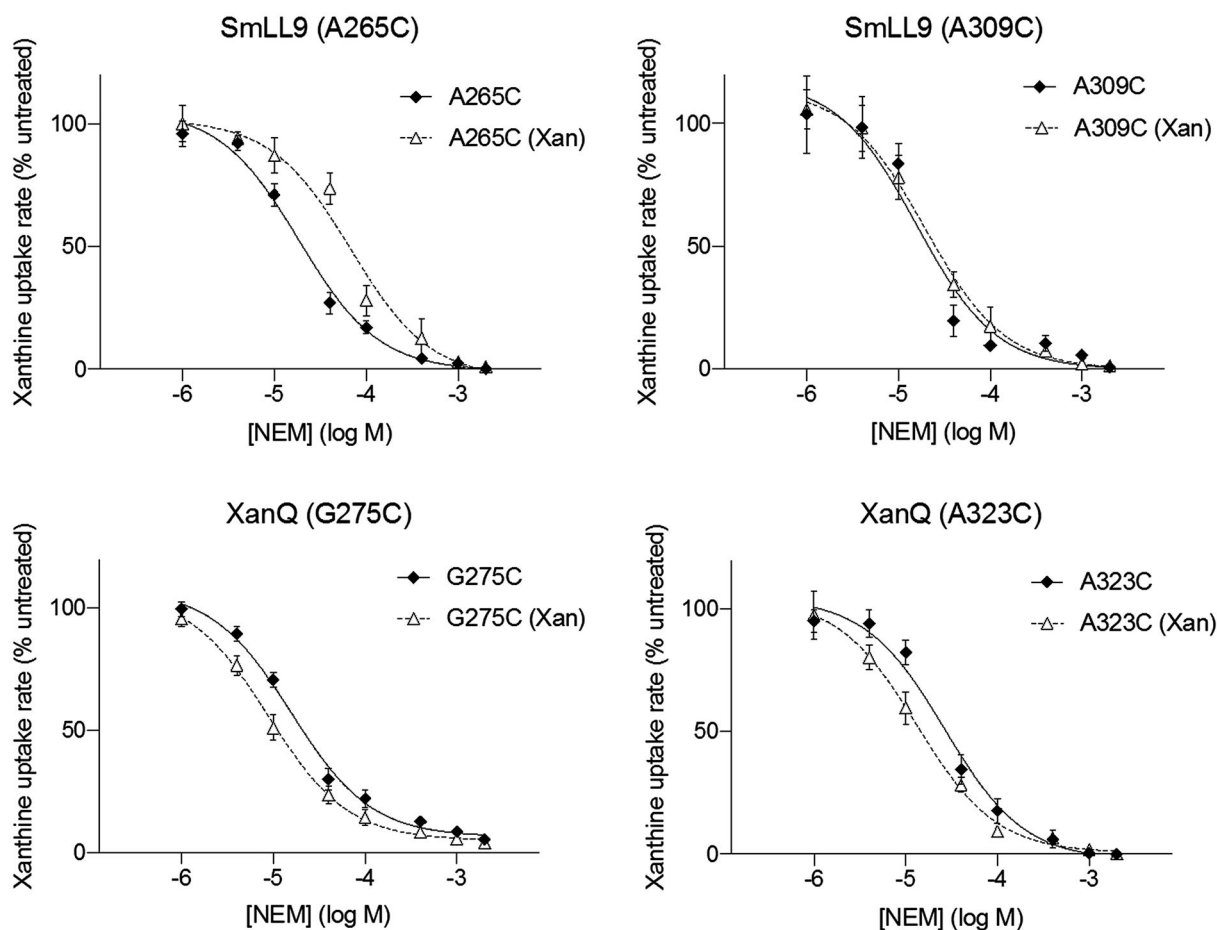
XanQ increases; the sensitivity of A309C does not change, whereas the sensitivity of A323C in XanQ increases) (Figure 5).

## 2.6 | Replacement of Ser93 with Asn converts SmLL9 to a xanthine-selective transporter

Our transport kinetics results show that several mutants display significant differences in affinity ( $K_M$  value) or efficiency ( $V_{max}/K_M$  value) from wild type for either xanthine or uric acid uptake or both (Table 3). Notably, S93N transports the uric acid with 10-fold lower efficiency relative to xanthine. Compromised efficiencies for uric acid are seen also with A265G, A309G, and A309S (6- to 7-fold lower  $V_{max}/K_M$  relative to xanthine), whereas D266E and D266N display lower efficiencies (10- to 12- and 15-fold, respectively, relative to wild type) for both the uric acid and xanthine and the same is true for A309C and I274S which transport xanthine with 12- and 15-fold lower efficiency and uric acid with 8- and 27-fold lower efficiency

relative to wild type. Six mutants (H31Q, S93N, D266E, D266N, A309G, V312C) display distinctively low affinity for uric acid (>3-fold higher  $K_M$  relative to wild type), whereas M291C and T304P have low affinity for both uric acid (2.5- to 3.5-fold higher  $K_M$ ) and xanthine (11-fold higher  $K_M$ ) relative to wild type.

Following up on the above results, we subjected the SmLL9 mutants to xanthine uptake assays in the presence of uric acid, 7-methylxanthine or 8-methylxanthine (three ligands that are not recognized by the xanthine-specific XanQ) in order to understand the effect of each replacement on the substrate profile of SmLL9. As shown in Table 5, most mutants are inhibited by each one of the three ligands with  $K_i$  values similar to wild type, but S93N displays 12-fold lower affinity (12-fold higher  $K_i$ ) for uric acid, four mutants (H31N, H31Q, A265C, T304P) are inhibited by uric acid with 3- to 4-fold higher  $K_i$ , and five mutants have lower affinity for 7-methylxanthine (2- to 2.5-fold higher  $K_i$ ) (A309C) or 8-methylxanthine (3- to 5-fold higher  $K_i$ ) (H31N, V312C) or both (H31Q, S93N). Of the above mutants, S93N stands out as the one which shows reduced affinity for uric acid



**FIGURE 5** Similar sensitivity to *N*-ethylmaleimide at key substrate-sensing positions between SmLL9 and XanQ. *Escherichia coli* T184 expressing the corresponding single-Cys mutants of SmLL9 or XanQ were analyzed for initial rates of [ $^3$ H]-xanthine (1  $\mu$ M) transport after preincubation with NEM for 10 min, at 25°C, in the absence or presence of xanthine (Xan) (1 mM), as indicated, and compared with control cells preincubated in the absence of NEM and xanthine (see Section 4 for further details). Each data point represents the mean and SD of triplicate measurements. Data were analyzed with Prism7 to obtain  $IC_{50}$  values. The results with the XanQ mutants replicate previously published results (G275C: Mermelekas *et al.*, 2010; A323C: Georgopoulou *et al.*, 2010). The  $IC_{50}$  values deduced (mean  $\pm$  SE) from the above experiments are: SmLL9 (A265C) 18.5  $\pm$  2.3  $\mu$ M, (Xan) 6.8  $\pm$  1.4  $\mu$ M; SmLL9 (A309C) 15.7  $\pm$  4.6  $\mu$ M, (Xan) 19.7  $\pm$  4.0  $\mu$ M; XanQ (G275C) 15.8  $\pm$  1.7  $\mu$ M, (Xan) 8.9  $\pm$  0.8  $\mu$ M; XanQ (A323C) 26.6  $\pm$  4.0  $\mu$ M, (Xan) 13.2  $\pm$  1.7  $\mu$ M

Permease	$K_i$ ( $\mu\text{M}$ ) for inhibition of [ $^3\text{H}$ ]-xanthine uptake by other nucleobases		
	Uric acid	7-Methylxanthine	8-Methylxanthine
SmLL9(wt)	12.8 ( $\pm$ 3.2)	114.8 ( $\pm$ 25.4)	164.5 ( $\pm$ 23.0)
H31N	46.8 ( $\pm$ 7.1)	161.8 ( $\pm$ 33.5)	443.0 ( $\pm$ 60.8)
H31Q	38.6 ( $\pm$ 4.4)	234.8 ( $\pm$ 27.3)	828.1 ( $\pm$ 147.9)
S93N	157.4 ( $\pm$ 27.3)	234.0 ( $\pm$ 43.1)	541.7 ( $\pm$ 107.7)
S251C	9.7 ( $\pm$ 0.7)	54.4 ( $\pm$ 10.2)	207.8 ( $\pm$ 33.4)
S251V	16.5 ( $\pm$ 3.6)	127.3 ( $\pm$ 18.5)	100.6 ( $\pm$ 27.2)
A265C	37.3 ( $\pm$ 4.5)	88.7 ( $\pm$ 8.1)	339.5 ( $\pm$ 40.2)
A265G	5.0 ( $\pm$ 0.5)	159.3 ( $\pm$ 30.7)	299.8 ( $\pm$ 44.6)
D266E	29.1 ( $\pm$ 4.5)	21.3 ( $\pm$ 2.4)	53.4 ( $\pm$ 4.9)
V270C	9.7 ( $\pm$ 1.0)	71.0 ( $\pm$ 7.4)	159.5 ( $\pm$ 18.5)
V270T	11.1 ( $\pm$ 1.5)	87.6 ( $\pm$ 11.9)	276.9 ( $\pm$ 66.8)
I274C	3.9 ( $\pm$ 0.2)	180.5 ( $\pm$ 28.2)	176.8 ( $\pm$ 30.7)
I274S	15.5 ( $\pm$ 0.7)	128.6 ( $\pm$ 44.6)	63.7 ( $\pm$ 8.0)
M291C	16.5 ( $\pm$ 3.4)	150.1 ( $\pm$ 28.4)	366.0 ( $\pm$ 102.3)
T304C	8.8 ( $\pm$ 0.9)	79.0 ( $\pm$ 6.9)	209.0 ( $\pm$ 21.2)
T304P	52.5 ( $\pm$ 10.1)	113.3 ( $\pm$ 17.1)	350.0 ( $\pm$ 70.5)
A309C	9.0 ( $\pm$ 0.8)	289.3 ( $\pm$ 43.8)	362.8 ( $\pm$ 81.4)
A309G	14.6 ( $\pm$ 0.8)	36.9 ( $\pm$ 2.3)	337.8 ( $\pm$ 27.0)
A309S	10.7 ( $\pm$ 0.8)	153.3 ( $\pm$ 13.6)	273.5 ( $\pm$ 47.5)
V312C	17.6 ( $\pm$ 1.6)	202.3 ( $\pm$ 19.4)	495.8 ( $\pm$ 34.2)
S417C	9.0 ( $\pm$ 0.6)	202.4 ( $\pm$ 33.8)	280.9 ( $\pm$ 22.0)
S417N	8.1 ( $\pm$ 0.8)	180.0 ( $\pm$ 46.5)	102.5 ( $\pm$ 19.1)

**TABLE 5** Specificity profiles of active SmLL9 mutants

Note: *Escherichia coli* T184 expressing the corresponding constructs were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) uptake at 25°C, in the absence or presence of 0.1–1,000  $\mu\text{M}$  of unlabeled nucleobases. The data were analyzed with *Prism7* to obtain  $\text{IC}_{50}$  and thereof  $K_i$  values. The values shown are the means of three determinations with SD shown in the parentheses.

(Table 5; Table 3) and, to a lesser extent, for the methyl analogs of xanthine at C8 or N7 (Table 5) and, also, very low efficiency for uric acid transport (Table 3; Figure 4b). The reciprocal mutant (N93S) in XanQ allows marginal uric acid uptake and low-affinity recognition of uric acid, 7-methyl and 8-methylxanthine (Karena and Frillingos, 2011). A comparison of the dose-response inhibitory effects of uric acid and 8-methylxanthine on SmLL9 and the relevant SmLL9 and XanQ mutants illustrates that the S93N replacement deteriorates the ability of SmLL9 to recognize uric acid and 8-methylxanthine mimicking in part the properties of XanQ or, more precisely, of the N93S mutant (Figure 6).

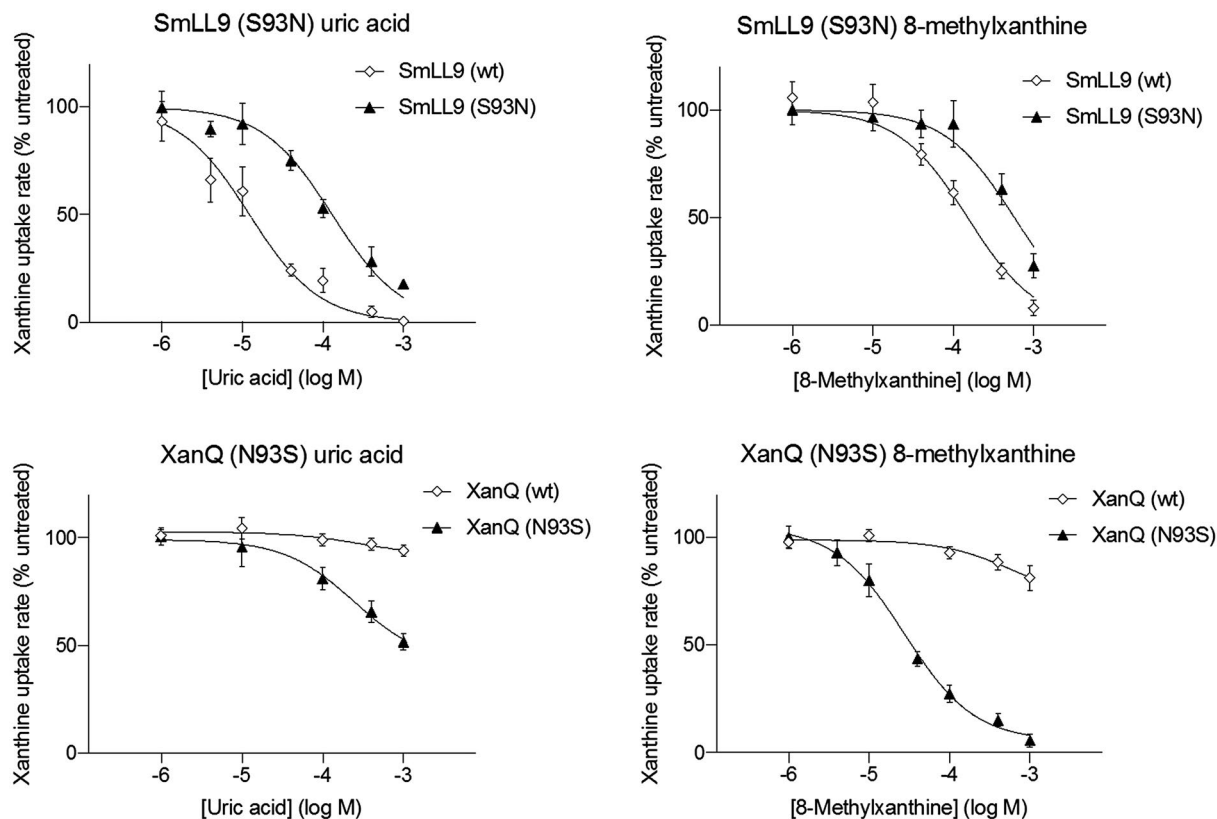
## 2.7 | IFD analysis indicates a role of Ser93 in binding uric acid

Based on the IFD calculations, SmLL9 binds xanthine similarly to UapA (Figure S7). Q310 side chain amide forms two H-bonds with xanthine O2 and N1 while E262 an H-bond with N7. C6=O forms an H-bond with the A309 backbone NH. The ligand is stabilized also by  $\pi$ - $\pi$  stacking interactions with F94 and F308 comparable to UapA

and UraA crystal structures (Figure 7a). In addition, the IFD calculations suggest the existence of an H-bond network in the vicinity of the substrate between S95-S93-S306 ending to the A307 backbone C=O.

Equivalently, XanQ interacts with xanthine in a slightly tilted orientation compared to SmLL9, forming H-bonds between N7 and E272 carboxylate, between N1-H, C2=O and Q310 amide side chain and between C6=O and the backbone NH of A323. F322 forms a  $\pi$ - $\pi$  stacking interaction with xanthine while F94 a T-stacking (Figure 7c). Further to the above interactions, it is interesting to note that during the MD simulations the N93 side chain displays a relative flexibility interacting with different residues in the vicinity, notably S95, S269, T320, and E272.

IFD calculations show that uric acid binds to SmLL9 with the same geometry and H-bonds as xanthine (Figure 7b), yet forming an extra hydrogen bond between C8=O and the S93 hydroxyl group connecting thus the substrate with the above mentioned S95-S93-S306 H-bond network. In contrast, uric acid binds to XanQ in a reverse orientation in the binding cavity with N7 and N3 directed toward E272, while C6=O is oriented toward F94 inversely to the corresponding xanthine carbonyl group which points to the E272 side chain (Figure 7d). Uric



**FIGURE 6** Impairment of the affinity of SmLL9 for uric acid (UA) by replacing Ser93 with Asn. *Escherichia coli* T184 expressing wild-type or the corresponding mutant of SmLL9 or XanQ were analyzed for initial rates of [ $^3$ H]-xanthine (1  $\mu$ M) in the presence of UA or 8-methylxanthine, as indicated, and compared with control cells assayed in the absence of ligands. Each data point represents the mean and SD of triplicate measurements. Data were analyzed with Prism7 to obtain  $IC_{50}$  and thereof  $K_i$  values. The results with the XanQ mutant N93S replicate previously published results (Karena and Frillingos, 2011). The  $K_i$  values deduced from the above experiments (mean  $\pm$  SE) are: SmLL9 (wt)  $9.4 \pm 1.5 \mu$ M (UA),  $116.0 \pm 13.5 \mu$ M (8-methylxanthine); SmLL9 (S93N)  $120.8 \pm 12.4 \mu$ M (UA),  $535.4 \pm 76.4 \mu$ M (8-methylxanthine); XanQ (N93S)  $18.7 \pm 2.7 \mu$ M (8-methylxanthine) (the  $K_M$  used for this mutant for derivatization of the  $K_i$  value is from reference Karena and Frillingos, 2011)

acid appears to form two H-bonds between Q324 amide and uric acid N1H and C2=O. However, the angle between the two planar systems, namely the Q324 amide and the substrate aromatic system, is  $\sim 45$  degrees, less optimal for the formation of strong H-bonds. In addition, the C8=O group does not interact with the N93 side chain which is directed toward the opposite side of the binding cavity.

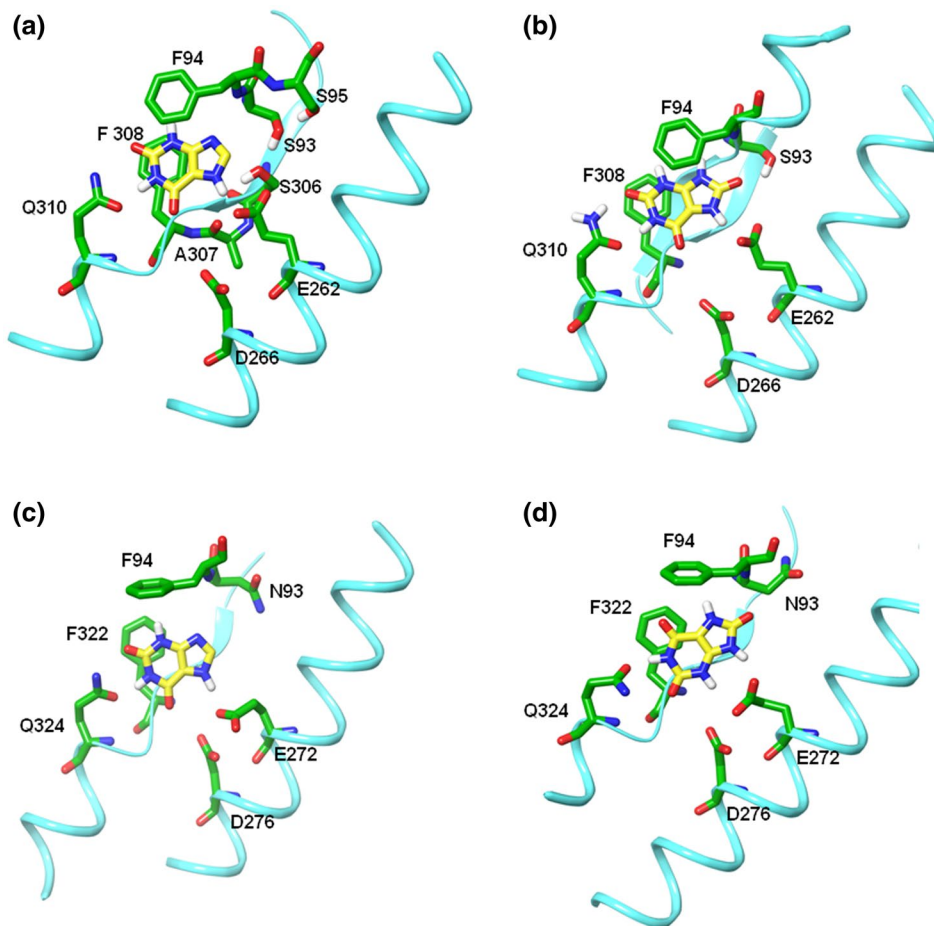
Overall, the docking calculations suggest that XanQ displays weaker interaction with uric acid as compared to xanthine. Uric acid cannot be accommodated in the binding cavity adopting a reverse orientation of the purine moiety compared to xanthine, while interaction with the tilted Q324 seems weaker and there is no interaction between C8=O and the N93 side chain. It is interesting that, although the major difference between XanQ and SmLL9 in the binding cavity is related to residue 93, this difference is enough to produce weaker XanQ-uric acid interaction and generate specificity. Uric acid needs the presence of the H-bond between C8=O and S93 hydroxyl group to bind to SmLL9 (Figure 7b). This interaction stabilizes the pose which is the most energetically favorable and geometrically suitable according to the UapA-xanthine based model, in agreement with the mutagenesis data showing that the S93N replacement leads specifically to impairment of the uric acid binding and uptake activity.

### 3 | DISCUSSION

The data in this study show that the model rhizobium *S. meliloti* contains a set of purine nucleobase transporters which display broader substrate profiles relative to their homologs from other proteobacteria and are uniquely present in the genus *Sinorhizobium* (*Ensifer*) in contrast to other rhizobia. SmLL9 and SmVC3 are also found in few other (non nodulating) genera of Rhizobiales, but the other four homologs are present only in *Sinorhizobium*. The complete set of six homologs is present in practically all strains of *S. meliloti* and *S. medicae* but not in *S. fredii*, *S. americanum* or any other related species which contain only SmVC3 and, less consistently, SmLL9, implying a recent and significant lineage-specific expansion of this group in the *S. meliloti*/*S. medicae* lineage (which diverged from the *S. fredii*/*S. americanum* lineage 35 million years ago; based on TimeTree). In support of this interpretation, the *SmLL9* and *SmLL8* genes are found in tandem in the *S. meliloti*/*medicae* genomes (Galibert *et al.*, 2001; Reeve *et al.*, 2010; Nelson *et al.*, 2018), indicative of an evolutionarily recent duplication event.

It is interesting that *S. meliloti* and *S. medicae* are exclusively associated with the effective nodulation of species of the Leguminosae





**FIGURE 7** Structural basis of the interaction of SmLL9 and XanQ with xanthine and uric acid (UA). Low energy structures of transporter-substrate complexes as resulted from the induced-fit docking analysis. (a) SmLL9-xanthine. (b) SmLL9-UA. (c) XanQ-xanthine. (d) XanQ-UA. The H-bond network between S95-S93-S306 is illustrated in (a) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

genera *Medicago*, *Melilotus*, and/or *Trigonella*, most of which contain *S. meliloti* or *S. medicae* as their primary or even exclusive nitrogen-fixing microsymbiont (Berrada and Fikri-Benbrahim, 2014; Andrews and Andrews, 2017). These three genera belong to Trifolieae, which diverged from the other legume lineages in the inverted repeat-lacking clade (IRLC) of the Papilionoideae (Faboideae) subfamily 36 million years ago (based on TimeTree) (see also Lavin *et al.*, 2005). The *Medicago/Melilotus/Trigonella* lineage is estimated to have evolved 16 million years ago. Therefore, it appears that the expansion of the NAT/NCS2 purine transporter genes in *S. meliloti* and *S. medicae* might have evolved in parallel with their associated host plants of the Trifolieae tribe.

Given that the presence of the above NAT/NCS2 homologs is confined in *S. meliloti* and few other alpha-proteobacteria, we cannot rule out the possibility that they have been acquired by a relatively recent horizontal gene transfer (HGT). However, the evidence for HGT is rather poor. First, these transporters do not share sequence identity higher than 80%–82% with any distant homolog outside alpha-proteobacteria; this identity score is in the same range with or lower than the identity of the same transporters with homologs in Rhizobiales. Second, their GC content is between 61% and 64%

which is almost average for core genes of *S. meliloti* and in the upper fourth for HT genes (Epstein *et al.*, 2014). Third, two of these genes are chromosomal (*SmLL9* and *SmLL8*), three (*SmX28*, *SmWQ0*, *SmVC3*) are in the chromid pSymB which is thought to largely derive from ancient HGT, and only one (*SmYE1*) is in the megaplasmid pSymA which contains the nodulation and nitrogen fixation genes and is considered to have originated through recent HGT (diCenzo *et al.*, 2018; Barnett *et al.*, 2001).

The type of legume-rhizobia symbiosis involving *S. meliloti* has been studied in the model legume *Medicago truncatula*. The *M. truncatula*–*S. meliloti* symbiotic system forms indeterminate-type symbiotic nodules in which the infected plant cells contain rhizobia that undergo a process of endoreduplication of their DNA as they differentiate to nitrogen-fixing bacteroids (Vinardell *et al.*, 2003; Udvardi and Poole, 2013; Penterman *et al.*, 2014). The increased needs for nucleotide building blocks during this process might be partially covered by the NAT/NCS2 transporters identified here. In particular, the gene for the adenine/guanine/hypoxanthine transporter *SmVC3* clusters together with several genes of adenine/guanine/xanthine metabolism and in tandem with an adenine/guanine phosphoribosyl transferase gene (Finan *et al.*,

2001), implying involvement of the purines transported by SmVC3 in purine salvage.

Of the six NAT/NCS2 homologs, SmLL9 and SmVC3 are the most efficient and display the broadest specificity for purine substrates relative to their paralogs (SmLL8 or SmX8, and SmYE1, respectively). The combined functions of SmLL9 (xanthine and uric acid transporter) and SmVC3 (adenine, guanine and hypoxanthine transporter) would be sufficient to allow transport of all natural purine nucleobase substrates. However, in the few cases of rhizobia other than *S. meliloti/medicae* that harbor both SmLL9 and SmVC3 (but not any of the other NAT/NCS2 homologs) in their genomes (Table S1), the relevant rhizobia are implicated in different-type SNF systems; *S. fredii* with Papilionoideae forming determinate nodules (*Glycine*, *Vigna*), *S. americanum* with Mimosoideae (*Acacia*), *R. tropici* CIAT499, and *R. freirei* with Papilionoideae forming determinate nodules (*Phaseolus*), *P. sophorae* with non-IRLC Papilionoideae forming indeterminate nodules (*Sophora*) (Berrada and Fikri-Benbrahim, 2014; and references therein). This observation implies that some function of the other NAT/NCS2 homologs might be particularly relevant to the symbiotic system between *S. meliloti* and the *Medicago* host. SmLL8 which offers an additional capacity for efficient, high-affinity uptake of uric acid (Table 1) might be relevant in this respect. An interesting possibility is that uric acid might need to be efficiently imported and catabolized in the bacteroid not just as a N source but also to provide the microsymbiont with more TCA intermediates through the glyoxylate cycle, deriving malate from the glyoxylate produced in the lower part of uric acid catabolism through the action of S(-)-ureidoglycolate hydrolase (Cusa *et al.*, 1999; Izaguirre-Mayoral *et al.*, 2018).

SmWQ0 is normally expressed in the plasma membrane of *E. coli* as a seemingly full-length protein (Figure 2a) but was not found to transport any nucleobase. SmWQ0 is related phylogenetically with SmLL9, SmLL8, SmX28 and other xanthine and/or uric-acid transporting homologs (Figure 1b), but does not conserve key residues shown to be important for substrate recognition by such homologs (Frillingos, 2012; Diallinas, 2016) and by other well-studied NATs/NCS2s (Yu *et al.*, 2017; Botou *et al.*, 2018). Namely, it lacks an aromatic residue in the binding site at positions of F94 (TM3) or F308 (TM10) (numbering according to SmLL9), it lacks the specificity-related S/T/N93 (TM3), and also lacks G/R/K319 in the NAT-signature motif (TM10) (Kourkoulou *et al.*, 2018) which has also been related with specificity (Figure S2). Another interesting deviation is that SmWQ0 does not retain the substrate-binding Q at TM10 which is a characteristic of the purine NAT/COG2233 transporters (Figure S2) but has an E instead, a characteristic of pyrimidine-transporting homologs (Yamamoto *et al.*, 2010; Lu *et al.*, 2011; Yu *et al.*, 2017; Botou *et al.*, 2018). In our experiments, we did not detect any statistically significant transport activity for uracil or other pyrimidines that could be analyzed kinetically. Overall, this evidence suggests that SmWQ0 may recognize and transport a different solute other than nucleobases.

SmLL9 and SmVC3 display substantially broader substrate profiles than their corresponding homologs from *E. coli* K-12, which uses separate transporters for xanthine or uric acid and adenine or guanine/hypoxanthine (Frillingos, 2013; Papakostas *et al.*, 2013).

SmVC3, in particular, can use analogs that are not used by either the adenine or the guanine/hypoxanthine transporters of *E. coli*. Among these additional purine analogs, SmVC3 recognizes with high affinity the cytokinin 6-benzylaminopurine (6-BAP or BA) as well as the cytokinin analog 6-benzoyladenine. BA and other adenine-type cytokinins are important regulators of the plant-rhizobium interaction leading to nodulation (Gamas *et al.*, 2017). They are produced by plant cells in response to rhizobial nodulation (Nod) factors but can also be produced by rhizobia or other rhizosphere bacteria and used directly as alternative or supplemental nodulation factors by rhizobial strains (Cooper and Long, 1994; Giraud *et al.*, 2007; Kisiala *et al.*, 2013).

SmLL9 was subjected to site-directed mutagenesis to reveal major determinants of its functional profile and its difference from the xanthine-specific XanQ of *E. coli*, taking advantage of the fact that XanQ is well studied experimentally (Frillingos, 2012) and that both transport proteins can be modeled on the structurally known homologs UapA (Alguel *et al.*, 2016) and UraA (Lu *et al.*, 2011; Yu *et al.*, 2017). The mutagenesis data show that SmLL9 and XanQ share several key functionally important residues which have similar roles in both transporters. This holds true for the xanthine-binding residues Glu262/272 and Gln310/324, residues involved in important H-bonds at the periphery of the binding site (His31/31, Asp266/276, Asp290/304, Asn311/325), but also residues of the binding-site region that are highly sensitive to alkylation by *N*-ethylmaleimide with similar IC<sub>50</sub> values (Ala265/Gly275, Ala309/Ala323). In addition, the kinetic properties of SmLL9 with respect to xanthine transport resemble those of XanQ (almost the same K<sub>M</sub> values, comparable K<sub>i</sub> values for inhibition by oxypurinol and other analogs). Based on the IFD studies, both transporters bind the xanthine substrate with equivalent interactions. Thus, SmLL9 has evolved the additional property of utilizing uric acid on a very similar molecular background with the xanthine-specific XanQ.

Ser93 in SmLL9 appears to be a major determinant for the transporter's additional ability to recognize and transport uric acid. The replacement of Ser93 with Cys results in unstable protein with no detectable expression in the membrane, but the replacement with Asn specifically impairs uric acid binding (>12-fold lower affinity relative to wild type) (Table 5; Figure 6) and uric acid transport (marginally detectable activity for uric acid while retaining highly significant xanthine uptake activity and 10-fold lower transport efficiency with uric acid relative to xanthine) (Table 3; Figure 4). These properties are corroborated by the IFD analysis for the binding-site interactions of SmLL9 and XanQ with xanthine and uric acid, highlighting that Ser93 in SmLL9 is needed to H-bond with C8=O of uric acid, an interaction that is absent in XanQ. Furthermore, the presence of Asn93 instead of Ser93 is enough to produce weaker interactions with uric acid in the binding cavity of XanQ (Figure 7b,d) consistent with the strict specificity of XanQ relative to SmLL9.

Our results on Ser93 confirm and extend previous findings on the importance of this residue in other homologous NAT/NCS2 transporters and its relation with specificity. A Ser at this position is conserved in all dual-specificity xanthine/uric acid NAT/NCS2

transporters in TM3 at the binding site region and this Ser has already been implicated with specificity from the mutagenesis studies of XanQ and UapA. However, the previous studies do not provide clear evidence that this Ser is important for binding and transport of uric acid. Although UapA is structurally known, the available crystal structure of UapA is in complex with xanthine (Alguet *et al.*, 2016) and its interaction with uric acid has not been studied with IFD analysis. The relevant mutants S154N and S154A are defective in transport of both xanthine and uric acid, retain significant affinity for uric acid although differing in substrate profiles from wild type, and their relative transport efficiencies for the two substrates have not been tested (Amillis *et al.*, 2011). In XanQ, the relevant mutant N93S shows hardly detectable ability to bind and transport uric acid in addition to xanthine (Karena and Frillingos, 2009; Karena and Frillingos, 2011). In UacT, the relevant mutant T100S shows very low uptake activity for uric acid and hardly detectable ability to transport xanthine in addition to uric acid (Papakostas and Frillingos, 2012). Thus, our study with SmLL9 is the first one to show that a Ser at this NAT/NCS2 position is important specifically for the recognition and transport of uric acid.

Other amino acid residues, external to the binding site, have been suggested to influence specificity in the fungal UapA by participating in gating elements that act as “selectivity gates” controlling the access of substrates to the binding site (Diallinas, 2014; Diallinas, 2016). Mutations at such residues were found to allow recognition of atypical substrates like hypoxanthine and adenine without affecting the recognition and transport of the normal substrates, xanthine or uric acid (Papageorgiou *et al.*, 2008; Kosti *et al.*, 2012; Alguet *et al.*, 2016). Evidence for such gating elements in bacterial NAT/NCS2 members is poor. The systematic Cys-scanning and site-directed mutagenesis of XanQ revealed very few mutations outside the binding site that allowed recognition of nonwild-type ligands in addition to xanthine and such nonwild-type ligands were never found to be recognized with high affinity or transported at a significant rate (Karena *et al.*, 2015). In the present study with SmLL9, one of the residues analyzed (S417) corresponds to T526 in UapA (TM14), which has been implicated as belonging to an outward-facing selectivity gate leading toward the binding site, based on the properties of the mutants T526M and T526L, either alone or combined with replacements in other sites (Q113L, R481G, Q408E), and MD simulations (Papageorgiou *et al.*, 2008; Alguet *et al.*, 2016). In our study, S417N and S417C were not found to differ from wild-type SmLL9 in the affinity for either xanthine or uric acid (Table 3) or specificity profile (Table 5) or display any detectable recognition of other nucleobases (data not shown). Similarly, mutations N430S and N430C at the corresponding position of XanQ had not revealed any deviation from the wild-type profile (Papakostas *et al.*, 2008; Papakostas and Frillingos, 2012). Taken together, the current data on XanQ and SmLL9 do not seem to support the idea that gating elements related with specificity exist in these bacterial NATs/NCS2s.

Based on the evolutionary and specificity profile of the homologs identified here, we propose that, of the *S. meliloti* (Sm) NAT/COG2233 transporters, which are closely related to UacT (Figure 1b; see also Karena *et al.*, 2015), SmLL8 and SmX28 be renamed to SmUacT1 and SmUacT2 (uric acid transporter 1 and 2), respectively,

and SmLL9 to SmXUacT (xanthine/uric acid transporter); of the AzgA-like/COG2252 transporters, SmVC3 be renamed to SmAGHxT (adenine/guanine/hypoxanthine transporter) and SmYE1 to SmAdeT (adenine transporter).

Apart from the NAT/NCS2 genes analyzed in this study, all *S. meliloti/medicae* genomes contain also one NCS1 homolog each (Table S1). Although NCS1 is annotated as a nucleobase-cation symporter-1 family, few bacterial members of this family have been studied in detail and substrates of known microbial members are more diverse than in NAT/NCS2 including also allantoin, hydantoins, pyridoxine, thiamine, or nicotinamide riboside (Patching, 2018). The *S. meliloti* homolog (as well as its closely related homologs from *Mesorhizobium*) do not cluster together with any functionally known member and are distantly related to YbbW of *E. coli* K-12 which is encoded in an allantoin-catabolizing operon (Cusa *et al.*, 1999) but remains unstudied to date.

Another set of genes encoding a candidate purine transport system might be the ATP-binding cassette (ABC) genes of an operon that includes also a gene for adenine deaminase (SMb21278) and a gene for nucleoside hydrolase (SMb21277). This ABC system, however, has been annotated as a putative spermidine/putrescine (polyamine) transport system which is widely conserved in rhizobacteria and other proteobacteria, its homologs including the GABA-transport system of *Rhizobium leguminosarum* bv. *viciae* (White *et al.*, 2009) and the spermidine/putrescine (PotABCD) transport system of *E. coli* (Igarashi *et al.*, 2001).

In conclusion, the NAT/NCS2 rhizobial purine transporters identified in this study may play unique roles in the *M. truncatula*–*S. meliloti* symbiotic system that are yet to be discovered. A comprehensive analysis of transcript profiles of the transporters and the relevant purine metabolism enzymes in the developing symbiotic nodules would be very helpful to this end. Since the organelle-like symbiosomes in the infected plant cell are enclosed in two membranes, one of rhizobial and one of plantal origin, an investigation of the functional and physiological relevance of *M. truncatula* purine transporters which are expressed in the symbiotic nodules would also be important. *M. truncatula* contains a multitude of potential nucleobase transporters, distributed in families NAT/NCS2, NCS1, the nucleoside transporter family ENT, and the plant-specific families PUP/POP and UPS/NBUT (Girke *et al.*, 2014), which remain experimentally untouched to date although they include some homologs that are highly expressed specifically in the root nodules (<http://mtgea.noble.org>; <http://www.jcvi.org/medicago>). Thus, our study opens up a new research direction for the evaluation of the role of purine transport and metabolism in symbiotic nitrogen fixation systems.

## 4 | EXPERIMENTAL PROCEDURES

### 4.1 | Designation of the NAT/NCS2 genes

The *S. meliloti* 1,021 genome (<http://www.cebitec.uni-bielefeld.de/CeBiTec/rhizogate/>) contains six NAT/NCS2 genes, four in subfamily

NAT/COG2233 (subfamily 1, in Cluster C1\_Xanthine-Uric Acid), and two in subfamily AzgA-like/COG2252 (subfamily 2/AzgA-like) (<http://bioinf.bio.uth.gr/nat-ncs2/>). The genes (with UniProt accession numbers for the predicted proteins in parentheses) are: *SMc02513* (Q92LL9), *SMc02512* (Q92LL8), *SM\_b20134* (Q92X28), and *SM\_b20289* (Q92WQ0) in NAT/COG2233; *SM\_b21281* (Q92VC3) and *SMa1717* (Q92YE1) in AzgA-like/COG2252. In this study, we use the designations SmLL9, SmLL8, SmX28, SmWQ0, SmVC3, and SmYE1 to highlight the fact that the corresponding orthologs are very well conserved (>96% sequence identity) in all the *S. meliloti/medicae* (Sm) strains.

## 4.2 | Materials and general considerations

[2,8-<sup>3</sup>H]-adenine (31.8 Ci mmol<sup>-1</sup>), [8-<sup>3</sup>H]-guanine (21.2 Ci mmol<sup>-1</sup>), [2,8-<sup>3</sup>H]-hypoxanthine (27.7 Ci mmol<sup>-1</sup>), [8-<sup>3</sup>H]-xanthine (22.8 Ci mmol<sup>-1</sup>), [8-<sup>14</sup>C]-uric acid (58 mCi mmol<sup>-1</sup>), [5,6-<sup>3</sup>H]-uracil (32.5 Ci mmol<sup>-1</sup>), [methyl-<sup>3</sup>H]-thymine (50.6 Ci mmol<sup>-1</sup>), and [<sup>3</sup>H]-cytosine (18.8 Ci mmol<sup>-1</sup>) were from Moravek Biochemicals. Non-radioactive nucleobases and analogs were from Sigma-Aldrich except 6-(dimethylamino) ethyladenine and 6-benzylaminopurine which were synthesized and kindly provided by Nicole Pouli lab at the Department of Pharmaceutical Chemistry, National and Kapodistrian University of Athens. The above two analogs as well as 6-benzoyladenine (Figure 3b) were prepared in dimethyl sulfoxide (DMSO). Carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) was also prepared in DMSO. Cell cultures were performed in Luria-Bertani broth (LB) in aerobic conditions. For all incubations in liquid media, cells were grown with shaking at 220 r.p.m. at 37°C (*E. coli* cultures) or at 180 r.p.m. at 28°C (*S. meliloti* 1,021). Oligodeoxynucleotides were synthesized from Eurofins Genomics GmbH. High-fidelity DNA polymerase was from Kapa Biosystems. Restriction endonucleases, alkaline phosphatase, and T4 DNA ligase were from Takara Clontech. Horseradish peroxidase (HRP)-conjugated streptavidin was from Millipore. All other reagents were of analytical grade and obtained from commercial sources.

## 4.3 | Proteobacterial strains, coding sequences, and plasmids

The six NAT/NCS2 genes were mobilized with PCR amplification from the genome of *S. meliloti* 1,021. The genomic DNA had been isolated from *S. meliloti* 1,021 cultures grown aerobically on Yeast Mannitol Broth (YMB) at 28°C. For heterologous expression in *E. coli* K-12, the coding sequence of each gene was transferred to a previously described version of plasmid vector pT7-5 which included the DNA sequence of the biotin-acceptor domain (BAD) of the oxaloacetate decarboxylase from *Klebsiella pneumoniae* as an insert between the *Apal* and *HindIII* sites. This vector is designated pT7-5/-BAD. After insertion of the NAT/NCS2 coding sequences in pT7-5/-BAD, at the appropriate orientation and frame, the resulting constructs contain the BAD sequence as a C-terminal tag of each NAT/NCS2.

Following expression of these genes, the gene products are biotinylated in vivo during bacterial growth and allow monitoring of the protein levels in the *E. coli* membrane by western blotting (Karatza and Frillingos, 2005).

*Escherichia coli* TOP10F' (Invitrogen) was used for initial propagation of recombinant plasmids to prepare plasmid DNA for sequencing. *E. coli* T184 (Teather *et al.*, 1980) was used for expression of the pT7-5/-BAD-borne genes from the *lacZ* promoter/operator by induction with isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) and for analysis of the uptake of xanthine (Karatza and Frillingos, 2005) and of uric acid (Papakostas and Frillingos, 2012). Similarly, *E. coli* JW3692 (Keio collection; *adeP* knock-out) was used for the analysis of the uptake of adenine and hypoxanthine and JW4025 (Keio collection; *ghxP* knock-out) was used for the analysis of the uptake of guanine (Papakostas *et al.*, 2013), whereas JW2482 (Keio collection; *uraA* knock-out) was used for the uptake of uracil and thymine (Botou *et al.*, 2018) and JW0327 (Keio collection; *codB* knock-out) for the uptake of cytosine (Danielsen *et al.*, 1992). The Keio collection single-gene knockout strains (Baba *et al.*, 2006) were provided from the *E. coli* Genetic Stock Center. All *E. coli* strains used were transformed according to Inoue *et al.* (1990).

## 4.4 | Molecular cloning

The coding sequences of NAT/NCS2 genes were amplified by PCR on the template of genomic DNA and transferred to pT7-5/-BAD by restriction fragment replacement between the *Bam*HI and *Apal* sites. All site-directed mutants of SmLL9 were constructed using two-stage (overlap extension) PCR (Heckman and Pease, 2007). The sequences of synthetic oligodeoxynucleotides used for these constructions are given in Table S5. The coding sequence of all constructs was verified by double-strand DNA sequencing (Eurofins Genomics GmbH).

## 4.5 | Growth of bacteria

*Escherichia coli* Keio collection strains harboring given plasmids were grown aerobically at 37°C in LB containing kanamycin (0.025 mg/ml) and ampicillin (0.1 mg/ml). *E. coli* T184 was grown at the same conditions except that streptomycin (0.01 mg/ml) was used instead of kanamycin. Fully grown cultures (1 ml) were diluted 10-fold, allowed to grow to mid-logarithmic phase to a cell density of OD<sub>600nm</sub> = 2.0, induced with IPTG (0.5 mM) for an additional 105 min at 37°C, and harvested for use in transport assays or western blotting.

## 4.6 | Transport assays and kinetic analysis

*Escherichia coli* T184, JW3692, or JW4025 were washed twice in KPi, 0.1 M, pH 7.5, normalized to an OD<sub>420nm</sub> of 10.0 (corresponding to 35  $\mu$ g of total protein per 50  $\mu$ l) in the same buffer and assayed for transport of radiolabeled substrates. Alternatively, *E. coli* JW2482



were prepared in MK buffer (MES 5 mM, pH 6.5, containing KCl, 0.15 M), as described (Botou *et al.*, 2018). Before initiating the transport reaction, the cells were energized by addition of glycerol to a final concentration of 20 mM and equilibrated in the assay buffer for 3 min at 25°C. All the transport reactions were performed at 25°C. After termination of the reactions, samples were rapidly filtered through Whatman GF/C filters, washed twice immediately with 3 ml of ice-cold KL buffer (KPi, 0.1 M, pH 5.5, LiCl, 0.1 M) and taken for liquid scintillation counting. To determine  $K_M$  and  $V_{max}$  values for the uptake of purine nucleobases, *E. coli* T184 were assayed for active transport of [<sup>3</sup>H]-xanthine at 5–20 s and of [<sup>14</sup>C]-uric acid at 5–30 s, *E. coli* JW3692 for active transport of [<sup>3</sup>H]-adenine at 5–20 s and of [<sup>3</sup>H]-hypoxanthine at 5–15 s, *E. coli* JW4025 for active transport of [<sup>3</sup>H]-guanine at 5–20 s, at a range of substrate concentrations, and data were fitted to the Michaelis–Menten equation using Prism7.

For competitive inhibition experiments, the uptake rate of [<sup>3</sup>H]-xanthine (1 μM) or [<sup>3</sup>H]-adenine (0.1 μM) was assayed at 5–20 s in the absence or presence of non-radiolabeled analogs. To obtain  $IC_{50}$  values, the data were fitted to the equation  $y = B + (T - B) / (1 + 10^{(\log IC_{50} - \log x)h})$  for sigmoidal dose-response (variable slope), using Prism7, where  $x$  is the concentration variable,  $y$  (the transport rate) ranges from  $T$  (top) to  $B$  (bottom) and  $h$  is the Hill coefficient. In all cases,  $h$  was close to  $-1$ , consistent with competition for a single binding site.  $K_i$  values were calculated from the  $IC_{50}$  values, based on the equation:  $K_i = IC_{50} / [1 + (S/K_M)]$  (where  $S$  is the concentration of [<sup>3</sup>H]-xanthine or [<sup>3</sup>H]-adenine used; and  $K_M$  is the  $K_M$  value obtained for the uptake of xanthine or adenine by the relevant permease or mutant) (Cheng and Prusoff, 1973). It should be noted that the  $K_i$  value is an affinity constant implying binding to the transporter but does not indicate whether the ligand is being transported across the membrane.

The effect of NEM on the [<sup>3</sup>H]-xanthine transport activity of single-Cys mutants was assayed as described previously (Karena *et al.*, 2015). When assaying the effect of NEM, transport assays were performed in the presence of 20 mM potassium ascorbate and 0.2 mM phenazine methosulfate (PMS), an artificial electron-donor system used to energize the *E. coli* membrane, as described previously (Barnes and Kaback, 1971; Konings *et al.*, 1971). *E. coli* T184 had been preincubated with NEM (ranging from 1 μM to 2 mM), for 10 min, at 25°C, and the reactions stopped by addition of a 10-fold molar excess of dithiothreitol (DTT). For incubations in the presence of substrate, xanthine (1 mM) was added in *E. coli* T184 prior to initiation of the NEM-reaction. After termination of the reaction, the excess of xanthine was removed by washing the samples twice with 40 ml of KPi, 0.1 M, pH 7.5, and thrice with 1 ml of the same buffer. Samples were finally normalized to an  $OD_{420nm}$  of 10 for the xanthine transport assay, as above.

#### 4.7 | Western blot analysis

*Escherichia coli* cells were washed twice in Tris-HCl (0.05 M), pH 8.0, containing NaCl (0.1 M) and Na<sub>2</sub>EDTA (1 mM), supplemented with 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) (0.2 mM), and used to prepare membrane fractions by osmotic shock,

treatment with EDTA/lysozyme, and sonication, as described (Karatzas and Frillingos, 2005). Membrane fractions prepared from 10 ml cell cultures were harvested by ultracentrifugation in an Optima MAX-XP Ultracentrifuge (Beckman Coulter), normalized to a protein concentration of 100 μg per 50 μl in sample loading buffer, and subjected to SDS-PAGE (12%). After electrophoresis, proteins were electroblotted to a polyvinylidene difluoride membrane (Parablot PVDF; Macherey-Nagel GmbH) and the BAD-tagged proteins were probed with HRP-conjugated streptavidin which was used at a dilution of 1:50,000. Signals were developed with enhanced chemiluminescence (ECL).

#### 4.8 | Bioinformatic and phylogenetic analysis

Comparative analysis of NAT/NSC2 sequences was based on BLAST-p search and sequence alignments using MUSCLE for multiple alignments and EMBOSS-Needle for pairwise alignments. The most recent genome annotations were used for retrieving sequence data. Phylogenetic trees were constructed with MEGA7 (Kumar *et al.*, 2016). For analysis of the phylogeny of alpha-proteobacteria (Figure 1a), we used multilocus sequence analysis (MLSA) of four housekeeping genes (*rpoB*, *atpD*, *recA*, 16S rDNA) (as in Mousavi *et al.*, 2015). Other phylogenetic/taxonomic analyses (Figure S1) were based on PhyloT and the TimeTree database (Kumar *et al.*, 2017). The analysis of the genetic locus of each transporter gene was based on information retrieved from the Joint Genome Institute Integrated Microbial Genomes & Microbiomes (JGI IMG/M) and the Kyoto Encyclopedia of Genes and Genomes (KEGG).

#### 4.9 | Homology modeling

Homology model building was performed using Prime 2018-4 (Schrödinger, LLC, New York, NY, 2018). (Jacobson *et al.*, 2002; Jacobson *et al.*, 2004).

#### 4.10 | Protein preparation for IFD

The protein was prepared for the IFD calculations using the Protein Preparation Protocol (Schrödinger Suite 2018, Protein Preparation Wizard) implemented in Schrödinger suite and accessible within the Maestro program (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018). Specifically, the hydrogen atoms were added. The orientation of hydroxyl groups of Asn, Gln, and the protonation state of His were optimized to maximize hydrogen bonding. Finally, the ligand – protein complex was refined with a restrained minimization performed by Impref utility, which is based on the Impact molecular mechanics engine (Impact, Schrödinger, LLC, New York, NY, 2016) and the OPLS2005 force field, setting a max rmsd of 0.30. Ligand preparation for docking was performed with LigPrep (LigPrep, version 2018-4, Schrödinger, LLC, New York, NY, 2018) application which consists of a series of steps that perform conversions, apply

corrections to the structure, generate ionization states and tautomers, and optimize the geometries.

#### 4.11 | Induced-fit docking

The IFD protocol (Schrödinger Suite 2018-4 IFD protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2018), is intended to circumvent the inflexible binding site and accounts for the side chain or backbone movements, or both, upon ligand binding. In the first stage of the IFD protocol, softened-potential docking step, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018-4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues having at least one atom within 8 Å of an atom in any of the 20 ligand poses. The Prime refinement starts with a conformational search and minimization of the side chains of the selected residues and after convergence to a low-energy solution, an additional minimization of all selected residues (side chain and backbone) is performed with the truncated-Newton algorithm using the OPLS parameter set and a surface Generalized Born implicit solvent model. The obtained complexes are ranked according to Prime calculated energy (molecular mechanics and solvation), and those within 50 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. In the final round, the ligands used in the first docking step are redocked into each of the receptor structures retained from the refinement step. The final ranking of the complexes is done by a composite score which accounts for the receptor – ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

#### 4.12 | Molecular Dynamics simulations

The models were minimized and equilibrated further to obtain the stable structures. Each model was inserted into a heterogeneous fully hydrated bilayer 120 Å × 120 Å × 120 Å, consisting of DPPC lipids using CHARMM-GUI and CHARMM36m (Huang *et al.*, 2017) force fields for lipids as well as protein. The protein orientation into the membrane was calculated using the PPM server. The structures were then solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. The assembled simulation system consisted of ~119,000 atoms. The biomolecular systems were simulated using GROMACS (Abraham *et al.*, 2018) software. Minimization was carried out for 2,000 steps with a step size of 0.001 kJ/mol applying a steepest descent and then, a conjugate gradient algorithm, and system was equilibrated for 15 ns while gradually heating and releasing the restraints to expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 323K for 40 ns using Nose-Hoover thermostat and Parrinello-Rahman semi-isotropic pressure coupling and compressibility at 4.5e-5 bar-1. The van der Waals interactions were

smoothly switched off at 1.2 nm by switching function as electrostatics, while long-range electrostatic interactions were calculated using particle mesh Ewald method.

#### ACKNOWLEDGMENTS

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek National Funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALES, Investing in knowledge society through the European Social Fund. I. Z. was supported by a “Stavros S. Niarchos Foundation” grant. We thank Katerina Kalliampakou from Fletmetakis lab for providing *S. meliloti* cultures at initial phases of the project.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTIONS

SF and EF conceived the study and SF with MB designed the experiments. PL and SF performed the phylogenetic analyses. MB performed the initial functional assays of all transporters and the detailed biochemical study and mutagenesis of SmLL9 and analyzed the results with SF. VC assisted in the construction and analysis of the mutants. KP performed the substrate profile analysis of SmLL9 and analyzed the results with MB and SF. VY performed the substrate profile study of SmVC3 and analyzed the results with SF. IZ performed the molecular simulations and analyzed the corresponding results with EM. SF wrote the manuscript with MB, EM, and EF.

#### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article as Supporting Information (Figures S1–S7 and Tables S1–S5).

#### ORCID

Maria Botou  <https://orcid.org/0000-0002-0675-6517>

Vassilis Yalelis  <https://orcid.org/0000-0001-9175-880X>

Stathis Frillingos  <https://orcid.org/0000-0002-2968-6136>

#### REFERENCES

- Abraham, M.J., van der Spoel, D., Lindahl, E. and Hess, B. (2018) GROMACS User Manual version 2018. Available at: [www.gromacs.org](http://www.gromacs.org)
- Alguel, Y., Amillis, S., Leung, J., Lambrinidis, G., Capaldi, S., Scull, N.J. *et al.* (2016) Structure of eukaryotic purine/H<sup>+</sup> symporter UapA suggests a role for homodimerization in transport activity. *Nature Communications*, 7, 11336.
- Amillis, S., Kosti, V., Pantazopoulou, A., Mikros, E. and Diallinas, G. (2011) Mutational analysis and modeling reveal functionally critical residues in transmembrane segments 1 and 3 of the UapA transporter. *Journal of Molecular Biology*, 411, 567–580.
- Andrews, M. and Andrews, M.E. (2017) Specificity in legume-rhizobia symbioses. *International Journal of Molecular Sciences*, 18, 705.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M. *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, 2006.0008.

- Barnes, E.M. Jr and Kaback, H.R. (1971) Mechanisms of active transport in isolated membrane vesicles: I. The site of energy coupling between D-lactic dehydrogenase and  $\beta$ -galactoside transport in *Escherichia coli* membrane vesicles. *Journal of Biological Chemistry*, **246**, 5518–5522.
- Berrada, H. and Fikri-Benbrahim, K. (2014) Taxonomy of the Rhizobia: current perspectives. *British Microbiology Research Journal*, **4**, 616–639.
- Botou, M., Yaelis, V., Lazou, P., Zantza, I., Papakostas, K., Charalambous, V. et al. (2020) Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*. Supporting Information (Figures S1–S7 and Tables S1–S5).
- Botou, M., Lazou, P., Papakostas, K., Lambrinidis, G., Evangelidis, T., Mikros, E. et al. (2018) Insight on specificity of uracil permeases of the NAT/NCS2 family from analysis of the transporter encoded in the pyrimidine utilization operon of *Escherichia coli*. *Molecular Microbiology*, **108**, 204–219.
- Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P., Barloy-Hubler, F. et al. (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9883–9888.
- Bueno Batista, M. and Dixon, R. (2019) Manipulating nitrogen regulation in diazotrophic bacteria for agronomic benefit. *Biochemical Society Transactions*, **47**, 603–614.
- Bürzle, M., Suzuki, Y., Ackermann, D., Miyazaki, H., Maeda, N., Cléménçon, B. et al. (2013) The sodium-dependent ascorbic acid transporter family SLC23. *Molecular Aspects of Medicine*, **34**, 436–454.
- Chaliois, A., Vlastaridis, P., Ntountoumi, C., Botou, M., Yaelis, V., Lazou, P. et al. (2018) NAT/NCS2-hound: a webserver for the detection and evolutionary classification of prokaryotic and eukaryotic nucleobase-cation symporters of the NAT/NCS2 family. *Gigascience*, **7**, giy133.
- Cheng, Y. and Prusoff, W.H. (1973) Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochemical Pharmacology*, **22**, 3099–3108.
- Clarke, V.C., Loughlin, P.C., Day, D.A. and Smith, P.M.C. (2014). Transport processes of the legume symbiosome membrane. *Frontiers in Plant Science*, **5**, 699. <https://doi.org/10.3389/fpls.2014.00699>
- Cooper, J.B. and Long, S.R. (1994) Morphogenic rescue of *Rhizobium meliloti* nodulation mutants by trans-zeatin secretion. *The Plant Cell*, **6**, 215–225.
- Cusa, E., Obradors, N., Baldomà, L., Badia, J. and Aguilar, J. (1999) Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli*. *Journal of Bacteriology*, **181**, 7479–7484.
- Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. and Neuhard, J. (1992) Characterization of the *Escherichia coli* codBA operon encoding cytosine permease and cytosine deaminase. *Molecular Microbiology*, **6**, 1335–1344.
- Diallinas, G. (2014) Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. *Frontiers in Pharmacology*, **12**, 207.
- Diallinas, G. (2016) Dissection of transporter function: from genetics to structure. *Trends in Genetics*, **32**, 576–590.
- diCenzo, G.C., Wellappili, D., Golding, G.B. and Finan, T.M. (2018) Inter-replicon gene flow contributes to transcriptional integration in the *Sinorhizobium meliloti* multipartite genome. *G3*, **8**, 1711–1720.
- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*, **32**, W665–W667.
- Epstein, B., Sadowsky, M.J. and Tiffin, P. (2014) Selection on horizontally transferred and duplicated genes in *Sinorhizobium (Ensifer)*, the root-nodule symbionts of *Medicago*. *Genome Biology and Evolution*, **6**, 1199–1209.
- Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F.J. et al. (2001) The complete sequences of the 1,683-kb pSymB megaplasmid from the  $N_2$ -fixing endosymbiont *Sinorhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9889–9894.
- Frillingos, S. (2012) Insights to the evolution of Nucleobase-Ascorbate transporters (NAT/NCS2) from the Cys-scanning analysis of xanthine permease XanQ. *International Journal of Biochemistry and Molecular Biology*, **3**, 250–272.
- Frillingos, S. (2013) Using Cys-scanning analysis data in the study of membrane transport proteins. In: D. Figurski (Ed.) *Genetic Manipulation of DNA and Protein – Examples from Current Research*. Rijeka: Intech, pp. 333–360.
- Frillingos, S., Sahin-Tóth, M., Wu, J. and Kaback, H.R. (1998) Cys-scanning mutagenesis: a novel approach to structure-function relationships in membrane proteins. *The FASEB Journal*, **12**, 1281–1299.
- Galibert, F., Finan, T.M., Long, S.R., Pühler, A., Abola, P., Ampe, F. et al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, **293**, 668–672.
- Gamas, P., Brault, M., Jardinaud, M.-F. and Frugier, F. (2017) Cytokinins in symbiotic nodulation: when, where, what for? *Trends in Plant Science*, **22**, 792–802.
- Georgopoulou, E., Mermelekas, G., Karena, E. and Frillingos, S. (2010) Purine substrate recognition by the nucleobase-ascorbate transporter motif in the YgfO xanthine permease: Asn-325 binds and Ala-323 senses substrate. *Journal of Biological Chemistry*, **285**, 19422–19433.
- Giraud, E., Moulin, L., Vallanet, D., Barbe, V., Cytryn, E., Avarre, J.C. et al. (2007) Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science*, **316**, 1307–1312.
- Girke, C., Daumann, M., Niopek-Witz, S. and Möhlmann, T. (2014) Nucleobase and nucleoside transport and integration into plant metabolism. *Frontiers in Plant Science*, **5**, 443.
- Gournas, C., Papageorgiou, I. and Diallinas, G. (2008) The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Molecular BioSystems*, **4**, 404–416.
- Heckman, K.L. and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols*, **2**, 924–932.
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L. et al. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, **14**, 71–73.
- Igarashi, K., Ito, K. and Kashiwagi, K. (2001) Polyamine uptake systems in *Escherichia coli*. *Research in Microbiology*, **152**, 271–278.
- Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23–28.
- Izaguirre-Mayoral, M.L., Lazarovits, G. and Baral, B. (2018) Ureide metabolism in plant-associated bacteria: purine plant-bacteria interactive scenarios under nitrogen deficiency. *Plant and Soil*, **428**, 1–34.
- Jacobson, M.P., Friesner, R.A., Xiang, Z. and Honig, B. (2002) On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology*, **320**, 597–608.
- Jacobson, M.P., Pincus, D.L., Rapp, C.S., Day, T.J., Honig, B., Shaw, D.E. et al. (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins*, **55**, 351–367.
- Kaback, H.R., Dunten, R., Frillingos, S., Venkatesan, P., Kwaw, I., Zhang, W. et al. (2007) Site-directed alkylation and the alternating access model for LacY. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 491–494.
- Karatza, P. and Frillingos, S. (2005) Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Molecular Membrane Biology*, **22**, 251–261.
- Karatza, P., Panos, P., Georgopoulou, E. and Frillingos, S. (2006) Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential, and Ile-329-Val-339 form an alpha-helix. *Journal of Biological Chemistry*, **281**, 39881–39890.
- Karena, E. and Frillingos, S. (2009) Role of intramembrane polar residues in the YgfO xanthine permease: His-31 and Asn-93 are crucial for affinity and specificity, and Asp-304 and Glu-272 are irreplaceable. *Journal of Biological Chemistry*, **284**, 24257–24268.



- Karena, E. and Frillingos, S. (2011) The role of transmembrane segment TM3 in the xanthine permease XanQ of *Escherichia coli*. *Journal of Biological Chemistry*, 286, 39595–39605.
- Karena, E., Tatsaki, E., Lambrinidis, G., Mikros, E. and Frillingos, S. (2015) Analysis of conserved NCS2 motifs in the *Escherichia coli* xanthine permease XanQ. *Molecular Microbiology*, 98, 502–517.
- Kisiala, A., Laffont, C., Emery, R.J.N. and Frugier, F. (2013) Bioactive cytokinins are selectively secreted by *Sinorhizobium meliloti* nodulating and nonnodulating strains. *Molecular Plant-Microbe Interactions*, 26, 1225–1231.
- Konings, W.N., Barnes, E.M. Jr and Kaback, H.R. (1971) Mechanisms of active transport in isolated membrane vesicles: III. The coupling of reduced phenazine methosulfate to the concentrative uptake of  $\beta$ -galactosides and amino acids. *Journal of Biological Chemistry*, 246, 5857–5861.
- Kosti, V., Lambrinidis, G., Myrianthopoulos, V., Diallinas, G. and Mikros, E. (2012) Identification of the substrate recognition and transport pathway in a eukaryotic member of the nucleobase-ascorbate transporter (NAT) family. *PLoS ONE*, 7, e41939.
- Koukaki, M., Vlant, A., Goudela, S., Pantazopoulou, A., Gioule, H., Tournaviti, S. et al. (2005) The nucleobase-ascorbate transporter (NAT) signature motif in UapA defines the function of the purine translocation pathway. *Journal of Molecular Biology*, 350, 499–513.
- Kourkoulou, A., Pittis, A.A. and Diallinas, G. (2018) Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family. *Microb Cell*, 5, 280–292.
- Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874.
- Kumar, S., Stecher, G., Suleski, M. and Hedges, S. (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, 34, 1812–1819.
- Lavin, M., Herendeen, P.S. and Wojciechowski, M.F. (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology*, 54, 575–594.
- Liu, A., Contador, C.A., Fan, K. and Lam, H.-M. (2018) Interaction and regulation of carbon, nitrogen, and phosphorus metabolisms in root nodules of legumes. *Frontiers in Plant Science*, 9, 1860. <https://doi.org/10.3389/fpls.2018.01860>
- Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G. et al. (2011) Structure and mechanism of the uracil transporter UraA. *Nature*, 472, 243–246.
- Mermelekas, G., Georgopoulou, E., Kallis, A., Botou, M., Vlantos, V. and Frillingos, S. (2010) Cysteine-scanning analysis of helices TM8, TM9a, and TM9b and intervening loops in the YgfO xanthine permease: a carboxyl group is essential at Asp-276. *Journal of Biological Chemistry*, 285, 35011–35020.
- Mousavi, S.A., Willems, A., Nesme, X., de Lajudie, P. and Lindström, C. (2015) Revised phylogeny of *Rhizobiaceae*: proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations. *Systematic and Applied Microbiology*, 38, 84–90.
- Mus, F., Crook, M.B., Garcia, K., Garcia Costas, A., Geddes, B.A., Kouri, E.D. et al. (2016) Symbiotic nitrogen fixation and the challenges of its extension to nonlegumes. *Applied and Environment Microbiology*, 82, 3698–3710.
- Nelson, M., Guhlin, J., Epstein, B., Tiffin, P. and Sadowsky, M.J. (2018) The complete replicons of 16 *Ensifer meliloti* strains offer insights into intra- and inter-replicon gene transfer, transposon-associated loci, and repeat elements. *Microbial Genomics*, 4, e000174.
- Pantazopoulou, A. and Diallinas, G. (2006) The first transmembrane segment of UapA contains determinants necessary for expression in the plasma membrane and purine transport. *Molecular Membrane Biology*, 23, 337–348.
- Papageorgiou, I., Gournas, C., Vlant, A., Amillis, S., Pantazopoulou, A. and Diallinas, G. (2008) Specific interdomain synergy in the UapA transporter determines its unique specificity for uric acid among NAT carriers. *Journal of Molecular Biology*, 382, 1121–1135.
- Papakostas, K. and Frillingos, S. (2012) Substrate selectivity of YgfU, a uric acid transporter from *Escherichia coli*. *Journal of Biological Chemistry*, 287, 15684–15695.
- Papakostas, K., Botou, M. and Frillingos, S. (2013) Functional identification of the hypoxanthine/guanine transporters YjcD and YgfQ and the adenine transporters PurP and YicO of *Escherichia coli* K-12. *Journal of Biological Chemistry*, 288, 36827–36840.
- Papakostas, K., Georgopoulou, E. and Frillingos, S. (2008) Cysteine-scanning analysis of putative helix XII in the YgfO xanthine permease: Ile-432 and Asn-430 are important. *Journal of Biological Chemistry*, 283, 13666–13678.
- Patching, S.G. (2018) Recent developments in nucleobase cation symporter-1 (NCS1) family transport proteins from bacteria, archaea, fungi and plants. *Journal of Biosciences*, 43, 797–815.
- Penterman, J., Abo, R.P., De Nisco, N.J., Arnold, M.F., Longhi, R., Zanda, M. et al. (2014) Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 3561–3566.
- Reeve, W., Chain, P., O'Hara, G., Ardley, J., Nandesena, K., Bräu, L. et al. (2010) Complete genome sequence of the *Medicago* microsymbiont *Ensifer (Sinorhizobium) medicae* strain WSM419. *Standards in Genomic Sciences*, 2, 77–86.
- Rogers, C. and Oldroyd, G.E. (2014) Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *Journal of Experimental Botany*, 65, 1939–1946.
- Teather, R.M., Bramhill, J., Riede, I., Wright, J.K., Furst, M., Aichele, G. et al. (1980) Lactose carrier protein of *Escherichia coli*: construction and expression of plasmids carrying the Y gene of the *lac* operon. *European Journal of Biochemistry*, 108, 223–231.
- Trepollilli, J.J., Hood, G.A. and Poole, P.S. (2012) What determines the efficiency of  $N_2$ -fixing rhizobium-legume symbioses? *Advances in Microbial Physiology*, 60, 325–389. <https://doi.org/10.1016/B978-0-12-398264-3.00005-X>
- Udvardi, M. and Poole, P.S. (2013) Transport and metabolism in legume-rhizobia symbioses. *Annual Review of Plant Biology*, 64, 781–805.
- Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z. et al. (2003) Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. *The Plant Cell*, 15, 2093–2105.
- White, J.P., Prell, J., Ramachandran, V.K. and Poole, P.S. (2009) Characterization of a gamma-aminobutyric acid transport system of *Rhizobium leguminosarum* bv. viciae 3841. *Journal of Bacteriology*, 191, 1547–1555.
- Yamamoto, S., Inoue, K., Murata, T., Kamigaso, S., Yasujima, T., Maeda, J. et al. (2010) Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *Journal of Biological Chemistry*, 285, 6522–6531.
- Yu, X., Yang, G., Yan, C., Baylon, J.L., Jiang, J., Fan, H. et al. (2017) Dimeric structure of the uracil:proton symporter UraA provides mechanistic insights into the SLC4/23/26 transporters. *Cell Research*, 27, 1020–1033.







## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Botou M, Yalelis V, Lazou P, et al. Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*. *Mol Microbiol*. 2020;114:151–171. <https://doi.org/10.1111/mmi.14503>

Article

# Screening of Heteroaromatic Scaffolds against Cystathionine Beta-Synthase Enables Identification of Substituted Pyrazolo[3,4-c]Pyridines as Potent and Selective Orthosteric Inhibitors

Anna-Maria Fantel <sup>1</sup>, Vassilios Myriantopoulos <sup>1</sup> , Anastasios Georgoulis <sup>2</sup>, Nikolaos Lougiakis <sup>1</sup>, Iliana Zantza <sup>1</sup>, George Lamprinidis <sup>1</sup> , Fiona Augsburger <sup>3</sup> , Panagiotis Marakos <sup>1</sup>, Constantinos E. Vorgias <sup>2</sup>, Csaba Szabo <sup>3</sup> , Nicole Pouli <sup>1</sup>, Andreas Papapetropoulos <sup>1</sup>  and Emmanuel Mikros <sup>1,4,\*</sup> 

<sup>1</sup> Department of Pharmacy, National and Kapodistrian University of Athens, 157 74 Panepistimiopolis, Zografou, Greece; anmafantel@pharm.uoa.gr (A.-M.F.); vmyriant@pharm.uoa.gr (V.M.); nlougiak@pharm.uoa.gr (N.L.); izantza@pharm.uoa.gr (I.Z.); lamprinidis@pharm.uoa.gr (G.L.); marakos@pharm.uoa.gr (P.M.); pouli@pharm.uoa.gr (N.P.); apapapet@pharm.uoa.gr (A.P.)

<sup>2</sup> Department of Biology, National and Kapodistrian University of Athens, 157 01 Panepistimiopolis, Zografou, Greece; tgeorgoulis@med.uoa.gr (A.G.); cvorgias@biol.uoa.gr (C.E.V.)

<sup>3</sup> Pharmacology, Section of Medicine, University of Fribourg, Ch. du Musée 18, 1700 Fribourg, Switzerland; fiona.augsburger@unifr.ch (F.A.); csaba.szabo@unifr.ch (C.S.)

<sup>4</sup> PharmaInformatics Unit, "Athena" Research and Innovation Center, Artemidos 6, 151 25 Maroussi, Greece

\* Correspondence: mikros@pharm.uoa.gr; Tel.: +30-2107274813

Academic Editor: Jorge A.R. Salvador

Received: 30 June 2020; Accepted: 15 August 2020; Published: 16 August 2020



**Abstract:** Cystathionine  $\beta$ -synthase (CBS) is a key enzyme in the production of the signaling molecule hydrogen sulfide, deregulation of which is known to contribute to a range of serious pathological states. Involvement of hydrogen sulfide in pathways of paramount importance for cellular homeostasis renders CBS a promising drug target. An in-house focused library of heteroaromatic compounds was screened for CBS modulators by the methylene blue assay and a pyrazolopyridine derivative with a promising CBS inhibitory potential was discovered. The compound activity was readily comparable to the most potent CBS inhibitor currently known, aminoacetic acid, while a promising specificity over the related cystathionine  $\gamma$ -lyase was identified. To rule out any possibility that the inhibitor may bind the enzyme regulatory domain due to its high structural similarity with cofactor *s*-adenosylmethionine, differential scanning fluorimetry was employed. A sub-scaffold search guided follow-up screening of related compounds, providing preliminary structure-activity relationships with respect to requisites for efficient CBS inhibition by this group of heterocycles. Subsequently, a hypothesis regarding the exact binding mode of the inhibitor was devised on the basis of the available structure-activity relationships (SAR) and a deep neural networks analysis and further supported by induced-fit docking calculations.

**Keywords:** cystathionine  $\beta$ -synthase; hydrogen sulfide; docking-scoring calculations; 7-azido-4-methylcoumarin assay; pyrazolo[3,4-c]pyridine; Sitemap algorithm; back-propagation DNN; thermal shift assay; Bateman module

## 1. Introduction

Endogenously generated hydrogen sulfide, H<sub>2</sub>S, is a signaling molecule of pivotal importance and its biosynthesis in mammalian cells is facilitated by three enzymes, cystathionine  $\beta$ -synthase (CBS), cystathionine

$\gamma$ -lyase (CSE), and 3-mercaptopyruvate sulfurtransferase (3-MST) [1–4]. Cystathionine  $\beta$ -synthase, or CBS, is the first enzyme of the transsulfuration pathway in which the potentially toxic metabolite homocysteine is converted to cysteine [4]. Among others, human CBS catalyzes the condensation of serine and homocysteine to cystathionine and also utilizes cysteine as a substrate to yield H<sub>2</sub>S. H<sub>2</sub>S is considered as an important biological mediator in a similar fashion to nitric oxide (NO) and carbon monoxide (CO) and CBS is responsible for 20%–75% of total H<sub>2</sub>S production and, more specifically, up to 95% of its production in the brain [5]. The enzyme is located in the central nervous system as well as in other tissues (cardiovascular and gastrointestinal). CBS is regarded as a highly attractive drug target, as the development of potent and selective inhibitors may offer important therapeutic traits for a series of pathological states where H<sub>2</sub>S signaling has a key role [6–11]. CBS overexpression has been reported in patients with Down syndrome, and leads to perturbation of H<sub>2</sub>S levels in the human body [12,13]. Interestingly, the CBS gene is located on chromosome 21, the (complete or partial) trisomy which is the genetic basis of the Down phenotype [14,15]. Inhibition of enzymatic activity in these cases can prevent chronic exposure of Down syndrome patients to H<sub>2</sub>S and potential prevention of related neurological deficits [16,17]. Furthermore, CBS up-regulation has been detected in cancer cells including colon cancer, ovarian cancer and lung cancer, amongst others [2,8,18–21]. Under these conditions, tumor growth is facilitated through stimulation of cellular bioenergetics, cell proliferation and angiogenesis [2,22,23].

In terms of structure, CBS is a unique heme-containing pyridoxal 5'-phosphate (PLP)-dependent enzyme that is allosterically activated by *s*-adenosylmethionine (AdoMet or SAM) [24,25]. The full-length human CBS is a homotetramer consisting by 63 kDa subunits, with each subunit comprising 551 residues [26]. CBS adopts a three-domain structure encompassing an N-terminal heme-binding domain, a central catalytic domain (PLP cavity) which is accessible only via a narrow channel, and a smaller C-terminal allosteric regulatory domain (CBS motif, Bateman module) [27,28]. The cofactor PLP is deeply buried in a cleft between the N- and the C-terminal domains adjacent to the active site. On the other hand, the heme binding site is located 20 Å away from the active site. Although heme does not contribute to catalysis, it is believed to contribute to redox regulation of CBS. The enzyme regulatory domain is responsible for interaction with SAM. When SAM binds to this domain, it causes allosteric activation of CBS [21,29–31]. In mammals, the catalytic activity of CBS is increased up to 5-fold by allosteric SAM binding. In its functional form, CBS is believed to form a domain-swapped dimer where the C-terminal regulatory domain of one subunit is atop the N-terminal catalytic domain of the other (protein data bank code: 1JBQ) [24,32]. The domains of each subunit are reoriented upon SAM binding, leading to the aforementioned five-fold increase in catalytic activity.

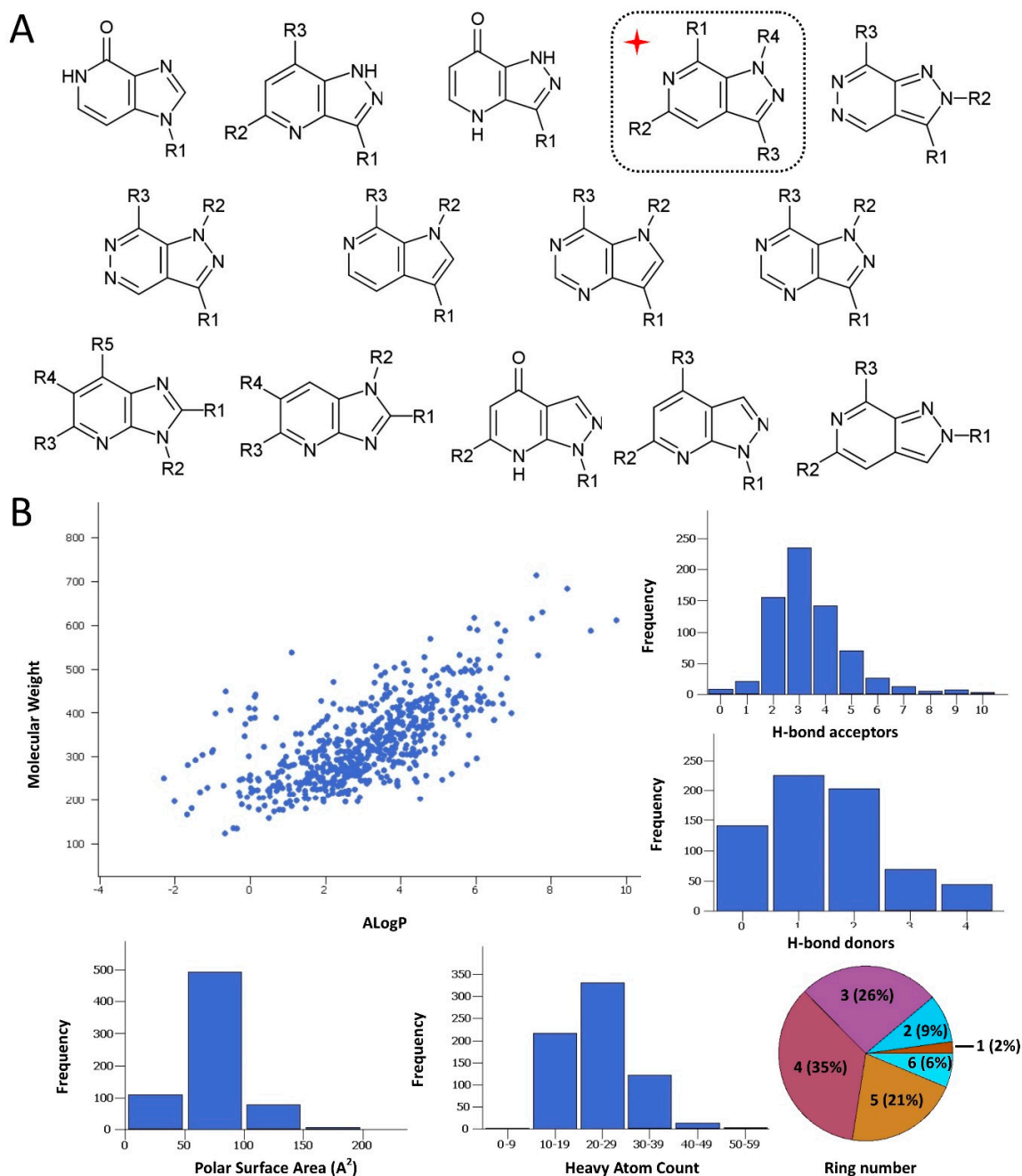
A series of different scaffolds showing a promising inhibitory potential against CBS have been reported so far, with many of them being structurally unrelated synthetic molecules or natural products including flavonoids, coumarins and marine metabolites [2,33–36]. Yet, the most potent CBS inhibitor is a remarkably simple molecule, aminoacetic acid (AOAA), believed to react with PLP toward the formation of a new external aldimine [3,28,37]. However, AOAA is not selective for CBS, as it also inhibits CSE and interacts with other PLP-dependent enzymes as well [9]. Such a simple organic molecule could marginally be considered as a tractable lead for drug development. In this direction, exploring the available chemical space with the objective to discover new scaffolds that can act as promising leads for achieving potent and selective CBS is deemed a highly interesting screening endeavor [38]. The present study describes the evaluation of a unique small molecule library particularly rich in heterocyclic drug-like compounds, the discovery of a pyrazolopyridine derivative as a new selective CBS inhibitor and the elucidation of its inhibitory mode of action. The main results are derived from a series of orthogonal biochemical and biophysical assays and additional support is offered by theoretical approaches, including artificial intelligence modeling and molecular simulations.

## 2. Results and Discussion

### 2.1. Rationale and Compound Selection

While there is no determined cocrystal structure of CBS with any of the published inhibitors, the available crystal structures and previous screening attempts regarding CBS have offered adequate structural data to support the notion that the protein could be characterized as a druggable target. This is true regardless of the relatively low hit rates and weak-to-medium inhibitory activities that have been so far recorded by the aforementioned screening endeavors against CBS. Though functional CBS comprises a complicated quaternary structure bound to several co-factors by means of their corresponding binding sites, the capacity of efficient ligand binding is deemed as particularly promising as to the active site of the enzyme, which comprises a topologically sound cavity for competitive inhibition by drug-like small molecule binding [10]. This was verified by a druggability analysis of the enzyme active site by implementing the Sitemap algorithm (Schrödinger LLC, New York, NY, USA, 2020). The algorithm characterizes binding sites in terms of structural and geometrical features affecting the affinity of potential small molecule binders. Such features are solvent exposure, degree of protein enclosure, hydrophobicity and hydrophilicity, as well as hydrogen bond capacity. Sitemap derives druggability scores (SiteScore and the variant Dscore), where a SiteScore above 0.8 indicates a druggable site and a value equal or higher than 1 denotes a particularly promising pocket. The analysis afforded a high score (SiteScore: 1.012; Dscore: 0.973), suggesting a predominantly promising cavity and further strengthening the hypothesis that, apart from molecules the size of the natural substrates, the enzyme might also interact with larger molecules that could hence function as modulators. In this direction, a focused compound collection was devised and a screening campaign was undertaken, aiming at the discovery of orthosteric CBS inhibitors. The collection was assembled from the in-house repository of the Pharmacy department of National and Kapodistrian University of Athens (NKUA). The NKUA repository (approximately 2000 entries) is a proprietary compound library assembled and enriched over the last years by synthetic molecules, natural products and semi-synthetic analogues derived from a number of phytochemical and synthetic projects. The library is characterized by high structural originality, as most of its components are natural products isolated from biodiversity hot-spots and their related semisynthetic analogues. In the present study, the objective was to sample the variety of heterocyclic scaffolds that are present in the abovementioned repository at the most efficient and timely fashion. The selection was mainly focused in choosing a representative ensemble of nitrogen-containing heteroaromatic scaffolds that could mimic the natural rings of purine or pyrimidine. Such systems comprise the main core of the repository and, most importantly, fall under regions of chemical space that are widely accepted as highly promising in terms of biological activity and privileged structural character. A number of diverse synthetic heterocycles that could be viewed as purine isosteres were considered, including derivatives that possess a central pyrazolopyridine, pyrrolopyridine, pyrazolopyrimidine, pyrrolopyrimidine, pyrazolopyrazole, pyrazinopyridine and pyrazolopyridazine scaffold (Figure 1A). The collection comprised ~600 molecules of high originality. To achieve higher diversity and increase the success rate, this in-house collection was further enhanced by the addition of top-ranked compounds originating from the National Cancer Institute (NCI) Repository and selected by means of virtual screening. To this end, *in silico* screening based on rigid docking of the NCI Repository was performed as a part of the current CBS inhibitor development effort. The virtual screen utilized our previously developed model [10] (1JBQ crystal structure of CBS) and Glide (Glide SP algorithm, Schrödinger LLC, New York, NY, USA, 2020) [39–41]. In this instance, the absence of any inhibitor-CBS co-crystal structure mandates for manual creation of the docking grid (please see Materials section). The 80 compounds which ranked higher in terms of the GScore scoring function were selected from the NCI Repository and ordered. The final collection (NKUA repository and NCI subset) was characterized by a highly satisfactory degree of drug likeness (Figure 1B), thus enabling efficient exploration of the bioactivity landscape of N-containing heterocycles (analyzed by Canvas, Schrödinger LLC, New York, NY, USA, 2020) [42,43]. More specifically, the majority of the compounds

were found to be in overall good compliance with the Rule-of-Five structural features (Figure 1B). As to the structural diversity of the collection, a considerable variety in the side chain decorations and substitution motifs was evident in the number of rings, with 88% of compounds carrying at least one additional ring system apart from the two-ring heteroaromatic core of purine and 1-ring of pyrimidine, respectively.



**Figure 1.** (A) The various substituted heteroaromatic scaffolds comprising the screened collection and the pyrazolo[3,4-c]pyridine core identified as a potent cystathionine  $\beta$ -synthase (CBS) inhibitor (inset denoted by a red cross). (B) Graphical evaluation of the collection drug-likeness in terms of key parameters related with the Rule-of-Five characteristics.

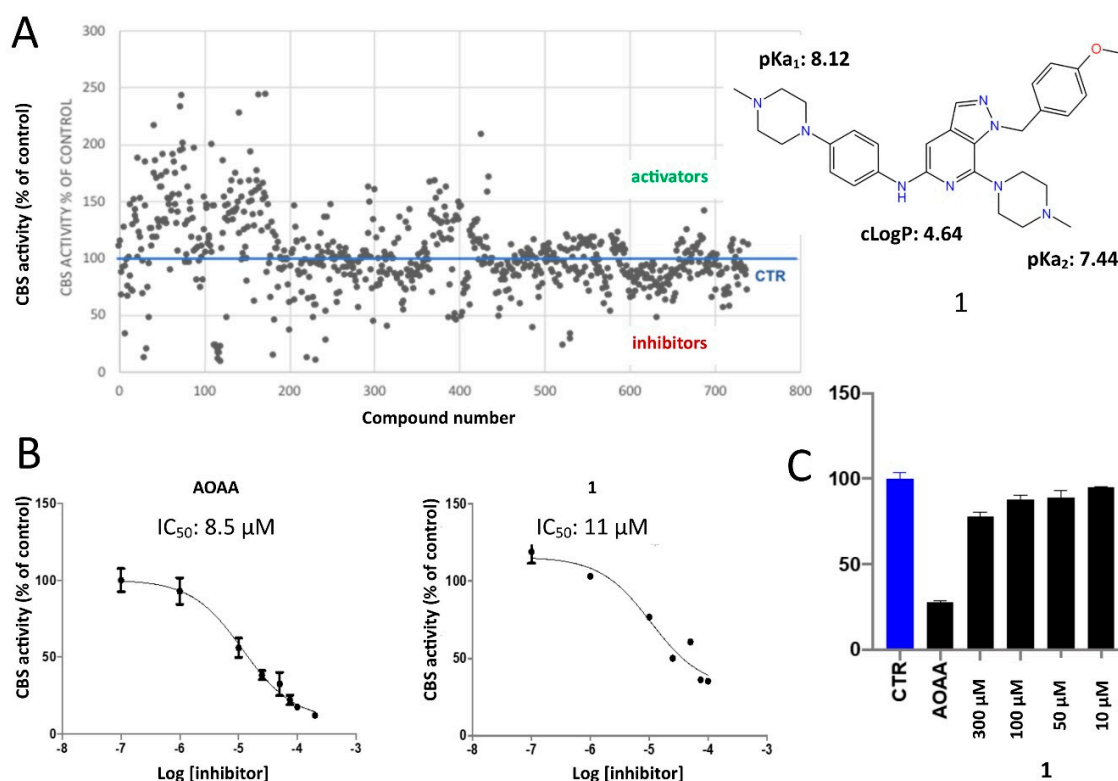


## 2.2. Protein Expression and In Vitro Evaluations

Expression and purification of CBS have been described as complicated because of the strong tendency of the enzyme to aggregate [44–47]. In the present study, the protein was expressed and purified as described earlier, whereas specific modifications of the isolation procedures were undertaken as a means to improve both yield and purity of the enzyme. An expression system of the full-length human CBS was constructed and human CBS was expressed as a glutathione S-transferase (GST)-fusion protein [48]. A three-step purification workflow was organized, first by using affinity chromatography (GSTrap FF column), then by changing buffer and pooling high concentration samples into an anion exchange column (Q Sepharose) and, finally, by gel filtration. Consequently, almost 7.5 mg of highly pure full-length active human CBS were obtained from 1 L of *E. coli* overnight culture. As a means to rule out any possibility that the identified hits could interact with the regulatory domain of CBS due to their structural resemblance with SAM (see discussion Section 2.3), the aforementioned domain was also expressed and purified accordingly. The total collection of the in-house heterocyclic structures (~600 molecules) merged with the top-ranked NCI compounds (80 molecules) were evaluated in vitro for their CBS inhibitory potency.

Several experimental settings, either biochemical or biophysical, have been described for assessing the inhibitory potential of small molecules against CBS. Among them, the methylene blue assay is considered as one of the most robust methods available [49–51]. In the present study, enzymatic activity of the CBS fusion protein was measured by the ability to produce H<sub>2</sub>S in a reaction employing L-cysteine and homocysteine substrates. Quantification of H<sub>2</sub>S was performed by using a standard curve and a H<sub>2</sub>S donor. To confirm the adequate complexation of PLP within the protein during the purification process, the assay was performed in the presence and absence of 0.01 mM PLP. In this study, to identify inhibitors, a quick first screening step was performed at a single inhibitor concentration of 50 μM and compounds that afforded higher than 30% CBS inhibition were further validated by additional assays and structurally analyzed. As expected, a moderate hit rate was determined by the screen, whereas several molecules emerged as CBS activators (Figure 2A). However, among the assayed molecules, the pyrazolopyridine derivative **1** was shown to be the most efficacious inhibitor. This specific molecule was previously synthesized in our lab as a potential inhibitor of angiogenesis [52]. In this molecule, the central pyrazolo[3,4-c]pyridine core is substituted by three functional groups which are present in many bioactive analogues, namely a N<sup>1</sup>-4-methoxybenzyl group attached to the pyrazole ring together with a 4-methylpiperazin-1-yl group and a 4-(4-methyl-piperazin-1-yl)phenylamino group connected to the nucleus (Figure 2B). To confirm the biological activity of the newly discovered pyrazolopyridine hit, the IC<sub>50</sub> value of the inhibitor was determined and directly compared with that of AOAA, calculated in an identical setting. Notably, the dose-response curves were constructed in the presence of 1 mM of L-cysteine and 1 mM homocysteine. The IC<sub>50</sub> value of **1** was 11 μM, whereas the corresponding value of AOAA was 8.5 μM (Figure 2B). Of interest, when the new inhibitor **1** was tested against the related H<sub>2</sub>S-producing enzyme CSE, it was found to possess considerably lower inhibitory activity (Figure 2C). The pyrazolopyridine inhibitor was tested against GST-CSE in three different concentrations in the presence of 1 mM L-cysteine and 0.01 mM PLP, resulting in no significant inhibitory effect.

For validating the most potent hit identified through the primary screen and rule out any possibility of undesirable interferences with the assay conditions leading to a false positive result, a parallel setting for H<sub>2</sub>S detection by the use of 7-azido-4-methylcoumarin (AzMC) was opted for [53]. The inhibitory effect of **1** on CBS was confirmed with this H<sub>2</sub>S detection method as well, although the IC<sub>50</sub> value of **1** against CBS by the AzMC assay was determined at 103 μM. This difference likely reflects intrinsic methodological variations between the two assays that turn to be critical when the assayed compounds are ionized with pK<sub>a</sub> values in very close range to the pH of each setting (pH values: 8.2 for methylene blue; 8 for AzMC assay).

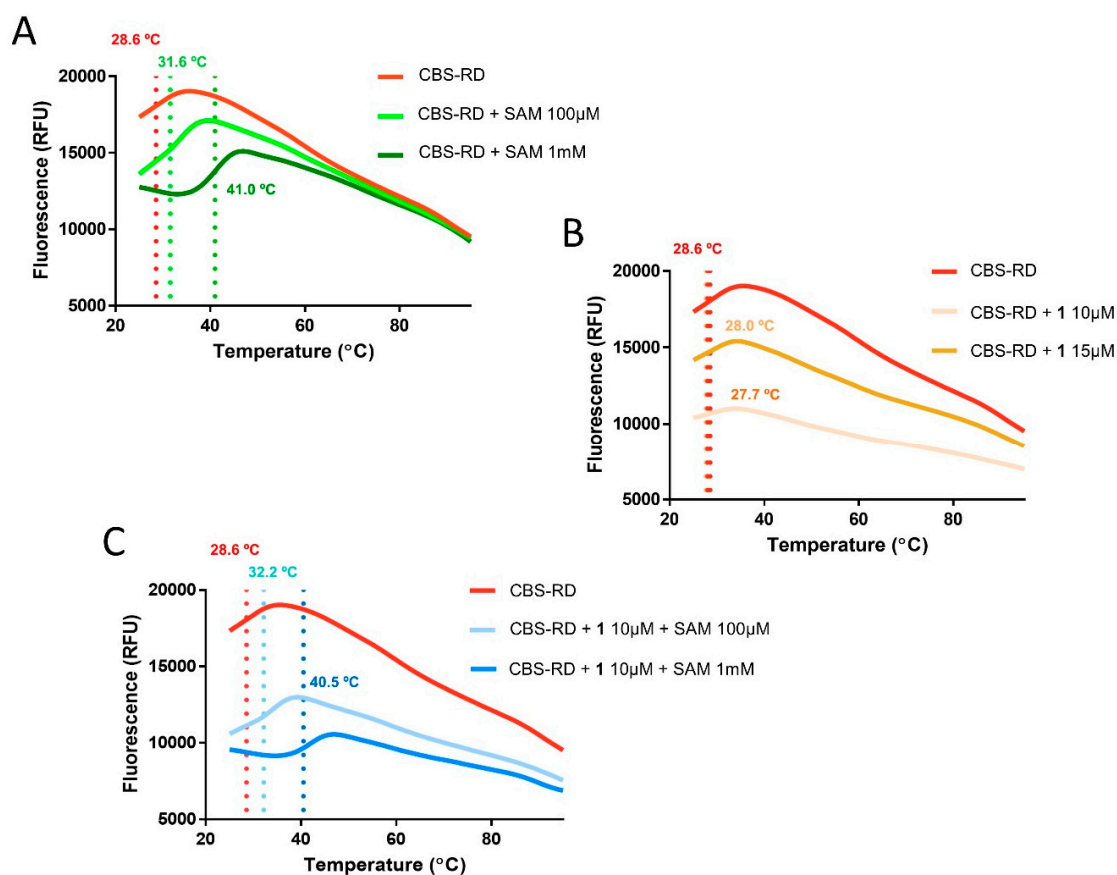


**Figure 2.** (A) A scatter plot summarizing obtained results from the single concentration screen against CBS. (B) The dose-response curves of **1** and AOAA show that the inhibitory activity of the pyrazolo[3,4-c]pyridine analogue is comparable to the most potent known CBS inhibitor, aminoxyacetic acid. (C) Evaluation of inhibitory potential of **1** against the related enzyme involved in H<sub>2</sub>S production cystathionine  $\gamma$ -lyase (CSE), showing specificity of **1** toward CBS as compared to AOAA.

### 2.3. Differential Scanning Fluorimetry

Even though the whole screening strategy aimed at discovering ligands that target the active site of CBS, the non-negligible resemblance of **1** with the regulatory domain co-factor SAM in terms of their heterocyclic scaffolds prompted the exploration of the possibility that the identified hit inhibits CBS via allosteric binding. To address this issue, differential scanning fluorimetry experiments were undertaken as a way to rule out the possibility for binding interactions between **1** and the regulatory domain of CBS (CBS-RD) [54]. The thermal melt results unambiguously reproduced the previously reported extensive stabilization that cofactor binding offers to the protein, with  $\Delta T_m$  values showing a dose-response increase in biologically relevant concentrations of SAM (+2.98 °C, 100  $\mu$ M SAM; +12.43 °C, 1 mM SAM; Figure 3A), but failed to show a statistically significant shift for two different concentrations of **1** (10  $\mu$ M and 15  $\mu$ M; Figure 3B), whereas addition of both SAM and **1** resulted in melting sigmoidals and respective  $\Delta T_m$  values that were highly similar to the corresponding of the SAM/CBS-RD system (+3.62 °C, 100  $\mu$ M SAM, 10  $\mu$ M **1**; +11.94 °C, 1 mM SAM, 15  $\mu$ M **1**; Figure 3C) [30]. The lack of stabilization upon thermal denaturation of either CBS-RD or the SAM/CBS-RD complex in the presence of **1** was a clear indication that no significant binding occurs between the inhibitor and the regulatory component of the enzyme, thus providing validity to the suggestion that **1** is an orthosteric CBS inhibitor.





**Figure 3.** Differential scanning fluorimetry was implemented to identify a possible binding interaction of the inhibitor **1** with the regulatory domain of CBS (CBS-RD). A series of representative melting curves as those shown above, suggest that **1** does not bind CBS-RD. (A) Addition of the cofactor s-adenosylmethionine (SAM) (100  $\mu\text{M}$ , light green,  $T_m$ : 31.6  $^{\circ}\text{C}$ ; 1 mM, dark green,  $T_m$ : 41.0  $^{\circ}\text{C}$ ) results in a dose-response stabilization effect on CBS-RD as compared to the apoprotein (red,  $T_m$ : 28.6  $^{\circ}\text{C}$ ). (B) In contrast, no stabilization is monitored by increasing concentrations of **1** (10  $\mu\text{M}$ , light brown,  $T_m$ : 28.0  $^{\circ}\text{C}$ ; 15  $\mu\text{M}$ , dark brown,  $T_m$ : 27.7  $^{\circ}\text{C}$ ). (C) Co-administration of SAM and **1** (100  $\mu\text{M}$  SAM, 10  $\mu\text{M}$  **1**, light blue,  $T_m$ : 32.2  $^{\circ}\text{C}$ ; 1 mM SAM, 10  $\mu\text{M}$  **1**, dark blue,  $T_m$ : 40.5  $^{\circ}\text{C}$ ) results in a stabilization and response that are practically identical to the case where only SAM is bound to CBS-RD, showing that no binding or any kind of cooperativity takes place between **1** and the SAM-protein complex.

#### 2.4. Structure-Activity Relationships, Neural Networks Modeling and Theoretical Simulations

As a means to explore the structure-activity relationships around the newly discovered CBS inhibitory core more thoroughly, a sub-scaffold search was performed and seven derivatives showing high similarity to **1** were recovered from the in-house repository and successively screened (Table 1). Derivatives **2**, **3**, **4** and **8** showed moderate to weak enzyme inhibitory activity when assayed at 100  $\mu\text{M}$  and a dose-dependent biological response, whereas analogues **5**, **6** and **7** possessing an aromatic amino substituent at R1 were not active even at high concentrations (Table 1). Although none of those analogues showed an improved activity compared to **1**, the results provided preliminary yet interesting structure-activity relationships that were subsequently used in combination with theoretical simulations to suggest a hypothesis as to the structural basis of CBS inhibition by the aforementioned pyrazolopyridine inhibitors. Minor structural modifications were found to be of major importance for the observed biological activity, as in the case of analogue **2**. This is the second most active inhibitor, showing though a considerable decrease of inhibitory potency from 70% to only 30% at 100  $\mu\text{M}$  that follows a simple removal of the 4-methyl group of the piperazine substituent R1 as compared to **1**. The decrease in activity of **2** was thought of as an indication that the ionization and total charge of

studied compounds are factors of key importance for CBS inhibition. More specifically, the structural perturbation involving the transition from **1** to **2** is expected to confer a critical shift not only to the ionization potential of the two compounds, with the basic character of the less potent inhibitor being considerably increased, but also to the trend of protonations given that the emergence of a secondary amine at R1 of **2** reverses the relative order by which the two 4-piperazinyl nitrogens of positions R1 and R2, respectively, are predicted to be ionized (MarvinSketch, ChemAxon). Indeed, in **1** the  $pK_a$  values of the two 4-piperazinyl nitrogens are 7.44 (R1 substituent) and 8.12 (R2 substituent), whereas in **2** the corresponding numbers are 8.84 (R1 substituent, now a secondary amine) and 7.91 (R2 substituent, a tertiary amine). It is also worth noting that the previously described biological activity of a series of pyrazolo[3,4-c]pyridines including **1** and **2** as promising angiogenesis inhibitors was not correlated in terms of structure-activity relationships (SAR) with the CBS-inhibitory potential presented herein [52]. Indeed, the most potent CBS inhibitor **1** was shown to inhibit angiogenesis weakly, whereas **2** characterized by the presence of a 3-phenyl group was a very potent inhibitor of angiogenesis but a marginal inhibitor of CBS.

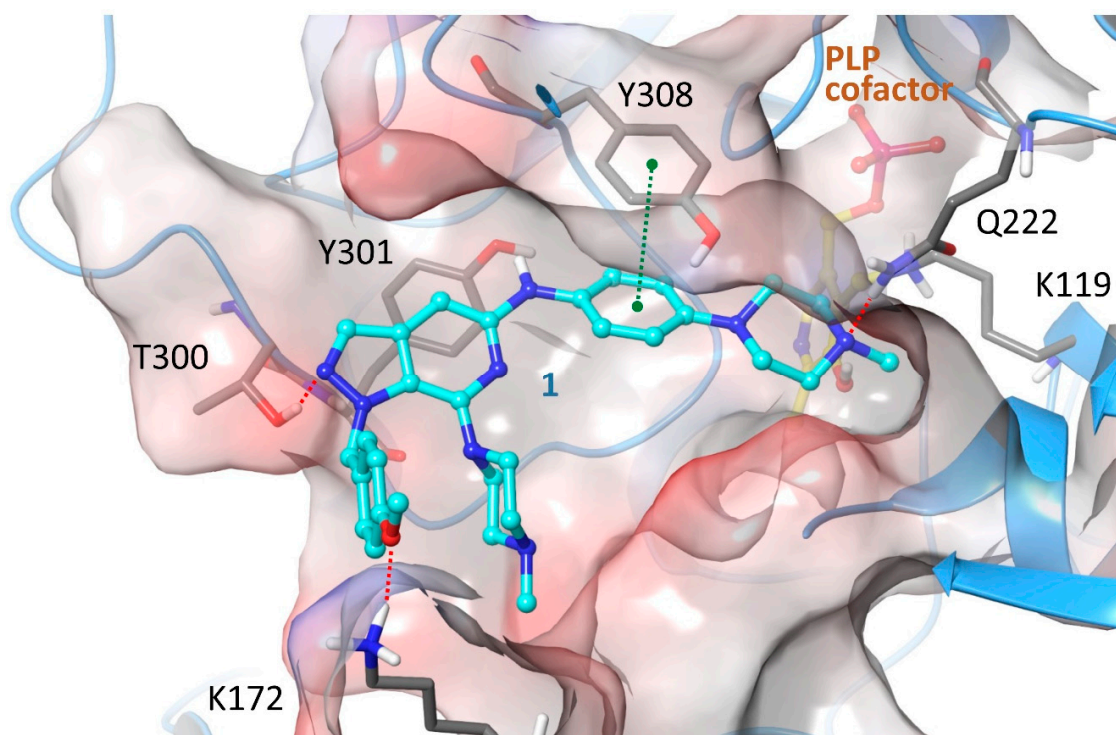
**Table 1.** The inhibitory activity of the 8 pyrazolo[3-c], pyridine derivatives against full length CBS. Percentage enzyme inhibition was measured at a final inhibitor concentration of 100  $\mu$ M.

Compound	R1	R2	R3	R4	Inhibition (%)
1			H		70
2			H		30
3					20
4			H		20
5		N≡C			No inhibition
6				H	No inhibition
7		N≡C	H	H	No inhibition
8		Cl	H		<20

In CBS the SAR observations and DSF results seem to favor a hypothesis where the active compounds bind CBS orthosterically and ligand ionization does not contribute to CBS binding through any direct interaction but, conversely, it opposes complex stabilization. Ionization can seriously impact activity either positively, by enhancing binding through charge-assisted hydrogen bonds and

electrostatic contacts, or negatively by hindering affinity through repulsions or, more frequently, by the large desolvation penalties involved in ionic ligand binding. With the objective to further explore this scenario, a QSAR model was created by utilizing deep neural networks (DNN) and a dataset comprising the hereby reported and previously published CBS inhibitors (60 actives in total) and 613 inactives. The DNN was constructed using one input layer ( $243 \times 673$  points), three hidden layers with 20 neurons each and an output layer with two neurons, namely [0 1] for actives compounds and [1 0] for non-active compounds. The learning workflow was stopped when at least 90% compounds from each group was correctly predicted. The final model resulted in 100% success for active and 94% success for non-active compounds, values that render the prediction significant. The weighting factors between the input and the first hidden layers depict the significance of each descriptor to final decision. Extraction of descriptors with weighting factors greater than 0.8 demonstrated that the most important for group discrimination were those connected to lipophilicity, suggesting it as a key factor for efficient CBS inhibition. This proceeding was in excellent accordance with the hypothesis that the active pyrazolo[3,4-c]pyridine derivatives show a preference to bind CBS as non-ionized molecules. To provide additional strength to the hypothesis, molecular simulations of the interaction between **1** and CBS were undertaken. As previously mentioned, no crystal structure of human CBS has been described in complex with any inhibitor, either substrate-competitive or allosteric. Thus, at present the exact mode of action of known inhibitors is not unambiguously determined, whereas in several cases kinetic results are complex indicating that a mixed-mode inhibition may finally prevail [33]. In this study, *in silico* experiments were carried out in the catalytic site of the enzyme for estimating the validity regarding the competitive inhibition hypothesis in direct comparison with the SAR notions derived via *in vitro* screening and, further, for evaluating the contribution of each of the chemical functionalities present in **1** in CBS binding. For enhancing sampling accuracy, induced-fit flexible protein docking (IFD, Schrödinger LLC, New York, NY, USA, 2020) was applied on top-ranked binding geometries determined by rigid docking (Glide, Schrödinger LLC, New York, NY, USA, 2020) [55–58]. Overall, docking calculations showed that the dominant binding geometry of **1** would agree, yet not unambiguously, with the scenario that charge-assisted interactions are not necessary for efficient **1** binding to CBS (Figure 4). More specifically, the inhibitor binds CBS active site through a network of hydrogen bonds formed between the N1 pyrazole nitrogen of **1** and the side chain hydroxyl of T400, the 4-methoxybenzyl moiety at R4 of the inhibitor and side chain of K172, the phenylamino nitrogen at R2 and Y301, as well as the N4 of piperazine ring at position R1 and the side chain of Q222, whereas the complex is additionally stabilized by a strong  $\pi$ - $\pi$  stacking contact between the phenylamino ring at R4 and Y308. A clear SAR finding was that replacement of the aliphatic piperazine of R1 by an aromatic ring has a detrimental effect on activity, with the phenylamino analogues **5**, **6** and **7** being inactive. The predicted binding geometry could account for this trend to a good extent. The aforementioned group is positioned in a solvent-exposed and highly polar area of the binding site where the existence of a hydrogen-bond acceptor could favor interaction with the water environment and stabilize the ligand through hydrogen bonds with adjacent polar residues E304 and S147. Moreover, the intramolecular interactions between R1 and R4 substituents are expected to influence free ligand energetics, with stronger stacking in the case of **5**, **6** and **7** likely resulting to a shift toward conformations not favoring protein binding, while an R3 substitution bulkier than H seems unfavorable for CBS inhibition, as **3** carrying a phenyl group at this position retains only a fraction of the activity compared to the otherwise commonly substituted **1**. The presence of Lys172 in a close proximity to the bound inhibitor suggests that a coulombic repulsion between the ionic side chain of Lys172 and the positively charged state of the ligand could be a likely explanation for the increased  $IC_{50}$  obtained for **2**, an analogue predicted to be slightly more basic than **1**. It ought to be emphasized though that a less favorable interaction of the cationic form of **1** with the enzyme cannot be excluded. Thus, docking calculations using the IFD protocol and the protonated state of **1** have been performed as well and the most reasonable pose is depicted in Figure S1. Poses where the ligand does not occupy

the inner side of CBS active site by a geometry which would permit contact of the inhibitor and PLP were not considered.



**Figure 4.** The dominant binding geometry of **1** inside the CBS active site as determined by the induced-fit algorithm for docking calculations. The inhibitor is anchored to the enzyme by a number of hydrogen bonds depicted as red dashed lines, whereas a strong stacking interaction between the phenylamino ring of **1** and Tyr308 (shown as green dashed line) further stabilizes the complex. The two piperazine rings of the inhibitor are deprotonated, whereas piperazine of R2 is positioned in close proximity to the catalytic Lys119, covalently attached in this structure to the cofactor pyridoxal aldehyde as a Schiff base. The protein is depicted as a blue ribbon and molecular surface colored according to the electrostatic potential (red: negative; blue: positive). The ligand and protein interaction geometry demonstrates high complementarity and an excellent occupancy of the pocket by the inhibitor.

To summarize, a rationally designed screen was devised for facilitating the exploration of biologically important subdomains in the chemical space that are represented by purine-like and pyrimidine-like heterocycles. This effort concluded in the discovery of a new scaffold with interesting CBS inhibitory properties. The active pyrazolopyridine analogue showed notable potency toward CBS, whereas a competitive inhibition mode was supported by biophysical analysis. The aforementioned activity was accompanied by moderate selectivity over the related enzyme CSE that is involved in sulfur metabolism as well and individually pursued as a drug target. While this selectivity seemed to be environmentally sensitive, it may provide a rationale and a structural basis for establishing CBS- or CSE-selective inhibition or concurrent CBS-CSE inhibition. Based on molecular simulations, fine-tuning of the ionization potential of the novel scaffold can possibly sustain the design of such enzyme-selective, optimized pyrazolopyridine-based inhibitors. Moreover, as the weak inhibitory properties of the discovered hit toward angiogenesis are likely not related to CBS inhibition, development of dual-specificity analogues seems to be a particularly promising perspective for this lead. To this end, the predictive aspect of the herein derived DNN model is expected to sustain optimization of the compound not only by guiding CBS activity enhancement, but also by rationalizing single- or dual-target specificity requisites. To this respect and in terms of chemistry, the pyrazolo[3,4-c]pyridine scaffold presented herein offers a versatile lead and a highly tractable



starting point for developing optimized, biologically active molecules that are in perfect agreement with the principles of druglikeness.

### 3. Materials and Methods

#### 3.1. Expression and Purification of Full Length GST-CBS and His Tag-CBS Regulatory Domain

Cystathionine beta synthase cDNA was cloned into the pGEX-Kg vector to create N-terminal GSH-S-Transferase (GST) fusion protein. For expression, *E. coli* expression cells (BL21 CodonPlus (DE3)) were transformed with the expression vector pGEX-Kg/GST-CBS. A fresh colony was chosen to grow a 5 mL starter culture in LB medium overnight at 37 °C and 180 rpm. All media were supplemented with 100 µg/mL ampicillin. The next day, the starter culture was transferred into 1 L of fresh LB medium containing 75 mg/L d-alanine, until cells reach an OD<sub>600</sub> of 0.6. For induction, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0, 1 mM. The culture was grown overnight at 30 °C and 180 rpm. The cells were subsequently harvested at 6000× g for 15 min. After centrifugation, the cell pellet was resuspended in an ice-cold buffer containing 50 mM Tris pH 8.0, 1 mM EDTA, 25 mM DTT, protease inhibitor, 1 mM PMSF, Triton 0.5%, 5 mM PLP and 500 mM NaCl. Cells were lysed by sonication on ice (sonication power ~35%, 20 s pulse, 45 s pause for 20 min). The cell extract was then centrifuged at 14,000× g for 30 min at 4 °C. After centrifugation, the soluble fraction containing GST-CBS was loaded onto a GSTrap FF 5 mL affinity column. The column was consecutively washed with five column volumes of binding buffer. The protein was eluted with five volume of elution buffer 50 mM Tris-HCL, 10 mM reduced GSH, pH 8.0. Fractions containing highly pure protein were pooled and immediately applied on anion exchange Q sepharose column fast flow. The column was washed with 20 mM Phosphate buffer pH 8.0, 1 mM EDTA and the bound proteins were eluted with a linear gradient of NaCl from 0 to 1.0 M in the same buffer at a flow rate of 2 mL/min. The protein was further purified by gel filtration (Superose 6 prep. 20 mM Phosphate buffer pH 8.0, 1 mM EDTA, 500 mM NaCl and 10% glycerol) Finally, the GST-CBS protein samples were dialyzed and concentrated in 10 mM Sodium Phosphate Buffer pH 8.2, DTT 1 mM and glycerol 5%. The purity of recombinant enzyme was checked by SDS-PAGE on 12% polyacrylamide gels. The regulatory CBS domain was expressed using the Addgene plasmid #73238 and purified as previously described [30].

#### 3.2. Sample Preparation and Library Administration

All tested samples were collected in powder and dissolved in 100% DMSO. The stock solutions were prepared by the use of assay buffer (50 mM Sodium Phosphate pH 8.2) at a final DMSO concentration of 10%. The structural and physicochemical assessment of the collection along with druggability analysis were performed by the use of Canvas software (Schrödinger LLC, New York, NY, USA, 2020).

#### 3.3. H<sub>2</sub>S Detection Using the Methylene Blue Assay

H<sub>2</sub>S detection production was measured using the methylene blue colorimetric assay. To test the inhibitory effect of the compounds against CBS, samples were prepared in 100 µL total volume containing 8 µg of recombinant CBS, 1 mM L-cysteine, 1 mM homocysteine, 0.01 mM PLP and 50 mM sodium phosphate buffer pH 8.2. Initially, after sample preparation all samples were incubated at 37 °C for 1 h. After 60 min, the samples were transferred on ice and 1% zinc acetate was added for trapping H<sub>2</sub>S, followed by the addition of 10% trichloroacetic acid for stopping the enzymatic reaction. Afterwards, freshly prepared solutions of N, N-dimethyl-p-phenylenediamine-sulfate in 7.2 M HCl and FeCl<sub>3</sub> in 1.2 M HCl were added followed by 15 min in the dark resulting in the formation of blue color. The sample solutions were transferred in transparent 96-well flat blank plates and the absorbance was measured at 650 nm. H<sub>2</sub>S quantification was carried out by a standard curve of Na<sub>2</sub>S (0–250 µM).

### 3.4. $H_2S$ Detection Using the 7-Azido-4-Methylcoumarin Assay

Using 50 mM Tris HCl pH 8 solution and black 96-well plate, 0.5  $\mu\text{g}/\text{well}$  of recombinant CBS was incubated 1 h at 37 °C in presence of various final concentrations of **1** in a total volume of 100  $\mu\text{L}$ . CBS substrates were then added to reach 200  $\mu\text{L}$  total volume, 2 mM L-cysteine, 2 mM homocysteine, 0.005 mM PLP and 0.5 mM SAM, as well as the probe 7-azido-4-methylcoumarin (AzMC) (Sigma-Aldrich, Saint Louis, MO, USA) at a final concentration of 10  $\mu\text{M}$  (pH 8.0). Fluorescence was measured in kinetic mode at 37 °C with an Infinite 200 Pro reader (Tecan), with excitation and emission wavelengths of 365 nm and 450 nm, respectively. The  $\text{IC}_{50}$  of the inhibitor was calculated using GraphPad Prism nonlinear fitting curve function.

### 3.5. Differential Scanning Fluorimetry

The regulatory domain of CBS was assayed at 2  $\mu\text{M}$  in a buffer consisting of 10 mM HEPES at pH 7.8, 150 mM NaCl and 10 $\times$  SYPRO orange. Ligand concentrations of 10  $\mu\text{M}$  and 15  $\mu\text{M}$  were assessed, while all experiments were performed without SAM or in the presence of SAM at 100  $\mu\text{M}$  or 1 mM. The BioRad CFX-Connect RT-PCR instrument and white BioRad 96-well plates were utilized. Relative fluorescence intensities were measured by increasing the temperature from 25 °C to 95 °C at 0.5 °C/min and the melting curves along with  $T_m$  values were calculated by non-linear fitting of fluorescence units over temperature by using a four-parameter logistic function as provided in GraphPad Prism v. 7 software.

### 3.6. Deep Neural Networks

The Deep Neural Network (DNN) algorithm employed is back propagation (BP). The dataset was based on 673 compounds tested in vitro and was divided to two subsets. The first contained all molecules ( $n = 60$ ) showing more than 50% inhibition, labeled as active compounds. The rest of molecules ( $n = 613$ ) were labeled non-active compounds. Topological descriptors were calculated from the molecular descriptors workflow as implemented in Schrödinger Suite 2019. (Schrödinger LLC, New York, NY, USA, 2020). Initially 277 topological descriptors were predicted. All descriptors with zero values for all molecules were removed, reducing the final number to 243. The values of each descriptor were normalized using the unit variance procedure. The training parameters for the DNN model were set as following: the number of hidden layers was 3; the number of neurons in each hidden layer was 20; the number of neurons in the output layer was 2, [1 0] for active compounds and [0 1] for non-active compounds; the activation and transfer functions were both sigmoid function; all weights of the network were initialized as random values; the number of iterations ranged from 1000 to 5000; during the gradient descent optimization procedure, the learning rate was 0.01. All mathematical calculations were run using the software package MATLAB v2018b developed by MathWorks, USA (<http://www.mathworks.com>). The descriptors with greater importance were those connected with lipophilicity, partial equalization of orbital electronegativity, and connectivity indexes such as Balaban-type connectivity index J.

### 3.7. Molecular Simulations

The CBS crystal structure deposited under PDB code 1JBQ was selected and prepared for calculations, utilizing the Protein Preparation module of Maestro software (Schrödinger LLC, New York, NY, USA, 2020). The docking grid was prepared accordingly and centered manually around three reference points in the active site of CBS, namely the cofactor PLP and the two active site residues Tyr223 and Gly307. A stepwise workflow was followed, with rigid docking calculations (Glide SP algorithm, Schrödinger LLC, New York, NY, USA, 2020) deriving ensembles of low-energy binding geometries for the studied compounds and the induced-fit algorithm (IFD, Schrödinger LLC, New York, NY, USA, 2020) implemented for optimally exploring the exact binding site of the most interesting analogues in terms of a flexible protein representation. Marvin was used for drawing, displaying and characterizing chemical

structures, substructures and reactions, Marvin v 17.13.0, 2017, ChemAxon (<http://www.chemaxon.com>). The assessment of ionization constants for the compounds was accomplished by MarvinSketch by using the macro mode and static prefix parameters and considering tautomerization and resonance.

### 3.8. Statistical Analysis

Statistics and graphs were created by the use of GraphPad Prism version 8.0. Evaluation and statistical analysis of in vitro screening experiments are represented as the means  $\pm$  SD. Furthermore, one-way ANOVA multiple comparison test was performed and *p*-values less than 0.05 were considered statistically significant.

**Supplementary Materials:** The Supplementary Materials are available online.

**Author Contributions:** Conceptualization, E.M. and A.P.; methodology, E.M., P.M. and A.P.; software, G.L.; validation, E.M., A.P., C.S. and V.M.; formal analysis, E.M., A.P., C.S. A.-M.F. and V.M.; investigation, A.-M.F., A.G., I.Z., F.A., N.L.; resources, E.M., A.P., C.S., N.P., C.E.V.; data curation, A.-M.F., A.G., I.Z., G.L., V.M., N.L.; writing—original draft preparation, E.M., A.P., V.M.; writing—review and editing, E.M., A.P., C.S., C.E.V., G.L., V.M.; visualization, V.M.; supervision, E.M.; project administration, E.M., A.P., V.M.; funding acquisition, E.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation-IKY (A.M.F.). We acknowledge support of this work by the project “INSPIRED” (MIS 5002550), under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) (E.M.). The APC were funded in the framework of a Stavros Niarchos Foundation (SNF) grant (I.Z.).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wang, R. Physiological implications of hydrogen sulfide: A whiff exploration that blossomed. *Physiol. Rev.* **2012**, *92*, 791–896. [[CrossRef](#)] [[PubMed](#)]
2. Szabo, C.; Papapetropoulos, A. International Union of Basic and Clinical Pharmacology. CII: Pharmacological Modulation of H<sub>2</sub>S Levels: H<sub>2</sub>S Donors and H<sub>2</sub>S Biosynthesis Inhibitors. *Pharmacol. Rev.* **2017**, *69*, 497–564. [[CrossRef](#)] [[PubMed](#)]
3. Singh, S.; Banerjee, R. PLP-dependent H<sub>2</sub>S biogenesis. *Biochim. Biophys. Acta* **2011**, *1814*, 1518–1527. [[CrossRef](#)] [[PubMed](#)]
4. Banerjee, R.; Zou, C.G. Redox regulation and reaction mechanism of human cystathionine-beta-synthase: A PLP-dependent hemesensor protein. *Arch. Biochem. Biophys.* **2005**, *433*, 144–156. [[CrossRef](#)]
5. Zhang, J.Y.; Ding, Y.P.; Wang, Z.; Kong, Y.; Gao, R.; Chen, G. Hydrogen sulfide therapy in brain diseases: From bench to bedside. *Med. Gas Res.* **2017**, *7*, 113–119. [[CrossRef](#)]
6. Keller, R.; Chrastina, P.; Pavlíková, M.; Gouveia, S.; Ribes, A.; Kölker, S.; Blom, H.J.; Baumgartner, M.R.; Bártl, J.; Dionisi-Vici, C.; et al. Newborn screening for homocystinurias: Recent recommendations versus current practice. *J. Inherit. Metab. Dis.* **2019**, *42*, 128–139. [[CrossRef](#)]
7. Wen, Y.D.; Wang, H.; Zhu, Y.Z. The drug developments of hydrogen sulfide on cardiovascular disease. *Oxidative Med. Cell. Longev.* **2018**, *2018*, 4010395. [[CrossRef](#)]
8. Wu, D.; Si, W.; Wang, M.; Lv, S.; Ji, A.; Li, Y. Hydrogen sulfide in cancer: Friend or foe? *Nitric Oxide Biol. Chem.* **2015**, *50*, 38–45. [[CrossRef](#)]
9. Asimakopoulou, A.; Panopoulos, P.; Chasapis, C.T.; Coletta, C.; Zhou, Z.; Cirino, G.; Giannis, A.; Szabo, C.; Spyroulias, G.A.; Papapetropoulos, A. Selectivity of commonly used pharmacological inhibitors for cystathionine beta synthase (CBS) and cystathionine gamma lyase (CSE). *Br. J. Pharmacol.* **2013**, *169*, 922–932. [[CrossRef](#)]



10. Druzhyna, N.; Szczesny, B.; Olah, G.; Módis, K.; Asimakopoulou, A.; Pavlidou, A.; Szoleczky, P.; Gerö, D.; Yanagi, K.; Törö, G.; et al. Screening of a composite library of clinically used drugs and well-characterized pharmacological compounds for cystathionine  $\beta$ -synthase inhibition identifies benserazide as a drug potentially suitable for repurposing for the experimental therapy of colon cancer. *Pharmacol. Res.* **2016**, *113*, 18–37. [[CrossRef](#)]
11. Hellmich, M.R.; Coletta, C.; Chao, C.; Szabo, C. The therapeutic potential of cystathionine beta-synthetase/hydrogen sulfide inhibition in cancer. *Antioxid. Redox Signal.* **2015**, *22*, 424–448. [[CrossRef](#)] [[PubMed](#)]
12. Panagaki, T.; Randi, E.B.; Augsburger, F.; Szabo, C. Overproduction of H<sub>2</sub>S, generated by CBS, inhibits mitochondrial Complex IV and suppresses oxidative phosphorylation in Down syndrome. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18769–18771. [[CrossRef](#)] [[PubMed](#)]
13. Marechal, D.; Brault, V.; Leon, A.; Martin, D.; Lopes Pereira, P.; Loaëc, N.; Birling, M.-C.; Friocourt, G.; Blondel, M.; Herault, Y. Cbs overdosage is necessary and sufficient to induce cognitive phenotypes in mouse models of Down syndrome and interacts genetically with Dyrk1 $\alpha$ . *Hum. Mol. Genet.* **2019**, *28*, 1561–1577. [[CrossRef](#)] [[PubMed](#)]
14. Gardiner, K.; Davisson, M. The sequence of human chromosome 21 and implications for research into Down syndrome. *Genome Biol.* **2000**, *1*, reviews0002.1. [[CrossRef](#)] [[PubMed](#)]
15. Skovby, F.; Krassikoff, N.; Francke, U. Assignment of the gene for cystathionine  $\beta$ -synthase to human chromosome 21 in somatic cell hybrids. *Hum. Genet.* **1984**, *65*, 291–294. [[CrossRef](#)] [[PubMed](#)]
16. Kamoun, P.P. Mental retardation in Down syndrome: Two ways to treat. *Med. Hypotheses* **2019**, *131*, 109289. [[CrossRef](#)]
17. Szabo, C. The re-emerging pathophysiological role of the cystathionine-beta-synthase-hydrogen sulfide system in Down syndrome. *FEBS J.* **2020**. [[CrossRef](#)]
18. Bhattacharyya, S.; Saha, S.; Giri, K.; Lanza, I.R.; Nair, K.S.; Jennings, N.B.; Rodriguez-Aguayo, C.; Lopez-Berestein, G.; Basal, E.; Weaver, A.L.; et al. Cystathionine beta-synthase (CBS) contributes to advanced ovarian cancer progression and drug resistance. *PLoS ONE* **2013**, *8*, e79167. [[CrossRef](#)]
19. Kashfi, K. The dichotomous role of H<sub>2</sub>S in cancer cell biology? Déjà vu all over again. *Biochem. Pharmacol.* **2018**, *149*, 205–223. [[CrossRef](#)]
20. Untereiner, A.A.; Pavlidou, A.; Druzhyna, N.; Papapetropoulos, A.; Hellmich, M.R.; Szabo, C. Drug resistance induces the upregulation of H<sub>2</sub>S-producing enzymes in HCT116 colon cancer cells. *Biochem. Pharmacol.* **2018**, *149*, 174–185. [[CrossRef](#)]
21. Modis, K.; Coletta, C.; Asimakopoulou, A.; Szczesny, B.; Chao, C.; Papapetropoulos, A.; Hellmich, M.R.; Szabo, C. Effect of S-adenosyl-L-methionine (SAM), an allosteric activator of cystathionine-beta-synthase (CBS) on colorectal cancer cell proliferation and bioenergetics in vitro. *Nitric Oxide Biol. Chem.* **2014**, *41*, 146–156. [[CrossRef](#)] [[PubMed](#)]
22. Giuffrè, A.; Tomé, C.S.; Fernandes, D.G.F.; Zuhra, K.; Vicente, J.B. Hydrogen sulfide metabolism and signaling in the tumor microenvironment. *Adv. Exp. Med. Biol.* **2020**, *1219*, 335–353. [[CrossRef](#)] [[PubMed](#)]
23. Zuhra, K.; Augsburger, F.; Majtan, T.; Szabo, C. Cystathionine- $\beta$ -synthase: Molecular regulation and pharmacological inhibition. *Biomolecules* **2020**, *10*, 697. [[CrossRef](#)] [[PubMed](#)]
24. Meier, M.; Janosik, M.; Kery, V.; Kraus, J.P.; Burkhard, P. Structure of human cystathionine beta-synthase: A unique pyridoxal 5'-phosphate-dependent heme protein. *EMBO J.* **2001**, *20*, 3910–3916. [[CrossRef](#)] [[PubMed](#)]
25. Ereno-Orbea, J.; Majtan, T.; Oyenarte, I.; Kraus, J.P.; Martinez-Cruz, L.A. Structural insight into the molecular mechanism of allosteric activation of human cystathionine beta-synthase by S-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3845–E3852. [[CrossRef](#)]
26. Pey, A.L.; Martinez-Cruz, L.A.; Kraus, J.P.; Majtan, T. Oligomeric status of human cystathionine beta-synthase modulates AdoMet binding. *FEBS Lett.* **2016**, *590*, 4461–4471. [[CrossRef](#)]
27. Ereno-Orbea, J.; Oyenarte, I.; Martinez-Cruz, L.A. CBS domains: Ligand binding sites and conformational variability. *Arch. Biochem. Biophys.* **2013**, *540*, 70–81. [[CrossRef](#)]
28. Catazaro, J.; Caprez, A.; Guru, A.; Swanson, D.; Powers, R. Functional evolution of PLP-dependent enzymes based on active-site structural similarities. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 2597–2608. [[CrossRef](#)]

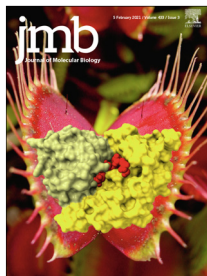
29. Koutmos, M.; Kabil, O.; Smith, J.L.; Banerjee, R. Structural basis for substrate activation and regulation by cystathionine beta-synthase (CBS) domains in cystathionine beta-synthase. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20958–20963. [[CrossRef](#)]
30. McCorvie, T.J.; Kopec, J.; Hyung, S.J.; Fitzpatrick, F.; Feng, X.; Termine, D.; Strain-Damerell, C.; Vollmar, M.; Fleming, J.; Janz, J.M.; et al. Inter-domain communication of human cystathionine beta-synthase: Structural basis of S-adenosyl-L-methionine activation. *J. Biol. Chem.* **2014**, *289*, 36018–36030. [[CrossRef](#)]
31. Hnizda, A.; Spiwok, V.; Jurga, V.; Kozich, V.; Kodicek, M.; Kraus, J.P. Cross-talk between the catalytic core and the regulatory domain in cystathionine beta-synthase: Study by differential covalent labeling and computational modeling. *Biochemistry* **2010**, *49*, 10526–10534. [[CrossRef](#)] [[PubMed](#)]
32. Janošik, M.; Kery, V.; Gaustadnes, M.; Maclean, K.N.; Kraus, J.P. Regulation of human cystathionine beta-synthase by S-adenosyl-L-methionine: evidence for two catalytically active conformations involving an autoinhibitory domain in the C-terminal region. *Biochemistry* **2001**, *40*, 10625–10633. [[CrossRef](#)] [[PubMed](#)]
33. Wang, L.; Cai, H.; Hu, Y.; Liu, F.; Huang, S.; Zhou, Y.; Yu, J.; Xu, J.; Wu, F. A pharmacological probe identifies cystathionine beta-synthase as a new negative regulator for ferroptosis. *Cell Death Dis.* **2018**, *9*, 1005. [[CrossRef](#)] [[PubMed](#)]
34. Niu, W.; Chen, F.; Wang, J.; Qian, J.; Yan, S. Antitumor effect of sikokianin C, a selective cystathionine beta-synthase inhibitor, against human colon cancer in vitro and in vivo. *MedChemComm* **2018**, *9*, 113–120. [[CrossRef](#)]
35. Niu, W.; Wu, P.; Chen, F.; Wang, J.; Shang, X.; Xu, C. Discovery of selective cystathionine beta-synthase inhibitors by high-throughput screening with a fluorescent thiol probe. *MedChemComm* **2017**, *8*, 198–201. [[CrossRef](#)]
36. Zuhra, K.; Sousa, P.M.F.; Paulini, G.; Lemos, A.R.; Kalme, Z.; Bisenieks, I.; Bisenieks, E.; Vigante, B.; Duburs, G.; Bandejas, T.M.; et al. Screening pyridine derivatives against human hydrogen sulfide-synthesizing enzymes by orthogonal methods. *Sci. Rep.* **2019**, *9*, 684. [[CrossRef](#)]
37. Nakai, T.; Nakagawa, N.; Maoka, N.; Masui, R.; Kuramitsu, S.; Kamiya, N. Structure of P-protein of the glycine cleavage system: Implications for nonketotic hyperglycinemia. *EMBO J.* **2005**, *24*, 1523–1536. [[CrossRef](#)]
38. Macarron, R.; Banks, M.N.; Bojanic, D.; Burns, D.J.; Cirovic, D.A.; Garyantes, T.; Green, D.V.; Hertzberg, R.P.; Janzen, W.P.; Paslay, J.W.; et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195. [[CrossRef](#)]
39. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)]
40. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [[CrossRef](#)]
41. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [[CrossRef](#)] [[PubMed](#)]
42. Duan, J.; Dixon, S.L.; Lowrie, J.F.; Sherman, W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* **2010**, *29*, 157–170. [[CrossRef](#)] [[PubMed](#)]
43. Sastry, M.; Lowrie, J.F.; Dixon, S.L.; Sherman, W. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 771–784. [[CrossRef](#)] [[PubMed](#)]
44. Bukovska, G.; Kery, V.; Kraus, J.P. Expression of human cystathionine beta-synthase in Escherichia coli: Purification and characterization. *Protein Expr. Purif.* **1994**, *5*, 442–448. [[CrossRef](#)]
45. Frank, N.; Kent, J.O.; Meier, M.; Kraus, J.P. Purification and characterization of the wild type and truncated human cystathionine beta-synthase enzymes expressed in E. coli. *Arch. Biochem. Biophys.* **2008**, *470*, 64–72. [[CrossRef](#)]
46. Kraus, J.; Packman, S.; Fowler, B.; Rosenberg, L.E. Purification and properties of cystathionine beta-synthase from human liver. Evidence for identical subunits. *J. Biol. Chem.* **1978**, *253*, 6523–6528.
47. Oyenarte, I.; Majtan, T.; Ereno, J.; Corral-Rodriguez, M.A.; Kraus, J.P.; Martinez-Cruz, L.A. Purification, crystallization and preliminary crystallographic analysis of human cystathionine beta-synthase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2012**, *68*, 1318–1322. [[CrossRef](#)]

48. Wingfield, P.T. Overview of the purification of recombinant proteins. *Curr. Protoc. Protein Sci.* **2015**, *80*, 6.1.1–6.1.35. [[CrossRef](#)]
49. Moest, R.R. Hydrogen sulfide determination by the methylene blue method. *Anal. Chem.* **1975**, *47*, 1204–1205. [[CrossRef](#)]
50. Reese, B.K.; Finneran, D.W.; Mills, H.J.; Zhu, M.-X.; Morse, J.W. Examination and refinement of the determination of aqueous hydrogen sulfide by the methylene blue method. *Aquat. Geochem.* **2011**, *17*, 567–582. [[CrossRef](#)]
51. Hartle, M.D.; Pluth, M.D. A practical guide to working with H<sub>2</sub>S at the interface of chemistry and biology. *Chem. Soc. Rev.* **2016**, *45*, 6108–6117. [[CrossRef](#)] [[PubMed](#)]
52. Michailidou, M.; Giannouli, V.; Kotsikoris, V.; Papadodima, O.; Kontogianni, G.; Kostakis, I.K.; Lougiakis, N.; Chatziioannou, A.; Kolisis, F.N.; Marakos, P.; et al. Novel pyrazolopyridine derivatives as potential angiogenesis inhibitors: Synthesis, biological evaluation and transcriptome-based mechanistic analysis. *Eur. J. Med. Chem.* **2016**, *121*, 143–157. [[CrossRef](#)] [[PubMed](#)]
53. Augsburger, F.; Randi, E.B.; Jendly, M.; Ascencio, K.; Dilek, N.; Szabo, C. Role of 3-mercaptopyruvate sulfurtransferase in the regulation of proliferation, migration, and bioenergetics in murine colon cancer cells. *Biomolecules* **2020**, *10*, 447. [[CrossRef](#)] [[PubMed](#)]
54. Niesen, F.H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protocols* **2007**, *2*, 2212–2221. [[CrossRef](#)]
55. Farid, R.; Day, T.; Friesner, R.A.; Pearlstein, R.A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic Med. Chem.* **2006**, *14*, 3160–3173. [[CrossRef](#)]
56. Sherman, W.; Beard, H.S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–84. [[CrossRef](#)]
57. Sherman, W.; Day, T.; Jacobson, M.P.; Friesner, R.A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553. [[CrossRef](#)]
58. Botou, M.; Yalalis, V.; Lazou, P.; Zantza, I.; Papakostas, K.; Charalambous, V.; Mikros, E.; Fliemetakis, E.; Frillingos, S. Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*. *Mol. Microbiol.* **2020**. [[CrossRef](#)]

**Sample Availability:** Samples of the compounds **1**, **3**, **4**, **5** and **7** are available ( $\leq 1$  mg) from the authors.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



# Context-dependent Cryptic Roles of Specific Residues in Substrate Selectivity of the UapA Purine Transporter

Anezia Kourkoulou<sup>1</sup>, Iliana Zantza<sup>2</sup>, Konstantina Foti<sup>1</sup>, Emmanuel Mikros<sup>2</sup> and George Diallinas<sup>1,3\*</sup>

**1** - Department of Biology, National and Kapodistrian University of Athens, Panepistimioupolis, Athens 15784, Greece

**2** - Department of Pharmacy, National and Kapodistrian University of Athens, Panepistimioupolis, Athens 15771, Greece

**3** - Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Heraklion, Greece

**Correspondence to George Diallinas:**\*Department of Biology, National and Kapodistrian University of Athens, Panepistimioupolis, Athens 15784, Greece. [diallina@biol.uoa.gr](mailto:diallina@biol.uoa.gr) (G. Diallinas)

<https://doi.org/10.1016/j.jmb.2021.166814>

Edited by Randy Stockbridge

## Abstract

Members of the ubiquitous Nucleobase Ascorbate Transporter (NAT) family are H<sup>+</sup> or Na<sup>+</sup> symporters specific for the cellular uptake of *either* purines and pyrimidines *or* L-ascorbic acid. Despite the fact that several bacterial and fungal members have been extensively characterised at a genetic, biochemical or cellular level, and crystal structures of NAT members from *Escherichia coli* and *Aspergillus nidulans* have been determined pointing to a mechanism of transport, we have little insight on how substrate selectivity is determined. Here, we present systematic mutational analyses, rational combination of mutations, and novel genetic screens that reveal cryptic context-dependent roles of partially conserved residues in the so-called NAT signature motif in determining the specificity of the UapA transporter of *A. nidulans*. We show that specific NAT signature motif substitutions, alone and in combinations with each other or with distant mutations in residues known to affect substrate selectivity, lead to novel UapA versions possessing variable transport capacities and specificities for nucleobases. In particular, we show that a UapA version including the quadruple mutation T405S/F406Y/A407S/Q408E in the NAT signature motif (UapA-SYSE) becomes incapable of purine transport, but gains a novel pyrimidine-related profile, which can be further altered to a more promiscuous purine/pyrimidine profile when combined with replacements at distantly located residues, especially at F528. Our results reveal that UapA specificity is genetically highly modifiable and allow us to speculate on how the elevator-type mechanism of transport might account for this flexibility.

© 2021 Elsevier Ltd. All rights reserved.

## Introduction

Solute transporters are transmembrane proteins that mediate the controlled translocation of nutrients, metabolites and drugs across biological membranes, and thus their activity is essential for cell nutrition, detoxification, signalling and homeostasis. The activity of transporters, similar to enzymes, is characterized by kinetic parameters such as affinity for substrate(s), rate of

transport, and specificity. In addition, the apparent functional capacity of transporters, also similar to enzymes, depends on regulation of cellular expression and turnover, and in many cases also on oligomerization and allostery.<sup>1</sup> Being polytopic transmembrane proteins, transporter function is also challenged by their continuous and dynamic interaction with membrane lipids, which drives their folding, subcellular trafficking, transport activity and turnover. This in turn signifies that transporter func-

tional evolution should, in principle, run in parallel with the evolution of lipid composition of cellular membranes, specific chaperones necessary for folding, or other effectors mediating proper subcellular trafficking and turnover.<sup>2</sup> Surprisingly, little effort has been dedicated in understanding how the function and specificity of transporters evolve. Addressing the molecular basis of transporter functional evolution, especially in respect to specificity, is interesting not only for understanding basic mechanisms governing protein–ligand interactions in distinct cellular contexts, but is also expected to be crucial in genome mining and predicting novel transporter functions of high biotechnical or biomedical interest. In that direction, we are interested in understanding how the function and specificity of members of the ubiquitous Nucleobase Ascorbate Transporter (NAT) family evolve.

The NAT family (also known as Nucleobase Cation Symporter 2 or NCS2 family) constitutes a large group of transporters present in all domains of life.<sup>3–6</sup> Functionally characterised NATs from bacteria, fungi and plants are specific for nucleobases, but rather surprisingly, mammals and other vertebrates possess NAT homologues that are specific for L-ascorbate transport, defining the evolutionary distinct SVCT1/SVCT2 (SLC23A1/SLC23A2) clade, in addition to nucleobase-specific members, grouped in the SVCT4/SLC23A4 clade.<sup>7–9</sup> Interestingly, humans and higher apes have lost the nucleobase-specific SVCT4/SLC23A4 group, conserving only the SVCT1/2 L-ascorbate-specific homologues.<sup>9</sup> Vertebrates include a third, phylogenetically distinct, highly conserved, clade (the SVCT3/SLC23A3-like group) that is expressed in several tissues, but mostly in the intestine and kidney (<http://www.informatics.jax.org/>), which remains of unknown function and specificity. Plants also include numerous NATs of unknown specificity, some of which seem to transport substrates other than nucleobases or L-ascorbate.<sup>10,11</sup> In conclusion, NATs constitute an excellent group for studying transporter molecular and functional evolution because: a) they are present in all major taxa, b) significant specificity shifts have occurred within the family, and c) a fungal NAT member (namely UapA from *Aspergillus nidulans*; see below) is among the best-studied *eukaryotic* transporters in respect to regulation of expression and turnover, subcellular trafficking, structure–function relationships, mechanism of transport, specificity, and role of lipid interactions.<sup>12,13</sup>

The UapA protein is a specific xanthine-uric acid/H<sup>+</sup> symporter, which can also transport xanthine analogues, including the commonly used drugs allopurinol and oxypurinol.<sup>14–16</sup> Specific single, double or triple mutations can convert UapA into a more promiscuous nucleobase transporter, able to transport or bind, in addition to uric acid and xanthine, adenine, hypoxanthine, guanine or uracil, as well as, several nucleobase analogues.<sup>13,17,18</sup> A high-

resolution crystal structure of UapA dimer, where each monomeric unit consists of a movable *core* and a rather stable *dimerization* domain, trapped in a cytoplasm-facing conformation, has been resolved.<sup>12</sup> This structure confirmed conclusions drawn from genetic studies in respect to residues involved in substrate binding and transport and led to a possible mechanism of transport via sliding of the core domain along the dimerization scaffold.<sup>13</sup>

In the most probable scenario, the mammalian L-ascorbate transporters have evolved from a clade of ancient nucleobase transporters.<sup>9</sup> This shift in substrate specificity, from nucleobases to L-ascorbate, which probably occurred before the divergence of cartilaginous fishes, is a dramatic one as these solutes have very different molecular structures and properties. Nucleobases are planar molecules with relatively low solubility, whereas L-ascorbate is non-planar and highly soluble in water. This difference should necessitate prominent changes in the architecture of the substrate-binding site and substrate translocation trajectory, but maybe also in other regions of the relevant NATs. We still have no insight on the molecular trajectories and mechanisms that led to this specificity shift. Here, we present rational and genetic approaches, combined with relative Molecular Dynamics, addressing the functional role of a conserved NAT motif that forms part of the substrate binding site, within the core domain of UapA.<sup>12,19,20</sup> Our results reveal an unexpected context-dependent cryptic role of specific residues in determining the specificity of UapA towards nucleobases. Our findings also show that genetic conversion of UapA to an L-ascorbate transporter might not be simply achieved by modifications of the NAT signature motif.

## Results

### Role of the NAT signature motif and rationale of the current study

Previous systematic mutational analysis in UapA have shown that a conserved amino acid motif, named the *NAT signature motif*, located in the 10th transmembrane segment (TMS10) of all NATs includes residues critical for substrate binding and specificity (Q408, R417) or transport catalysis (N409, G411, T416). These five residues are practically irreplaceable for wild-type UapA transport activity, although specific conserved substitutions might still possess low activity or substrate binding, albeit with modified kinetics.<sup>13,17,19,21</sup> Random genetic screens have further shown that a sixth residue of the motif, F406, can also moderately contribute to specificity.<sup>18</sup> In all UapA homologues that are specific for nucleobases the NAT signature motif conforms to the consensus sequence **T/S/A/V**<sup>405</sup>-**F/Y/S/L**<sup>406</sup>-**A/S/T/G/V**<sup>407</sup>-**Q/E**<sup>408</sup>-**N**<sup>409</sup>-**X-G**<sup>411</sup>-**X-X-X-X-T**<sup>416</sup>-**R/K/G**<sup>417</sup> (numbering refers to UapA), where highly

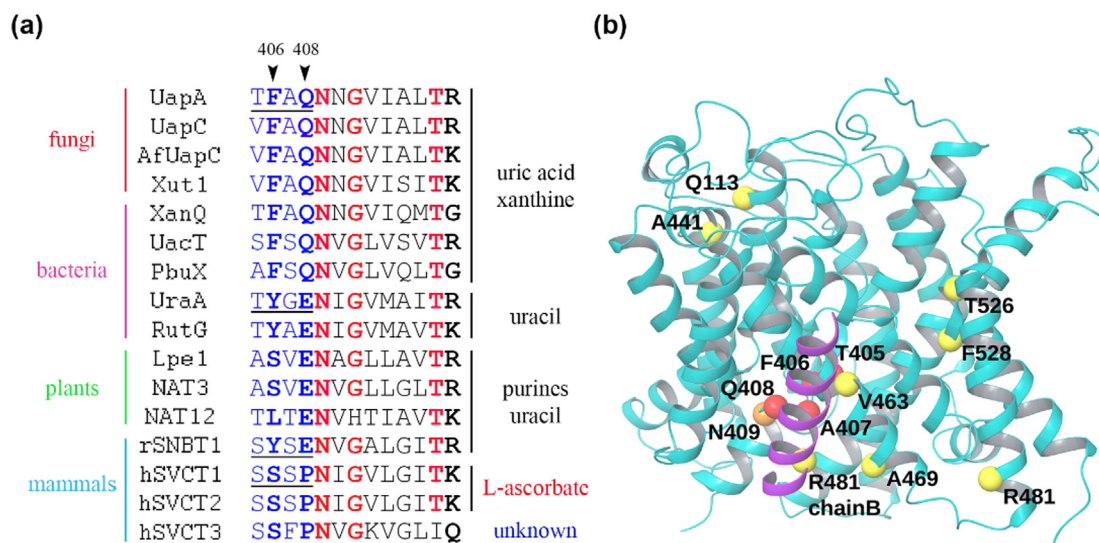


conserved residues are shown in bold (Figure 1(A)). This motif has been shown, via modelling and crystallography, to be a major element of the substrate binding site located in the core domain of UapA<sup>12,20</sup> or UraA, an *Escherichia coli* homologue specific for uracil.<sup>22,23</sup> In the UapA crystal, Q408 forms a strong H-bond via its side chains with the substrate (xanthine). A407 and F406 also interact with xanthine via a backbone nitrogen and  $\pi$ - $\pi$  stacking, respectively, while N409 and T416 seem to be involved in a network of dynamic intramolecular interactions with specific residues in TMS1, TMS3 and TMS8. The flexibility conferred by G411 also proved essential for transport as replacement by Val traps UapA in a non-functional substrate-occluded conformation facing the cytoplasm.<sup>12</sup> Finally, specific substitutions of R417 proved crucial in determining the high affinity of UapA for uric acid compared to xanthine. The aforementioned studies on UapA have been corroborated by mutational and modelling studies with a number of bacterial NATs.<sup>5,24,25</sup>

Interestingly, mammalian NATs that evolved to become highly specific for L-ascorbic acid and

their vertebrate orthologues (SVCT1/SVCT2 group), but also homologues of the vertebrate SVCT3/SLC23A3 group that remains functionally uncharacterized, possess modified versions of the canonical NAT signature motif present in nucleobase-specific NATs.<sup>9</sup> As shown in Figure 1 (A), the two critical differences concern substitutions in the first part of the motif. More specifically, the aromatic Phe/Tyr residue (F406 in UapA) is substituted by Ser, while the polar residue Gln/Glu (Q408 in UapA) is replaced by a Pro. Mutant versions of UapA with single (F406S and Q408P), double (F406S/Q408P) or triple substitutions (T405S/F406S/A407S and F406S/A407S/Q408P) introducing residues present in the L-ascorbate transporters proved to have practically no transport activity for either purines or L-ascorbate.<sup>9</sup> These findings pointed out that additional amino acids, possibly outside the major substrate binding site, might need to be substituted for genetically converting UapA into an efficient L-ascorbate transporter.

Interestingly, the most prominent specificity mutations in UapA, selected by direct genetic



**Figure 1.** Sequence of the NAT signature motif and topological model of UapA highlighting residues critical for specificity. (A) Alignment of UapA, UapC, AfUapC, Xut1, XanQ, UacT, PbuX, UraA, RutG, Lep1, NAT3, NAT12, rSNBT1 (SVCT4), SVCT1, SVCT2 and SVCT3. UapC (XP\_664334), AfUapC (XP\_748919) and Xut1 (XP\_715538) are xanthine-uric acid transporters from *A. nidulans*, *Aspergillus fumigatus* and *Candida albicans*. XanQ (P67444), UacT (Q46821) and PbuX (P42086) are xanthine and/or uric acid transporters from *Escherichia coli* and *Bacillus subtilis*. RutG (AAC74091) and UraA (P0AGM7) are uracil transporters from *E. coli*. NAT3 (NP\_180219) and NAT12 (NP\_850108) are purine-uracil transporters from *Arabidopsis thaliana*. Lpe1 (Q41760) is a maize xanthine-uric acid transporter also showing low affinity for uracil. rSNBT1 (SVCT4/SLC23A4) is a rat purine-uracil transporter. hSVCT1 (Q9UHI7) and hSVCT2 (NP\_005107) are human L-ascorbate transporters. hSVCT3 (NP\_001138361) is a human transporter of unknown specificity, unrelated to nucleobases, nucleosides or L-ascorbic acid.<sup>9</sup> (B) Topology of the UapA monomeric subunit showing residues affecting (enlarging) UapA specificity: V463, A469 located in TMS12, T526 and F528 at end of TMS14 (outer gate), R481 in start of TMS13 (inner gate), and Q113 and A441 at two flexible helical hinges linking the core and dimerization domains. UapA function as a homodimer (monomers A and B). Notice that R481 of monomeric subunit B (only part of TMS13 of this subunit shown in purple) affects specificity by regulating the translocation of substrates via the trajectory of the opposite monomeric subunit A. All other specificity mutations are located in the interface of the core and dimerization domain in each monomer, and thus affect the sliding of the core, including the substrate binding site, along the dimerization domain (for details see<sup>12</sup>).

screens, map outside the major substrate binding site and the NAT signature motif. These concern little conserved residues located at the interface of the core and dimerization domain (V463, A469 in TMS12), at putative outer (T526 and F528 at the end of TMS14) or inner (R481 at the start of TMS13) gates, or at flexible helical hinges linking the core and dimerization domains (Q113, A441) (Figure 1(B)).<sup>12,13</sup> Most of these mutations enlarge the specificity of UapA so that it can transport, besides its natural substrates (xanthine and uric acid), all natural purines, several purine-analogues and uracil. Notably however, no mutation converted UapA into a protein able to transport L-ascorbic acid.

In the present work we a) developed a novel strategy for the systematic genetic analysis of specific UapA residues that are variably conserved in the NAT signature motif of UapA, b) used unbiased genetic screens to isolate revertants of a specific UapA NAT signature motif mutant incapable of transporting its physiological substrates, c) run Molecular Dynamics (MDs) for corroborating major conclusions drawn from mutational and functional analyses.

### Context-dependent role of the NAT signature motif in determining UapA specificity

For the functional analysis of NAT signature motif mutants we used, for the first time, a genetic system that allows the characterization of an extended specificity profile of UapA. This system makes use of a) the strong *gpdA<sub>p</sub>* promoter for expressing *uapA* alleles, some of which might exhibit too little transport activity (<10% of wild-type) to be measured through expression via the native *uapA* promoter, and b) an *A. nidulans* strain, called  $\Delta 7$ , which genetically lacks all seven major endogenous nucleobase-related transporters, thus allowing direct functional assessment of any *uapA* allele introduced by transformation. More specifically, the  $\Delta 7$  strain includes total deletions of the endogenous *uapA*, *azgA*, *uapC*, *fcyB*, *furD*, *furA* and *cntA* genes. UapC is a paralogue of UapA exhibiting similar specificity for xanthine and uric acid, albeit distinct transporter kinetics (i.e. lower affinity for uric acid relative to xanthine, and lower transport capacity in general).<sup>26</sup> AzgA is a high-affinity hypoxanthine/adenine/guanine transporter.<sup>27</sup> FcyB is a high-affinity cytosine transporter that also contributes to purine uptake,<sup>28,29</sup> FurD and FurA are high-affinity uracil and allantoin transporters<sup>30,31</sup> and CntA is the single general nucleoside transporter.<sup>32</sup> Due to deletions in the aforementioned genes, strain  $\Delta 7$  has no detectable transport activity for purines, pyrimidines, nucleosides, allantoin and nucleobase toxic analogues.<sup>33</sup> Thus,  $\Delta 7$  offers the advantage of scoring, based on simple growth tests and subsequently direct transport assays, whether UapA mutant versions, introduced by genetic transformation, can recog-

nize and transport any nucleobase-related solute. Since pyrimidines, unlike purines, cannot be used as N sources in *A. nidulans*, their accumulation by UapA mutants can be scored indirectly in growth tests based on sensitivity of pyrimidine toxic analogues (e.g. 5-fluorouracil/5FU or 5-fluorocytosine/5FC).<sup>33</sup> Allantoin and purine-nucleosides can be used as N sources in *A. nidulans* and thus UapA mutations can be scored by conferring growth on these substrates. Pyrimidine-nucleoside transport can be scored by sensitivity to 5-fluorouridine (5FUd). Expression via the *gpdA<sub>p</sub>* promoter and the use of the  $\Delta 7$  strain have been previously used to reveal cryptic functions in the Fur and Fcy groups of the NCS1 family of transporters.<sup>29,31,33</sup>

Using the above system, we functionally analyzed seven UapA versions mutated in the first four variable residues of the NAT signature motif. These are: UapA-TYAQ, -TFAE, -TYAE, -SYSQ, -SYSE, -SSSQ and -SSSP, where in bold are substitutions relative to the TFAQ (T<sup>405</sup>-F<sup>406</sup>-A<sup>407</sup>-Q<sup>408</sup>) motif present in the wild-type UapA. Among them, UapA-TFAE, -TYAQ and -SSSQ have been previously functionally analyzed via expression from the *uapA* native promoter and in a genetic background (*uapA<sup>-</sup> azgA<sup>-</sup> uapC<sup>-</sup>*) that did not permit testing of nucleobase-related substrates other than purines. This previous analysis showed UapA-TFAE sequence is functional and additionally has acquired the ability to bind (but not to transport) hypoxanthine/guanine, albeit with relatively high affinity,<sup>19</sup> whereas UapA-TYAQ has acquired ability for recognition of all natural purines, but with extremely low binding affinities ( $K_i \geq 1$  mM).<sup>18</sup> Thus, before the present work, it was known that residues at positions 406 and especially 408 (F/Y<sup>406</sup> or Q/E<sup>408</sup>) are critical for UapA specificity. Finally, the UapA-SSSQ mutant has been shown to lack any measurable purine transport capacity.<sup>9</sup>

UapA versions carrying TYAQ, TFAE, TYAE, SYSE, SYSQ, SSSQ or SSSP NAT motif versions were used to examine the role of specific residues at critical positions 406 and 408 (i.e. F/Y/S<sup>406</sup> and Q/E/P<sup>408</sup>) in substrate recognition and transport in an otherwise wild type UapA context, but also in variable combinations with neighboring substitutions at positions 405 or 407, basically introducing amino acids present in NATs specific for a L-ascorbic acid (i.e. Ser or Pro) (see Figure 1 (A)). Notice that SYSE is a sequence frequently present in fungal and metazoan NATs of unknown function and the one predicted to exist in phylogenetically reconstructed ancestral fungal and metazoan NATs (George Diallinas, unpublished analysis). Notice also that SSSP introduces the full set of residues found in the signature motif of L-ascorbic acid NATs.<sup>9</sup> By analyzing these mutants we primarily asked which residues are critical for restricting or enlarging UapA



specificity towards nucleobase-related metabolites, but also tried to understand whether changes in the NAT signature motif can lead to a relative shift of specificity from nucleobases to L-ascorbate.

For each of the mutations studied we selected, purified and analyzed several single-copy transformants. In all cases *uapA* mutations were constructed on a plasmid that carries a fully functional chimeric *uapA-gfp* gene so that UapA subcellular localization and stability could also be analyzed by fluorescent microscopy. **Figure 2** shows growth tests of mutants and control strains on purines, purine-nucleosides or allantoin as N sources, and on media containing toxic concentrations of 5FU, 5FC and 5FUd. As expected, the negative control strain  $\Delta 7$  does not grow on purines, purine-nucleosides or allantoin and is resistant to 5FU, 5FC and 5FUd, which is nearly the mirror image of the growth phenotype of a standard wild-type *A. nidulans* strain. The positive control strain ( $\Delta 7::uapA^+$ ) grows on uric acid and xanthine, but not on other nucleobase-related compounds tested, and is resistant to 5FU, 5FC and 5FUd, as expected given the specificity of UapA. Importantly, none of the mutations analyzed conferred instability or lack of proper localization of UapA in the plasma membrane (**Figure 2**, lowest panel), showing that the mutations studied have not affected UapA folding and subcellular trafficking, which in turn strongly suggested that changes in growth reflect direct effects on UapA activity.

The NAT motif versions that allowed wild-type (i.e. TFAQ-like) growth on physiological substrates of UapA (i.e. uric acid or xanthine) were those including the TYAQ, TFAE and **SYSQ** sequences. Notably, **SYSQ** conferred growth also on high adenine concentration, while TYAQ led to very weak growth on hypoxanthine. In addition, UapA-TYAQ and more prominently UapA-**SYSQ**, led to increased sensitivity to 5FU. This confirmed that position 406 in UapA (F/Y<sup>406</sup>) is critical for specificity, and showed for the first time that the presence of Tyr residue increases recognition of 5FU. Combinations of mutations further revealed the contribution of all NAT residues studied in specificity. For example, the single substitution Q408E (e.g. TFAE) led to no effect on transport specificity, at least within the limit of growth tests, but in combination with other neighboring substitutions (i.e. in TYAE or **SYSE**) led to distinct outcomes. UapA-TYAE could not confer growth on any purine or other nucleobase-related solutes and led to resistance to toxic nucleobase analogues, except from some evidence for relatively increased 5FU sensitivity, as compared to wild-type UapA (UapA-TFAQ). UapA-**SYSE**, despite showing no apparent transport capacity for purines, nucleosides or 5FC and 5FUd, conferred high sensitivity to 5FU, suggesting that it acts as a 5FU carrier. Notably, UapA versions

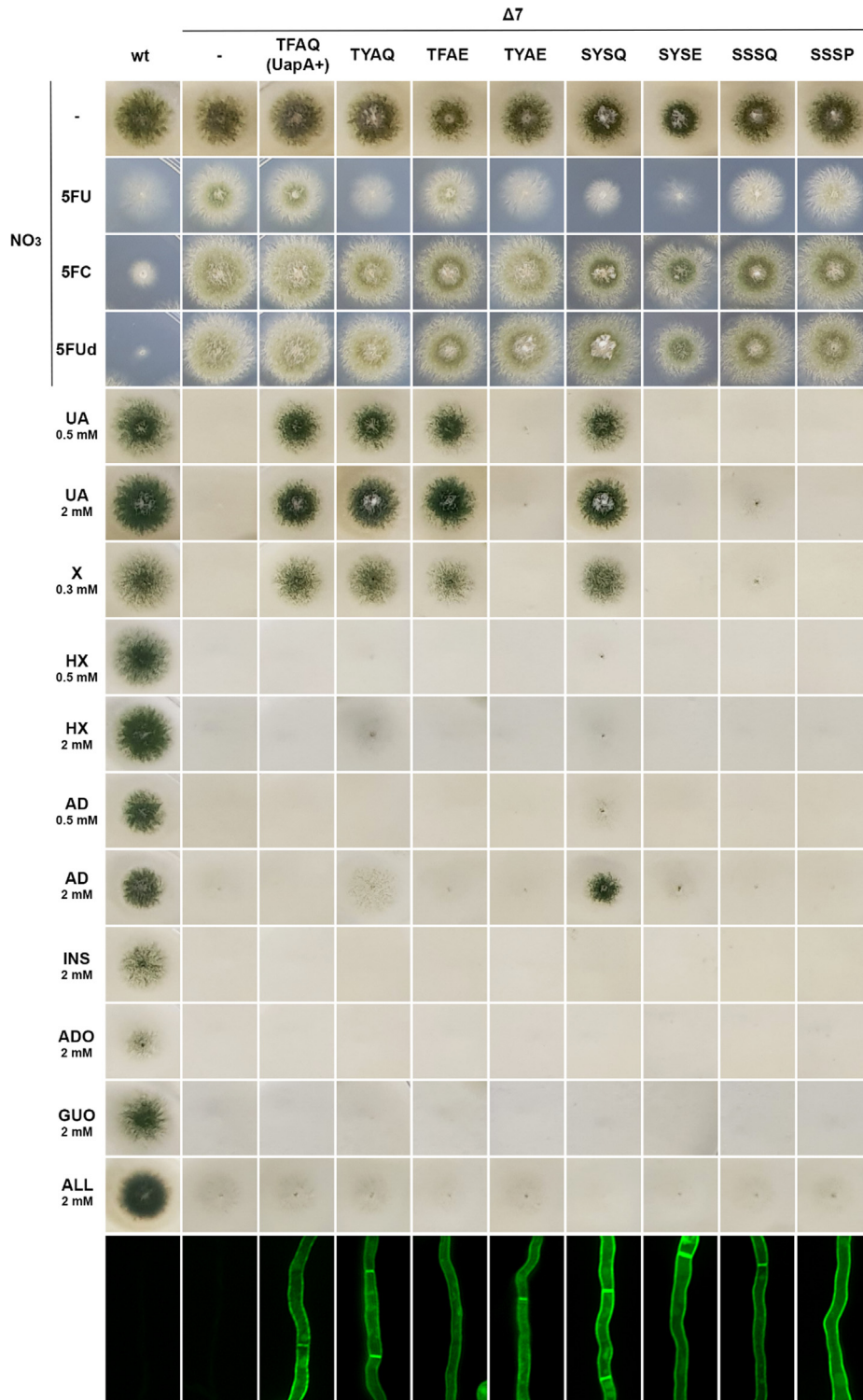
carrying NAT sequences mimicking the motif of L-ascorbic acid transporters showed no apparent capacity for nucleobase transport (e.g. **SSSP**), although the one that conserves Q408 (e.g. **SSSQ**) seemed to confer some growth on high concentration of UA. Both, however, seem to confer some low level accumulation of 5FU not seen in the wild-type UapA control (**Figure 2**). Thus, all mutant versions of UapA studied seem to retain some transport capacity, albeit for distinct substrates. In addition, the growth tests strongly suggested that all four partially conserved residues of the first part of the NAT signature motif of UapA contribute, albeit at different degrees, to specificity.

We performed direct uptake assays, using radiolabelled xanthine, in all mutants studied. **Figure 3(A)** shows that, in line with growth tests, UapA-TYAQ and UapA-**SYSQ** conserve high capacity for xanthine transport (>70% of wild-type UapA), whereas all other mutants show very low relative transport ( $\leq 10\%$  of wild-type UapA). In the UapA mutants that conferred highest sensitivity to 5FU, namely UapA-**SYSQ** and UapA-**SYSE**, we also measured their capacity for accumulation of radiolabelled uracil, which is structurally very similar to 5FU. Surprisingly, none of the two mutants showed measurable UapA-mediated accumulation of uracil. This is discussed further later. We estimated the  $K_i$  values for purines and uracil in **SYSQ** and TYAQ mutants, which conserved considerable transport levels of radiolabeled xanthine, but also in TFAE. **Figure 3 (B)** shows that the binding affinity for the physiological substrates xanthine and uric acid in all three mutants did not differ significantly from those of wild-type UapA (i.e. at the 2–15  $\mu\text{M}$  range). However, unlike wild-type UapA, the mutants exhibited variable, mostly very low but measurable, binding affinity for adenine, hypoxanthine or uracil.

We took advantage of the fact that several mutants tested are sensitive to 5FU to investigate whether some of them, despite being unable to transport efficiently purines, can still bind purines or purine-related compounds. For this, we performed *in vivo* competition growth assays where we scored the reversion of 5FU inhibition in the presence of excess purines or pyrimidines in the growth medium (**Figure 3(C)**). This test showed that 5FU sensitivity mediated by UapA-**SYSQ** is significantly reversed by most nucleobases tested, while that of UapA-**SYSE** is also reduced, but at a lower degree. Notice also that UapA-TYAQ, which, otherwise, is very similar in phenotype with UapA-**SYSQ** (**Figs. 2, 3A and 3B**), appears to be less efficient in reversing the 5FU sensitivity with some nucleobases (e.g. xanthine). Finally, the moderate 5FU sensitivity conferred by UapA-**SSSQ** and UapA-**SSSP** was also reversed by uracil and less so by uric acid.

In an effort to investigate whether any of the mutants analyzed also alters the capacity of UapA in recognizing L-ascorbic acid, we tested the growth of all mutants in the presence of relatively high concentration of L-ascorbic acid (50 mM or 1%). Notice that L-ascorbic acid is not taken up by

*A. nidulans* cells as there is neither a specific endogenous transporter, nor evidence for passive diffusion (G. Diallinas, unpublished observations). Thus, we thought that, if any of the UapA mutations conferred a capacity for L-ascorbate accumulation, this might be reflected in an effect



on the growth phenotype of relative mutants, given the well-known anti- or pro-oxidant activity of L-ascorbate.<sup>34</sup> None of the UapA mutants altered the mild toxic effect (i.e. reduced conidiation) of L-ascorbic acid on *A. nidulans* growth (not shown). This result was not however conclusive on whether any of the UapA mutants can transport or at least bind L-ascorbic acid (discussed later).

Based on all aforementioned results, Table 1 summarizes the function and specificity of the different UapA versions analyzed. Notice that UapA-SYSQ is the most efficient and promiscuous UapA version, being able to recognize and transport purines and 5FU, and it can also at least bind uracil. In contrast, UapA-SYSE, which differs only at a single residue with UapA-SYSQ (i.e. Q408E), proves to be a mutant that has lost its capacity for efficient purine transport, but acquired a prominent high capacity for 5FU accumulation. However, also notice that if we compare the relative transport function and specificity of UapA-TFAQ versus UapA-TFAE, which differ in the same residue as UapA-SYSQ differs from UapA-SYSE, the outcome is different, as replacement Q408E does not lead to total loss of purine uptake or acquisition of high capacity for 5FU transport. This observation clearly shows that UapA specificity is determined by complex interactions of residues in the NAT signature motif, difficult to predict *a priori*.

### Combination of the SYSE sequence with selected UapA mutations leads to distinct transport profiles

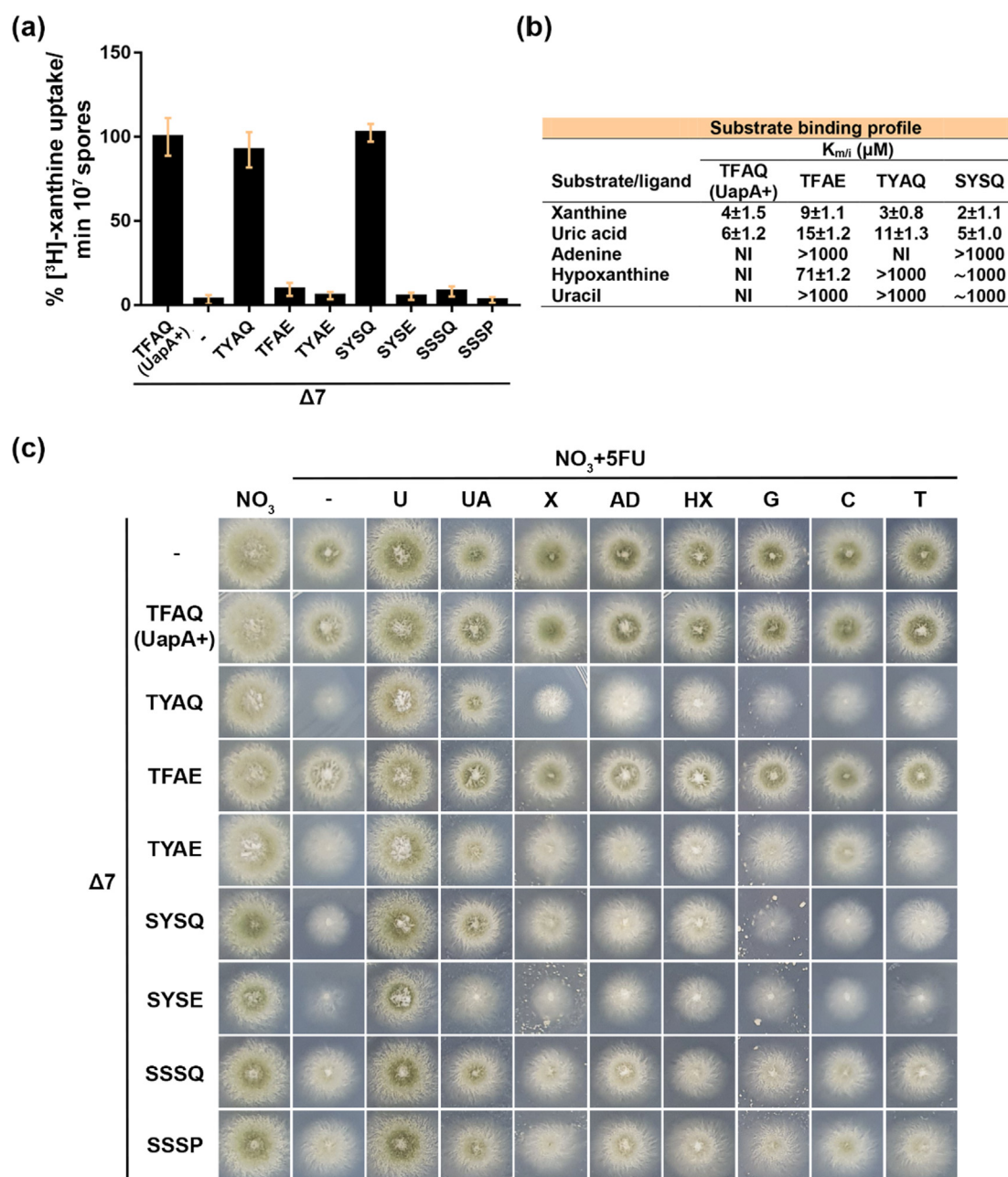
A SYSE consensus sequence is naturally present in several NATs, including the metazoan rSNBT1 that functions as promiscuous nucleobase transporter.<sup>7</sup> However, a SYSE sequence within the context of UapA led to inability for purine transport and acquisition of a capacity for 5FU accumulation, strongly suggesting that the NAT motif interacts functionally and in a context-dependent manner with other residues in NATs. To address this issue, we combined the SYSE sequence with

mutations R481G, T526L, F528M, R481G/T526L and R481G/F528M. These mutations enlarge UapA specificity to transport, besides its natural substrates uric acid and xanthine, all purines, several purine analogues and uracil, albeit with low affinity.<sup>17,18</sup> Topological evidence based on the UapA crystal structure and Molecular Dynamics support that the relevant residues are parts of outer (T526 and F528) or inner (R481) gating elements located at the two sides of the substrate translocation trajectory in UapA.<sup>12</sup>

Figure 4(A) summarizes growth tests of combinations of SYSE with these mutations. Notice that in all cases the combined mutations did not affect the proper localization of UapA to the PM (see right column). As a general picture, the increased promiscuity of UapA in respect to purine accumulation caused by F528M, T526L, R481G, R481G/F528M or R481G/T526L (see rows 3–7) was significantly suppressed in the presence of SYSE (see rows 9–13). Only F528M/SYSE conferred detectable growth on purines, especially on high concentrations of adenine. In respect of 5FU toxicity, which was very high in UapA-SYSE and some specificity mutants (T526L, R481G/T526L or R481G/F528M), the simultaneous presence of the SYSE sequence and specificity mutations led to variable toxicity levels that could have not been predicted *a priori*. For example, the presence of R481G led to partial suppression of the ability of UapA-SYSE to accumulate 5FU, as evidenced by increased resistance to this toxic analogue exhibited by UapA-R481G/SYSE. The triple mutant UapA-R481G/T526L/SYSE showed moderate sensitivity to 5FU, similar to UapA-SYSE, thus ‘masking’ the hypersensitivity of the parental double UapA-R481G/T526L mutant, while UapA-R481G/F528M/SYSE remained hypersensitive to 5FU, similar to the relative double mutants UapA-R481G/F528M or UapA-F528M/SYSE. Furthermore, mutation F528M increased sensitivity to 5FU in the context of SYSE or/and R481G, while when it stands alone conferred reduced sensitivity to this drug. Notice also that

**Figure 2.** Context-dependent role in UapA specificity of variably conserved residues in the NAT signature motif. The image shows growth tests of UapA mutants and a control strain (referred by the sequence of their NAT motif, e.g. wild-type UapA is TFAQ, etc.) on toxic nucleobase/nucleoside analogues at 100  $\mu$ M 5FU, 50  $\mu$ M 5FC, 10  $\mu$ M 5FUd (5FU is 5-fluorouracil; 5FC is 5-fluorocytosine; 5FUd is 5-fluorouridine), or on different purines (UA is uric acid; X is xanthine, HX is hypoxanthine; AD is adenine), nucleosides (INS is inosine; ADO is adenosine, GUO) is guanosine) or allantoin (ALL), as sole N sources. Concentrations of purine-related substrates used as N sources are in the range of 0.3–2.0 mM. Toxic analogues are scored in the presence of 10 mM sodium nitrate ( $\text{NO}_3$ ) as nitrogen source. The lower panel shows the subcellular localization of UapA mutants, as evidenced via a GFP tag, in all strains analyzed. All UapA mutant versions shown are analyzed in the isogenic background of  $\Delta 7$  strain (see text). A standard wild-type (wt) strain possessing all relative endogenous nucleobase-related transporters is also shown for comparison in the first column. Growth tests were performed at 37 °C and pH 6.8. Lower panel: Inverted fluorescence microscopy images showing localization of the GFP-tagged UapA constructs. Samples were grown for 18 h at 25 °C on MM with  $\text{NO}_3$  as N source.





**Figure 3.** Transport activity and specificity of UapA NAT motif mutant versions (A) Comparative transport of 0.1  $\mu\text{M}$  of radiolabelled  $^3\text{H}$ -xanthine expressed as percentages of initial uptake rates (V) compared to the wild-type (UapA) rate taken as 100%, performed at 37  $^{\circ}\text{C}$  (see Materials and methods). Results shown are the average of triplicate measurements from 3 independent assays with a SD shown as error bars. (B)  $K_m$  (xanthine) or  $K_i$  (uric acid, adenine, hypoxanthine, uracil) values estimated via competitive inhibition assays of radiolabelled xanthine uptake (Kryptou and Diallinas<sup>26</sup>; see also Materials and methods) in selected mutants. NI stands for no evidence of radiolabeled xanthine transport inhibition in the simultaneous presence of excess ‘competitor’ nucleobase supplied to up to 1 mM. >1000 means detectable inhibition, but the  $K_i$  value could not be determined given that nucleobases are little soluble at concentrations  $\geq 2$  mM. Results shown are averages of triplicate measurements from 3 independent assays with SD values shown. (C) *In vivo* competition assays scoring the reversion of 5FU (100  $\mu\text{M}$ ) growth inhibition in the presence of excess purines or pyrimidines (2 mM) in the growth medium. Strains and growth conditions are as described in Figure 2. U is uracil, G is guanine, C is cytosine and T is thymine.

mutants UapA-F528M, -R481G and -**SYSE** are significantly less sensitive to 5FU than their relative double or triple mutants. Finally, while

single mutation T526L confers hypersensitivity to 5FU, this is compromised in the context of R481G/**SYSE**. In conclusion, the functional effect

Table 1 Transport function and specificity of UapA NAT motif mutants

NAT motif version	UA	XA	AD	HX	5FU (U)	5FC (C)	5Fud (Ud)	AA	Comments on UapA function
TFAQ	+++	+++	-	-	-	-	-	-	Wt function/specificity
TYAQ	+++	+++	+	+	++	-	-	-	increased <i>binding/transport</i> promiscuity
TFAE	++	++	+/-	++	+/-	-	-	-	increased <i>binding/transport</i> promiscuity, reduced affinity and transport of UA, XA
TYAE	-	-	-	-	+/-	-	-	-	loss of purine <i>binding/transport</i> , but gain of moderate 5FU transport
SSSQ	+/-	-	-	-	+	-	-	-	loss/reduction of purine <i>binding/transport</i> , but gain of 5FU transport
SYSE	-	-	-	-	+++	-	-	-	loss/reduction of purine transport but gain of efficient 5FU transport
SYSQ	+++	+++	++	+/-	++	-	-	-	increased promiscuity for purines and 5FU
SSSP	-	-	-	-	+/-	-	-	-	loss of purine, some gain of very low 5FU transport

Underlined are amino acid replacements in mutant versions of UapA.

+++ signifies wild type-like substrate binding and transport activity. – stands for  $\leq 5\%$  transport capacity of wild-type UapA and no indication of binding of nucleobases (up to 1 mM) or toxic nucleobase analogues (50–100  $\mu\text{M}$ ).

In between values, +/-, +, ++ reflect compromised binding/transport (10–50% of wild-type UapA).

UA is uric acid, XA is xanthine, AD is adenine, HX is hypoxanthine, , 5FU is 5-fluorouracil, U is uracil, 5FC is 5-fluorocytosine, C is cytosine, 5Fud is 5-fluorouridine, Ud is uridine, AA is L-ascorbic acid (1%).

This table is based on functional studies (growth tests, subcellular localization uptake measurements and transport competition assays) of different mutants expressed in a  $\Delta 7$  genetic background (see Figs. 2 and 3).

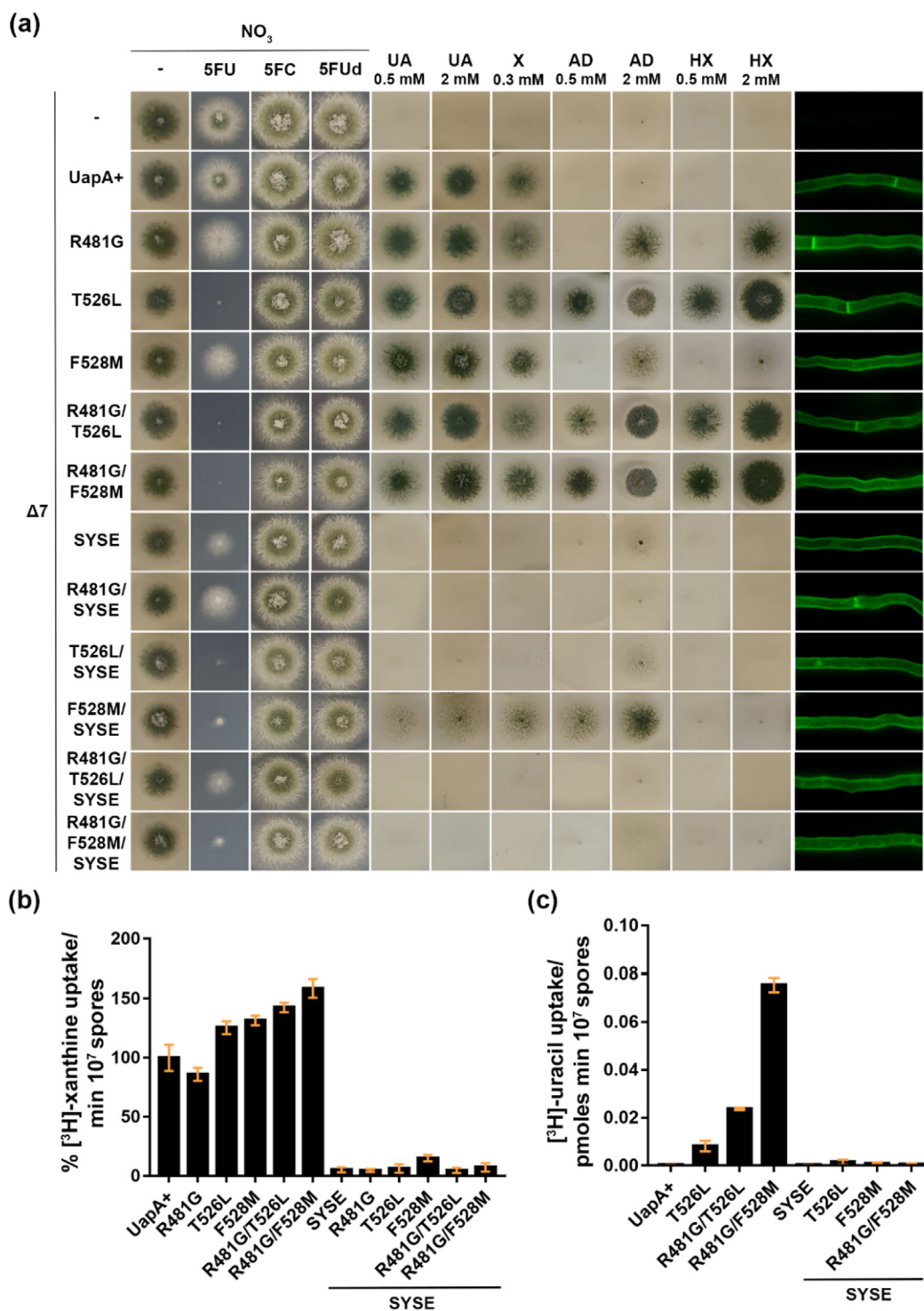
None of the above mutant confers AA, 5FU or nucleoside or allantoin transport

of the **SYSE** sequence in the context of UapA proved highly dependent on the presence of mutations at residues on R481, T526 or F528. Notably, none of the above combinations of mutations conferred to UapA an ability to transport nucleosides or allantoin. Additionally, we did not detect any change in *A. nidulans* growth of strains containing the aforementioned combinations of mutations in media containing 0.1–1% (w/v) of L-ascorbic acid (not shown).

To better rationalize the above observations, we performed radiolabeled xanthine transport assays in selected mutants (Figure 4(B)). The presence of the **SYSE** sequence dramatically reduced UapA-mediated xanthine transport, despite the presence of specificity mutations, which when present in an otherwise wild-type NAT context (i.e. TFAQ instead of **SYSE**) not only enlarged specificity, but also lead to increased xanthine transport rates. Given that R481G, T526L, F528M, **SYSE** mutations, alone or in specific combinations, conferred increased capacity for 5FU accumulation, we tested whether the corresponding strains also have a capacity for UapA-mediated transport of uracil, which is very similar in structure with its toxic analogue 5FU. Results in Figure 4(C) showed that the **SYSE** sequence does not confer to UapA measurable transport activity for uracil, and also seems to suppress detectable uracil transport in UapA-T526L, -R481G/T526L or -R481G/F528M mutants, which possess a wild-type NAT signature motif. Taken together, results shown in Figure 4 (A–C) suggest that introducing the **SYSE** sequence in the NAT signature motif of UapA

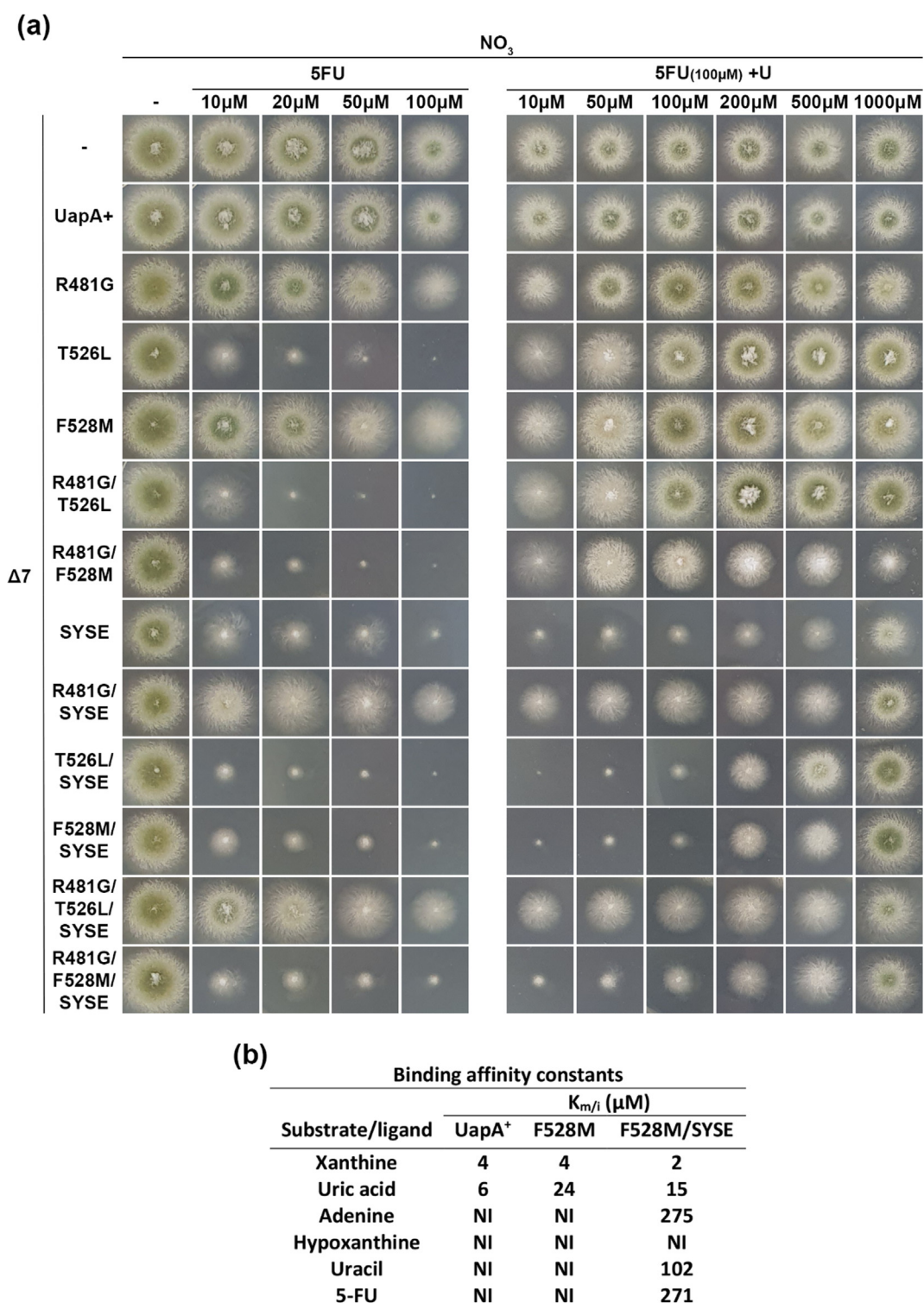
leads to reduced activity for purine or uracil transport. This negative effect of **SYSE** seems only partially suppressed in the context of F528M. On the other hand, versions of UapA carrying **SYSE** alone or in combination with other mutations lead to accumulation of 5FU, thus revealing that the transporter is transport-active for this toxic pyrimidine analogue.

The most probable explanation for the observed apparent inability of UapA-**SYSE** to transport uracil, despite its capacity to lead to significant 5FU accumulation, is that this UapA version has high-affinity specifically for 5FU, but low affinity for uracil and nucleobases. Similar situations, where transporters have increased affinities for chemical analogues of physiological substrates are known to exist, and seem to be due to additional interactions of substituted positions in the analogues.<sup>35,36</sup> To test this possibility, and given that radiolabelled 5FU is not available, we performed *in vivo* competition assays in growth tests where we follow the level of 5FU reversion of toxicity by increasing concentrations of uracil. For this, we first established the concentration range that 5FU toxicity is best visualized. We performed these tests not only for UapA-**SYSE**, but also for several of the specificity mutants used in this work. The left panel of Figure 5(A) shows that most mutants (T526L, R481G/T526L, R481G/F528M, **SYSE**, T526L/**SYSE**, F528M/**SYSE**, R481G/F528M/**SYSE**) were very sensitive to 5FU at concentrations as low as 10  $\mu\text{M}$ . Some mutants (F528M, R481G, R481G/**SYSE** or R481G/T526L/**SYSE**) were more mildly sensitive to 5FU, showing significantly reduced growth only at 100  $\mu\text{M}$  of the toxic analogue. In gen-



**Figure 4.** Combination of the **SYSE** NAT motif sequence with specificity mutations leads to functionally distinct UapA versions (A) Growth tests of strains expressing UapA versions carrying the **SYSE** sequence in the NAT motif combined with selected specificity mutations (F528M, T526L, R481G, R481G/F528M or R481G/T526L). Growth conditions and control strains are as described in Figure 2. Inverted fluorescence microscopy images confirming the proper localization of all UapA mutant versions to the PM is shown on the last right column. (B) Comparative <sup>3</sup>H-xanthine (0.1 μM) transport rates in strains expressing UapA versions combining **SYSE** with selected specificity mutations (F528M, T526L, R481G, R481G/F528M or R481G/T526L). Details are as in Figure 3A. Standard deviation is depicted with error bars. (C) Comparative <sup>3</sup>H-uracil (0.1 μM) transport rates in strains expressing UapA mutant versions that conferred increased capacity for 5FU accumulation. Details are as in Figure 3A. Standard deviation is depicted with error bars.





**Figure 5.** Relative recognition of 5FU versus uracil as established by growth tests in UapA mutants. (A) Left panel: Concentration-dependent 5FU toxicity in the mutants. Right panel: *In vivo* competition of 100  $\mu$ M 5FU toxicity by increasing concentration of uracil (range 10–1000  $\mu$ M) in the mutants. Growth conditions are as described in Figure 2. (B)  $K_m$  (xanthine) or  $K_i$  (uric acid, adenine, hypoxanthine, uracil, 5FU) values estimated via competitive inhibition assays of radiolabelled xanthine uptake (<sup>26</sup>; see also Materials and methods) in selected mutants. Details are as described in Figure 3B. Results shown are averages of triplicate measurements from 3 independent assays with a SD of < 20%.



eral, mutants possessing the **SYSE** sequence are more sensitive to 5FU, compared to wild-type UapA. We thus used 100  $\mu\text{M}$  5FU to perform *in vivo* competition assays by uracil. The right panel of [Figure 5\(A\)](#) shows that uracil competed 5FU toxicity, in all cases, but the level of reversion of toxicity was different in the mutants analysed. In mutants that are either highly (T526L, R481G/T526L or R481G/F528M) or moderately (R481G or F528M) sensitive to 100  $\mu\text{M}$  5FU due to mutations outside the NAT motif, uracil suppressed toxicity at a concentration range of 50–100  $\mu\text{M}$ , which suggested that 5FU and uracil are recognized with comparable affinities. However, in most mutants having the **SYSE** sequence only 10-fold excess uracil (1000  $\mu\text{M}$ ) could significantly suppress the toxicity of 100  $\mu\text{M}$  of 5FU. These results suggested that the presence of the SYSE sequence might lead to either higher affinity for 5FU or lower recognition of uracil, or both.

To address this issue directly, we estimated the  $K_i$  values for purines, uracil and 5FU in F528M/**SYSE**, which was the only combination of **SYSE** with other mutations that possessed measurable transport of radiolabeled xanthine, thus enabling relative competition assays. As a control for the effect **SYSE** on substrate recognition we also estimated the  $K_i$  values of the single mutant F528M and wild-type UapA. [Figure 5\(B\)](#) shows that UapA-F528M/**SYSE** conserves high affinity (low  $\mu\text{M}$  range) for physiological substrates (xanthine and uric acid), similar to wild-type UapA, but additionally recognizes adenine, uracil and 5FU with moderate affinities (275, 102 271  $\mu\text{M}$ , respectively). Notice that F528M alone can lead to very low apparent accumulation of adenine or hypoxanthine (see [Figure 4\(A\)](#)), but in this case transport should operate via very low affinity, as suggested by the lack of competition of radiolabeled xanthine by excess of these purines (see [Figure 5\(B\)](#)). Thus, the presence of **SYSE**, at least in the context of F528M, contributes to better binding of several non-physiological substrates (adenine, uracil, 5FU), but does not make a significant distinction between uracil and 5FU. Thus, the simplest explanation for the observation that 10-fold excess uracil did not fully suppress 5FU toxicity in mutants possessing **SYSE** (seen in [Figure 5A](#), right panel), is that moderate affinity binding and low transport of 5FU is still sufficient to cause cytotoxicity.

We also combined the quadruple **SYSE** substitution (residues 405–408) with mutations N410I and N410V. The rationale for this was the following. We observed and verified via correlated sequence conservation algorithms (BIS2Analyzer, see Materials and methods) that in NAT proteins the residue analogous to position 410 of UapA is conserved as an Asn *only* when the residue at position 408 is occupied by a Gln (see [Figure 1 \(A\)](#)). In these cases, the NAT protein functions as a transporter specific for xanthine or uric acid and

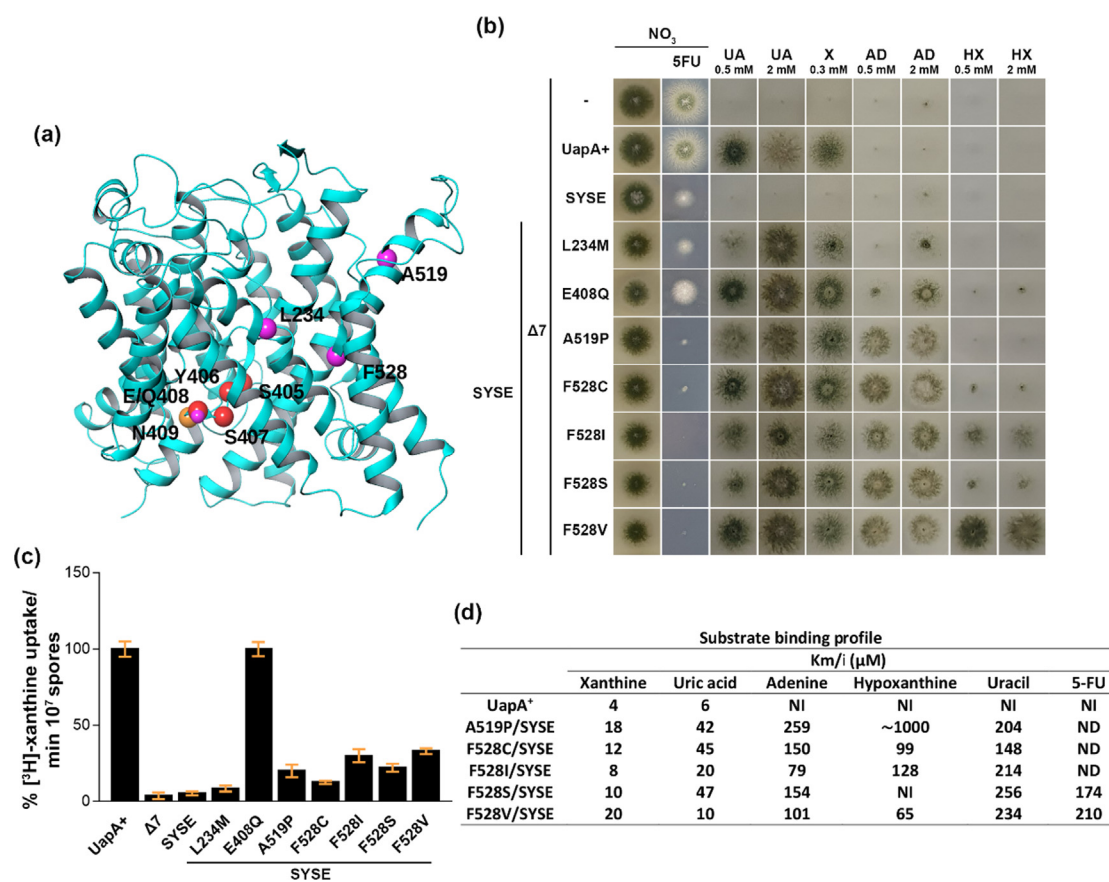
xanthine. In cases where the position equivalent to 410 in UapA is occupied by variable aliphatic amino acids, while positions 408 is a Glu, the proteins show reduced specificity towards uric acid/xanthine, being able to transport most purines and uracil. In other words, position 410 might have a ‘cryptic’ role in determining the specificity and in particular in contributing to acquisition of increased affinity for uric acid/xanthine binding in a context-dependent manner. The functional importance of N410 in UapA has not been tested before, while mutational studies in the bacterial XanQ homologue have suggested that residue N326 (equivalent to N410 in UapA) is not essential for xanthine transport.<sup>24,25,37</sup> Here we show that the single N410V mutation led to loss of apparent UapA transport activity of purines, while N410I had no effect on UapA activity and specificity, as based on growth assays with uric acid, adenine and hypoxanthine as sole N sources. However, either mutation, when combined with the **SYSE** sequence, led to loss of any apparent transport of nucleobases, including a significant reduction in 5FU toxic accumulation ([Supplementary Figure S1](#)). This shows that the presence of aliphatic residues at 410 is not compatible with SYSE in the UapA sequence context, contrasting what is observed in other NATs which naturally possess a SYSE sequence. Essentially “mirror-image” replacements at the position equivalent to 410 have been studied in bacterial xanthine/uric acid-transporting homologues that have a Q408-V410 combination. In those cases, replacement of the aliphatic residue Val with Asn leads to functional inactivation: the single-replacement mutants V312N in UacT<sup>37</sup> and V320N in SmXUacT<sup>38</sup> are present at wild-type levels in the membrane, but are functionally inactive. Thus, re-establishment of the combination Q408-N410 in the sequence context of those bacterial UapA-homologs is not compatible with (xanthine or uric acid) transport activity. This further highlights the complexity of epistatic interactions that control the fine functioning of the substrate binding site in NATs.

### Suppressor mutations conferring uric acid transport in UapA-SYSE support the functional importance of F528 and reveal new residues critical for specificity

To further understand the role of the NAT signature motif in specificity, we took advantage of the fact that UapA-SYSE has no capacity to transport purines to isolate revertants, after U.V. mutagenesis, that confer growth on uric acid as sole N source. [Table 2](#) and [Figure 6\(A\)](#) summarize the molecular identity and position of 33 revertants, which all proved to be within the *uapA* orf. Revertants concerned 7 amino acid changes in four residues: L234, E408, A519 and F528. All mutations, except F528I, were isolated

Table 2 Profile of SYSE suppressors

Mutation	TMS*	Domain	Codon change	Number of isolates
L234M	5	Dimerization	CTG → ATG	2
E408Q	10	Core	GAG → CAG	5
A519P	14a	Dimerization	GCG → CCG	3
F528S	14b	Dimerization	TTT → TCT	10
F528S	14b	Dimerization	TTT → TCA	1
F528C	14b	Dimerization	TTT → TGT	6
F528V	14b	Dimerization	TTT → GTT	5
F528I	14b	Dimerization	TTT → ATT	1

\* Based on.<sup>12</sup>

**Figure 6.** Characterization of mutations allowing uric acid transport in the UapA-SYSE context (A) Topological model showing the position of suppressor mutations relative to the NAT motif in UapA (B) Growth tests of suppressors. Growth conditions and control strains are as described in Figure 2. (C) Comparative <sup>3</sup>H-xanthine (0.1 μM) transport rates in suppressors. Details are as described in Figure 3A. Standard deviation is depicted with error bars. (D)  $K_{m/i}$  values (μM) determined using <sup>3</sup>H-xanthine uptake competition in suppressors. Details are as described in Figure 3B. ND indicates not determined. Results shown are the averages of triplicate measurements from 3 independent assays with a SD of <20%.

several times (2–10 times), showing that mutagenesis was rather saturated. Mutations in E408 introduced a Gln residue, thus reverting the SYSE sequence to SYSQ, which is a mutation expected to be isolated in this screen given the fact that UapA-SYSQ was earlier shown to confer growth on uric acid. Mutations in F528 were: F528C, F528I, F528S and F528V.

F528 has been previously shown to act as major specificity element as mutations introducing aliphatic (Ala, Met) or polar amino acids (Ser, Thr, Gln or Asn) converted UapA into a promiscuous purine-uracil transporter (Papageorgiou *et al.*<sup>17</sup> and Amillis *et al.*<sup>39</sup> but also herein). Notably, while F528 specificity mutations conferred very low affinity transport (>1 mM) for non-physiological purine

substrates, when combined with Q408E increased significantly the affinity of UapA for these 'new' substrates. Based on the crystal structure of UapA and Molecular Dynamics<sup>12,20</sup> (see also later) F528 is a residue located at the outward-facing gate of the substrate translocation trajectory, in the middle of TMS14 of UapA. Mutation A519P concerned also a residue in TMS14, but one that has never appeared in previous genetic screens for specificity mutants. Introduction of a Pro residue at this position might well modify the local structure to TMS14, and in particular the positioning of T526 and F528, both residues being critical for specificity.<sup>17</sup> Interestingly, the mutation concerning residue L234 (L234M in TMS5) has been isolated before in a screen searching for revertants of UapA mutations that led to loss-of-function because of inefficient dimerization.<sup>40</sup> It has then been proposed that L234M stabilizes the structure of monomeric UapA units, and thus lead to increased steady state levels of UapA. How such a mutation can also affect specificity opens new issues on the role of dimerization in finely regulating specificity, as already discussed in<sup>1,12</sup>.

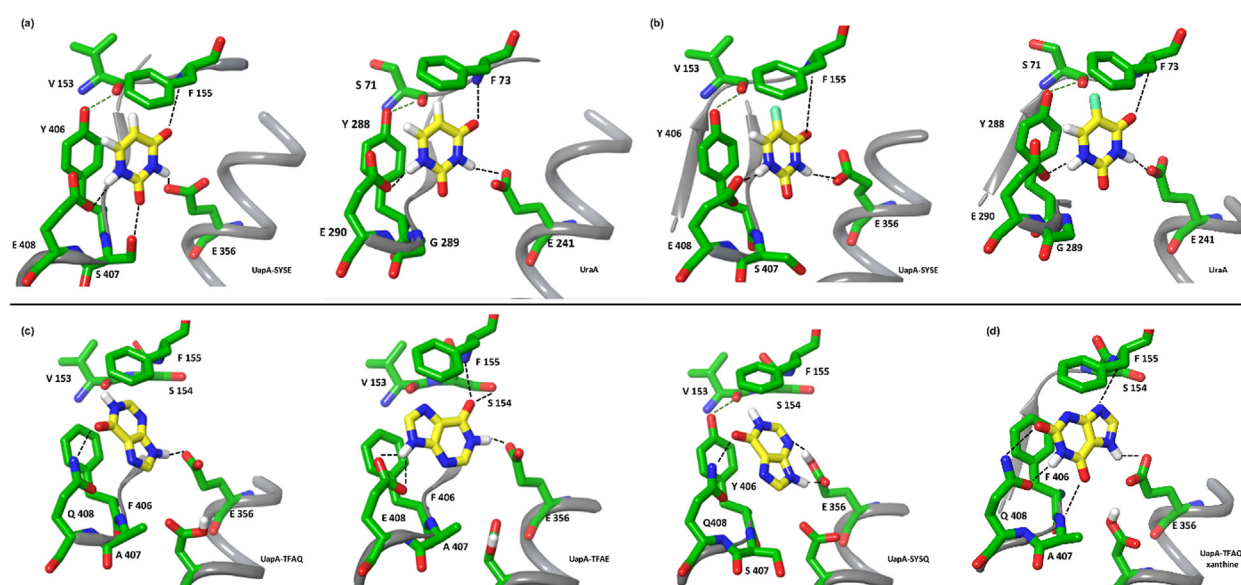
We analysed further the specificity profile of the revertants isolated by growth tests and uptake assays, as described earlier. Figure 6(B) shows that all revertants grow on uric acid (0.5 mM), as expected given their selection was carried out on this purine, but showed different levels of growth. E408Q, F528C and F528V were similar to the control strain expressing wild-type UapA. F528S, F528I, A519P and L234M showed relatively reduced growth, which however improved to nearly wild-type levels when uric acid was supplied at 4-fold higher concentration (2 mM). This suggests that the latter set of mutants might exhibit reduced uric acid binding affinity (see later). Most revertants grew equally well on xanthine, except from L234M, which seemed to be less efficient in xanthine accumulation. Most mutants also grew well on adenine, except from E408Q and L234M, which showed little growth on 0.5 mM adenine. Increased supply of adenine restored significant growth of E408Q, but only partially that of L234M. Most mutants, except F528V showed very little or moderate capacity for growth on hypoxanthine, both at standard and high concentrations. In fact, L234M and A519P mutants showed no evidence for hypoxanthine growth at all, scoring similarly to the original **SYSE** mutant. Finally, some revertants increased the apparent capacity for 5FU accumulation (F528S, F528V, F528C, F528I and A519P) or did not seem to affect significantly (L234M) the existing level of sensitivity of the original UapA-**SYSE** strain, with only E408Q leading to decreased capacity for 5FU accumulation.

We performed transport assays with radiolabelled xanthine in all revertants isolated. Figure 6(C) shows that E408Q has wild-type

transport rates justifying its full growth on uric acid or xanthine (Figure 6B). Most other mutants analyzed, except L234M, grew relatively well on uric acid or xanthine, despite reduced apparent transport rates of xanthine (~15–28% of wild-type levels), suggesting that this level of transport is apparently sufficient for growth (Notice also that in growth tests the concentration of substrates is 1000-fold higher than the concentration of radiolabeled substrates in transport assays; for details see<sup>26</sup>). Lastly, L234M that showed the lowest transport rate (8% of wild-type), also showed the lowest capacity to grow on xanthine and uric acid. We also estimated the  $K_i$  values for purines, uracil and 5FU in suppressors (Figure 6(D)). In general, all suppressors conserved relatively high or moderate affinity for xanthine or uric acid (8–47 M), albeit reduced compared to wild-type UapA. The moderate reduction in affinity for the physiological substrates is probably due, at least in the case of suppressors concerning F528 substitutions, to the presence of the **SYSE** sequence. This is concluded due to previously analyzed single F528 mutations that do not affect significantly the affinity of UapA for xanthine and uric acid,<sup>17,41</sup> but also when considering reduced growth on xanthine or uric acid of relative mutants including **SYSE**, as shown in Figure 4(A). Most suppressors also showed moderate affinities for other purines, uracil and 5FU (in the range of 65–259  $\mu$ M), with few exceptions of certain mutants towards specific substrates. Given that single mutations in F528 do not modify significantly the affinity for substrates,<sup>17,41</sup> the **SYSE** sequence should, in principle, be the reason of acquisition of moderate/low affinity binding of purines, uracil and 5FU. Thus, while the **SYSE** sequence when it stands alone in an otherwise wild-type UapA context leads to a transporter that is only able to accumulate low but toxic amounts of 5FU, when combined with suppressor mutations removing the aromatic Phe residue at position 528, it may lead to generation of UapA versions recognizing and transporting all purines, uracil and 5FU.

A similar, but not identical, promiscuous profile has been previously reported when specific mutations in F528, T526 or R481 are combined with mutation Q408E (i.e. sequence TFAE in the NAT motif; see<sup>17,41</sup>). Thus, the critical substitution that converts UapA from being highly specific for xanthine/uric acid to become a more general nucleobase transporter is mutation Q408E, a replacement topologically located in the substrate binding site and specifically concerning a residue known to interact strongly with substrates.<sup>12,13</sup> However, the potential effect of substitution Q408E to specificity is revealed only in the simultaneous presence of specific substitutions of other residues, distant from the *bona fidae* binding site, in this case amino acids R481, T526 or F528.





**Figure 7. UapA mutants-substrate interactions as resulted by Molecular Simulations.** (A) UapA-SYSE-uracil and UraA-uracil interactions. The UapA-SYSE-uracil interaction is stabilized by the formation of four H-bonds (depicted as dashed lines) between N1H and N3H with E408 and E356 carboxylates, respectively, C2=O with S407 hydroxyl group and C4=O with F155 backbone, as well as, a  $\pi$ - $\pi$  stacking formed between uracil with both F155 and Y406. Notice also that Y406 hydroxyl group forms a H-bond with V153 backbone. The UraA-uracil crystallographic structure is also presented on the right for comparison (23). (B) UapA-SYSE-5FU and UraA-5FU interactions. In both structures 5FU is oriented similarly to uracil with the F atom located in a hydrophobic cleft and C-F bond parallel to both F155 and Y406 rings. (C) Hypoxanthine interactions with UapA, UapA-TFAE and UapA-SYSQ. Hypoxanthine forms two H-bonds with the UapA binding cavity, one between N9H and E356 carboxylate and the second between C6=O and Q408 amide group. In UapA-TFAE, hypoxanthine N9H and N1H both form H-bonds with E408, and E356 carboxylates, respectively, while C6=O forms H-bond with S154 hydroxyl group and F155 backbone. In UapA-SYSQ-hypoxanthine interaction E356 is protonated. Hypoxanthine H-bond interactions with the binding cavity are between N9H, N3 and E356, as well as, between C6=O and Q408. (D) Wild-type UapA-xanthine crystal structure is included for comparison.

### Rationalizing specificity modifications via alteration in NAT signature motif

To better understand the specificity alterations of UapA, molecular simulations of protein-substrate interactions were performed (Figure 7). To this end, the models of UapA-SYSE, UapA-SYSQ, and UapA-TFAE have been constructed on the basis of the structure of UapA crystallized in the inward-open conformation. Important differences were observed in the modelled structures. These are related to the binding cavity of the transporter due to the presence of the F406Y mutation but also to the calculated pKa of acidic residues critical for substrate binding (e.g. the invariable E356 and D360 in TMS8 and E408 in TMS10 of UapA-SYSE or UapA-TFAE). More specifically, in UapA-SYSE the calculated pKa of E356 rises from 7.5 in the wild-type (UapA-TFAE) to 10.0, while that of D360 drops from 9.4 to 7.4. Similar pKa values to UapA-SYSE were calculated in UapA-SYSQ. Thus, in both UapA-SYSE and UapA-SYSQ E356 is considered protonated, while D360 is thought to be negatively charged (given

transport assays are performed at pH 6.8). On the other hand, E408 in both UapA-SYSE and UapA-TFAE exhibits a theoretical pKa of 6.8. Alterations in the acidic character in side chains of UapA-SYSQ and UapA-SYSE as compared to wild-type (UapA-TFAE) or UapA-TFAE are expected to modify the polar nature of the substrate binding side, and thus might affect substrate binding and transport. The other important observation related to the binding cavity of these mutants versus the wild-type UapA concerns the formation of a hydrogen bond when F406 is mutated to Tyr between the phenolic hydroxyl and the V153 backbone carbonyl (see Figure 7(A), (B)). This hydrogen bond stabilizes the main part of the binding cavity between TMS3 and TMS10, which is known to form the characteristic beta-sheet structure in the middle of the transmembrane domain where the substrate is embedded. This stabilization does not seem to affect transport activity (see Figure 3(A)), but apparently contributes to enlarged specificity as UapA-TFAE recognizes 5FU, hypoxanthine and adenine

significantly better than UapA-TFAQ (see Figs. 2 and 3(B), (C))

In a second step, we used Induced Fit Docking (IFD) calculations to try to understand the nature of increased apparent affinity of UapA-SYSE for uracil and 5FU. The most energetically favourable binding structure of uracil exhibits a remarkable similarity to the crystal structure of uracil bound by its natural transporter UraA,<sup>23</sup> as presented in the left and right panels of Figure 7(A). In UapA-SYSE, uracil is well stabilized between the two acidic side chains E408 and E356 forming also p-p stacking interactions with F155 and Y406 plus one more H-bond with S407 side chain. The latter is not present in UraA having a glycine at this position while UapA has an alanine. A quite similar positioning resulted by the IFD calculations also for 5FU, in both UapA-SYSE and UraA (Figure 7(B), left and right panels, respectively). Notice that the F atom is located in a hydrophobic cleft of the binding pocket where the C-F bond is parallel to both F155 and Y406 rings, without affecting proper orientation of the uracil ring. Importantly, in wild-type UapA, which does not seem to transport uracil or 5FU, these ligands are not similarly positioned in the binding pocket, when compared to the docking position in UraA or UapA-SYSE, as in their best fit position seem to interact via a mirror-image orientation (Supplementary Figure S2). Thus, it becomes evident that uracil/5FU docking in UraA or UapA-SYSE, both of which recognize and can transport these ligands, is distinct from that in wild type UapA, which cannot transport these pyrimidines. Our docking analysis cannot however rationalize why, despite similar docking positions, UraA recognizes uracil and 5FU with very high affinity (sub micromolar concentration), whereas UapA-SYSE recognizes these two substrates with moderate affinities (148–259  $\mu$ M, depending on the combination of the SYSE sequence with different F528 substitutions). A probable explanation for this difference, other than considering that binding affinities cannot be well predicted by modeling approaches, is that residues outside the binding site, crucial for recognition, might be distinct in UraA and UapA-SYSE.

In an effort to further rationalize the specificity of UapA mutants against hypoxanthine IFD calculations were performed considering wild type (UapA-TFAQ) and mutant versions UapA-SYSQ and UapA-TFAE. Notice again here that functional assays show that wild type (UapA-TFAQ) does not bind hypoxanthine, whereas UapA-TFAE binds with relatively high affinity (60–70  $\mu$ M) but does not transport hypoxanthine, and UapA-SYSQ recognizes hypoxanthine with low affinity (~1 mM). The lowest energy binding modes of hypoxanthine of all three forms are presented in Figure 7(C). It can be easily noticed that none of them is similar to the xanthine-UapA binding mode where Q408 forms two H-bonds with N1H and C2=O (Figure 7(D)). In the wild type UapA-

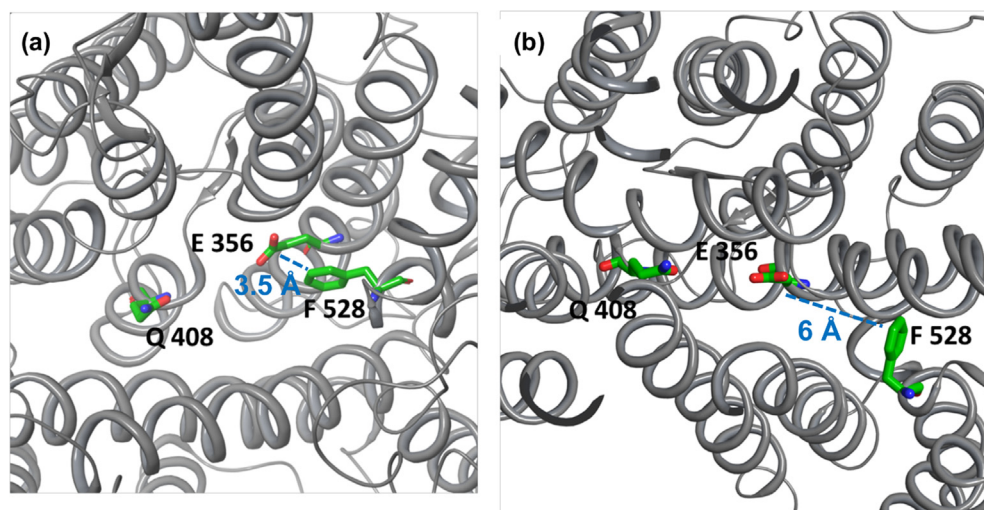
hypoxanthine structure N9H, but not N1H, interacts with E356, while Q408 forms a H-bond with C6=O forcing the hypoxanthine scaffold in a completely different orientation compared to the orientation of physiological substrates (e.g. xanthine). In UapA-TFAE, hypoxanthine appears to be oriented in such a way that the two positive charges on N9H and N1H<sup>42</sup> interact with the two negative charges of E408 and E356, respectively, while the C6=O group forms a H-bond with S154. This could explain the relatively high binding affinity of hypoxanthine in this UapA mutant. However, probably due to its topological disorientation within the binding site, this ligand is not transported. Finally, in UapA-SYSQ, where E356 is protonated, as deduced from pKa calculations, can now interact with N9H and N3 while C6=O interacts with Q408 amide group. This might explain the binding of hypoxanthine by UapA-SYSQ, even with low affinity, as compared to wild-type UapA that makes fewer interactions.

### F528 dynamically interacts with the substrate binding site

The structural explanation of the functional role of all different UapA residues studied herein is beyond the scope of this article and under investigation in our group. However, in order to rationalize the importance of F528 mutations isolated several times in this work, but also previously, as a key element affecting specificity alone and in combination with NAT signature motif substitutions, the model of the outward conformation of UapA has been constructed by homology modelling to the anion exchanger domain of human erythrocyte Band3 crystal structure.<sup>43</sup> Band3 is a human structural homologue of UapA that functions as bicarbonate ( $\text{HCO}_3^-$ ) transporter, which has been crystalized in the outward conformation and represents an excellent model, if not unique, for the outward conformation of transporters following the elevator mechanism. Figure 8 shows the position of F528 relatively to the position of E356 and Q408 in the modelled outward conformation and the inward crystal structure of UapA. As can be seen, in the outward topology the two residues are in very close distance interacting through p-electrons. The MD calculations shows that the distance of the two side chains remains stable ~3.50 Å through the 100 ns simulation, suggesting a quite stable interaction. In contrast, in the inward crystal structure this distance becomes ~6.00 Å. This suggests a dynamic interaction of F528 with the substrate binding site that could explain its critical role in specificity (see Discussion).

## Discussion

Despite the fact that UapA and bacterial NATs have been extensively analysed by a plethora of



**Figure 8.** UapA outward conformation. (A) The model constructed using Band3 as template is shown mainly focusing on the interaction between F528 and E356. The E356 carboxylate group remains almost parallel to the F528 phenyl ring (3.5 Å) during the 100 ns MD simulation. (B) The corresponding distance in the inward conformation is much higher (pdb structure 5i6c).

mutational studies we still do not know how specificity is determined by residues outside the binding site. The recent determination of the crystal structures of UapA and UraA, corresponding to inward-facing and occluded topologies, respectively, has led to formal identification of a major substrate binding site, but provided little hints on the issue of specificity. MDs using UapA, UraA and a handful of NAT structural homologues (e.g. anion exchanger 1 AE1/Band3 protein and SLC26Dg) caught in distinct conformations (e.g. outward-open) led to proposing a *sliding-elevator* type of transport mechanism<sup>44</sup> and defined a putative substrate translocation pathway in UapA.<sup>12</sup> In addition, based on MDs, the most prominent specificity mutations concerned residues R481, T526 or F528, which are located along the proposed sliding trajectory of the core domain in the UapA dimer. Still however, how these mutations affect specificity is poorly understood.

Here we decided to re-address how UapA specificity is determined by trying to better understand the contribution of partially conserved residues of the NAT signature motif, which is known to be part of the substrate binding site located in sliding elevator core domain in all NATs. Thus, we performed directed mutational and functional analyses of residues located at positions 405–408 of the NAT signature motif of UapA (i.e. TFAQ). The rationale of the mutations designed was to introduce residues that are present in other NATs showing distinct specificities. An important novelty of the present study was the use of a strain ( $\Delta 7$ ) that genetically lacks not only a wild-type UapA endogenous copy, but also all other major transporters related to

purine, pyrimidine, nucleoside or allantoin transport. The  $\Delta 7$  strain has practically no uptake transport activity for purine-related solutes and thus any UapA mutant introduced by genetic transformation in this strain could be analyzed for its ability to transport any purine-related substrate in a ‘clean’ background.

Based on the above system we revealed context-dependent cryptic roles of specific partially conserved residues located in the first part of the NAT signature motif. Most impressively, introducing the **SYSE** sequence in UapA leads to a transporter that lost its capacity to transport its physiological substrates or any other purine, but gained a capacity for toxic 5FU accumulation. In addition, combination of **SYSE** with F528 mutations led to a transporter able to translocate uracil or 5FU, in addition to purines. These findings are in line with the observation that functionally characterized bacterial NATs known to transport uracil and 5FU (e.g. UraA or RutG)<sup>22,45</sup> conserve the Tyr and Glu residues (e.g. possess the sequence **TYG/AE**), while the plant and metazoan NATs that recognize purines and uracil,<sup>6,10,11</sup> also conserve the Glu residue (e.g. **XXXE**). In other words, the presence of a Glu residue seems to dramatically increase uracil transport capability, while the Tyr residue also contributes to that. Our relative MDs have showed why this might be so. In UapA-**SYSE** uracil is well stabilized between E408 and E356 forming also pi-pi stacking interactions with F155 and Y406 and one more H-bond with S407 side chain. This docking position of uracil is not seen with the wild-type UapA. The important context-dependent role of E408 in UapA specificity is corroborated by comparing UapA-**SYSE** and UapA-**SYSQ**, the former being rather specific for



5FU accumulation and the latter promiscuous (transport of purines-uracil-5FU). Relative to these observations, our MD analysis also revealed possible interactions that rationalize why UapA-TFAE binds, but does not transport hypoxanthine, or why UapA-SYSE with UapA-SYSQ, have very distinct transport activities and specificities compared to wild-type UapA.

Among the UapA mutations functionally analyzed here two were specifically designed to introduce amino acids found in the NAT signature motif of mammalian L-ascorbate transporters, namely **SSSP** and **SSSQ** (see Figure 1(A)). The first one introduces the full set of 4 residues found in L-ascorbate transporters (SVCT1/2), and the second conserves the essential for purine binding Q408. Thus, we tested whether UapA could be converted to a transporter capable of L-ascorbate transport, or at least improve its capacity for binding L-ascorbate. We have recently shown that wild-type UapA might recognize with extremely low affinity ( $K_i \sim 20$  mM) L-ascorbic acid,<sup>9</sup> but we were unable to convincingly show any relative transport activity. Mutant UapA-**SSSP**, as probably expected due to replacement of residues critical for purine recognition (F406 and Q408), lost any detectable transport activity for xanthine and uric acid or purines in general. UapA-**SSSQ** also showed little, if any, purine transport activity, despite conserving Q408. This strongly suggests that the introduction of a series of polar Ser residues upstream of Q408 is detrimental for purine transport, at least within the context of an otherwise wild-type UapA sequence. Surprisingly, unlike wild-type UapA, both UapA-**SSSP** and UapA-**SSSQ** seem to be capable of accumulating low subtoxic levels of 5FU, judged from the reduction in conidiospore production when compared to wild-type UapA (see Figure 2(A)). However, we failed to obtain any evidence that UapA-**SSSP** and UapA-**SSSQ** recognize better or transport L-ascorbate. Given however the limitations of our system for detecting low-affinity L-ascorbate transport, we cannot exclude the possibility that UapA-**SSSP** and UapA-**SSSQ** might have an ameliorated capacity for recognizing this metabolite. It seems that engineering L-ascorbate transport capacity in UapA may not be feasible by simply introducing residues that mimic the presumed binding site of L-ascorbate transporter homologues (e.g. SVCT1/2).

The present work also provides novel findings concerning how F528, a residue outside the substrate binding site, might function as a key amino acid in determining UapA specificity. The great majority of suppressor mutations that restored purine (e.g. uric acid) transport in UapA-SYSE replace F528 with aliphatic or polar residues. These suppressors in general enlarge UapA specificity. In previous studies, similar F528 replacements also enlarged UapA specificity, so that it can transport, albeit with low affinity, all

purines and uracil. The only known substitution in this residue that conserves a wild-type functional profile is F528Y,<sup>41</sup> strongly suggesting that an aromatic ring at residue 528 is responsible for conferring specificity for uric acid or xanthine. How removing the aromaticity of residue 528 allows UapA to transport, in addition to its physiological substrates, all purines and uracil is not clear. However, our present MD analysis tried to shed some light into this question. By constructing the outward topology of UapA we were able to see that F528 approaches very close and probably interacts via pi-pi stacking with E356, which is a major substrate-binding residue. This interaction is apparently dynamic, as it is not present in the inward-facing UapA crystal. One might thus hypothesize that in the apo state (i.e. no substrate bound) the F528-E356 interaction 'locks' UapA in its outward topology. Upon substrate (and probably H<sup>+</sup>) loading and stabilization between E356 and Q408, the F528-E356 interaction is lost, and UapA is allowed to shift to its inward topology via sliding of the elevator domain. Replacing F528 by small aliphatic amino acids seems to genetically unlock the sliding mechanism in a way that becomes less dependent on substrate/H<sup>+</sup> binding, and thus convert UapA into a more promiscuous transporter for low-affinity substrates. Given that mutations other than those concerning F528, as shown previously,<sup>13</sup> but also here (e.g. L234 and A519), can similarly convert UapA into a more promiscuous transporter for weakly-binding substrates, these might also modify ('loosen') the sliding mechanism. A more rigorous understanding of how specificity is determined in UapA and other NATs, but also in elevator-type transporters in general,<sup>46</sup> will require more structural data and relative MDs involving several distinct topologies.

## Materials and Methods

### Media, strains and growth conditions

Standard complete (CM) and minimal media (MM) for *A. nidulans* growth were used. Media and supplemented auxotrophies were used at the concentrations given in <http://www.fgsc.net>.<sup>47</sup> Glucose 1 % (w/v) was used as carbon source. 10 mM sodium nitrate (NO<sub>3</sub>) was used as a standard nitrogen source. Nucleobases, nucleosides and toxic analogues were used at the following final concentrations: 5FU, 5FC or 5FUd at 10–100 μM, uric acid, adenine, hypoxanthine, guanine, cytosine, thymine, inosine, adenosine, guanosine, allantoin at 0.5–2.0 mM, and xanthine at 0.3 mM. L-ascorbic acid was added at 0.1–1% in the presence of sodium nitrate as N source. All media and chemical reagents were obtained from Sigma-Aldrich (Life Science Chemilab SA, Hellas) or Appli-Chem (Bioline Scientific SA, Hellas). A  $\Delta furD::riboB \Delta furA::riboB \Delta fcyB::argB \Delta azgA \Delta uapA \Delta uapC::AfpyrG \Delta cntA::riboB pabaA1 pantoB100$  mutant

strain, named  $\Delta 7$ , was the recipient strain in transformations with plasmids carrying *uapA* alleles based on complementation of the pantothenic acid auxotrophy *pantoB100*.<sup>26</sup> *pabaA1* is a paraminobenzoic acid auxotrophy. *A. nidulans* protoplast isolation and transformation was performed as previously described.<sup>48</sup> Growth tests were performed at 37 °C for 48 h, at pH 6.8. All strains used in this work are described in Table S1.

### Standard molecular biology manipulations and plasmid construction

Genomic DNA extraction from *A. nidulans* was performed as described in FGSC (<http://www.fgsc.net>). Plasmids, prepared in *E. coli*, and DNA restriction or PCR fragments were purified from agarose 1% gels with the Nucleospin Plasmid Kit or Nucleospin ExtractII kit, according to the manufacturer's instructions (Macherey- Nagel, Lab Supplies Scientific SA, Hellas). Standard PCR reactions were performed using KAPATaq DNA polymerase (Kapa Biosystems). PCR products used for cloning, sequencing and re-introduction by transformation in *A. nidulans* were amplified by a high fidelity KAPA HiFi HotStart Ready Mix (Kapa Biosystems) polymerase. DNA sequences were determined by VBC-Genomics (Vienna, Austria). Site directed mutagenesis was carried out according to the instructions accompanying the Quik-Change® Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene). The principal vector used for most *A. nidulans* mutants is a modified pGEM-T-easy vector carrying a version of the *gpdA* promoter, the *trpC* 3' termination region and the panB selection marker.<sup>31</sup> Mutations were constructed by oligonucleotide-directed mutagenesis or appropriate forward and reverse primers (Table S2). Transformants arising from single copy integration events with intact *uapA* ORFs were identified by Southern and PCR analysis

### Correlated sequence conservation algorithm

The BIS2Analyzer (<http://www.lcqb.upmc.fr/BIS2Analyzer/>) was used for detected residues co-conserved with specific residues of the NAT motif. It is an openly accessible server providing online analysis of co-evolving amino-acid pairs in protein alignments, especially designed protein families which typically display a small number of highly similar sequences.<sup>49</sup>

### Uptake assays

Kinetic analysis of wild-type and mutant UapA was measured by estimating uptake rates of [<sup>3</sup>H]-xanthine uptake (40 Ci mmol<sup>-1</sup>, Moravek Biochemicals, CA, USA), as previously described in.<sup>26</sup> In brief, [<sup>3</sup>H]-xanthine uptake was assayed in *A. nidulans* conidiospores germinating for 4 h at 37 °C, at 140 rpm, in liquid MM, pH 6.8. Initial veloc-

ities were measured on 10<sup>7</sup> conidiospores/100  $\mu$ L by incubation with concentrations of 0.2–2.0  $\mu$ M of [<sup>3</sup>H]-xanthine at 37 °C. For the competition experiments, initial uptake rates of [<sup>3</sup>H]-xanthine were measured in the simultaneous presence increasing concentrations (5  $\mu$ M – 1 mM) of various putative nucleobase-related inhibitors or L-ascorbic acid at 1%, as indicated. [<sup>3</sup>H]-uracil uptake was also measured similarly. The time of incubation (1 or 2 min) was defined through time–course experiments. All transport assays were carried out at least in two independent experiments and the measurements in triplicate. Standard deviation was <20%. Results were analyzed in GraphPad Prism software.

### Isolation and characterization of suppressor mutations

Suppressor mutations of 10<sup>9</sup> conidiospores of the strain UapA-SYSE were obtained after 3 min 45 sec exposure at a standard distance of 20 cm from an Osram HNS30 UV-B/C lamp and subsequent selection of colonies capable of growing on MM containing uric acid as sole nitrogen source, at 37 °C. Spores from positive colonies were collected after 4–8 days and further isolated on the same selective medium that was used to obtain the original colonies. Genomic DNA from 34 purified colonies was isolated and the ORF of UapA was amplified and sequenced. In all cases the amplified fragments contained a new mutation.

### Epifluorescence microscopy

Samples for standard epifluorescence microscopy were prepared as previously described.<sup>50,51</sup> In brief, sterile 35 mm l-dishes with a glass bottom (Ibidi, Germany) containing liquid minimal media supplemented with NaNO<sub>3</sub> and 1% glucose were inoculated from a spore solution and incubated for 16 h at 25 °C. The samples were observed on an Axioplan Zeiss phase contrast epifluorescent microscope and the resulting images were acquired with a Zeiss-MRC5 digital camera using the AxioVs40 V4.40.0 software. Image processing and contrast adjustment were made using the ZEN 2012 software while further processing of the TIFF files was made using Adobe Photoshop CS3 software for brightness adjustment, rotation and alignment.

### Molecular simulations

**Protein Model Construction.** Models of UapA-TFAE, UapA-SYSE and UapA-SYSQ were generated by mutating the specific residues on Maestro platform (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018), on the basis of the structure of UapA crystallized in the inward-open conformation (PDB ID: 5i6c). In order to construct the model of UapA outward conformation Band3 transporter (4yzf) was used

as template.<sup>43</sup> Band3 is an anion exchanger of 14 transmembrane helices, crystallized in the outward conformation. The model was constructed using Prime Homology in Prime 2018-4 (Schrödinger, LLC, New York, NY, 2018), by superimposing helices one by one and aligning all the important residues.

**Protein Preparation.** The protein was prepared using the Protein Preparation Protocol implemented in Schrödinger suite (Schrödinger Suite 2018, Protein Preparation Wizard) and accessible within the Maestro program (Maestro, version 2018-4, Schrödinger, LLC, New York, NY, 2018). Specifically, at first hydrogen atoms were added followed by an optimization of the orientation of hydroxyl groups of Asn, Gln, and of the protonation state of His in order to maximize hydrogen bonding. The final step was that of minimization of the protein, using the OLS3 force field.<sup>52</sup>

**Ligand preparation** Ligand preparation for docking was performed with LigPrep application (LigPrep, version 2018-4, Schrödinger, LLC, New York, NY, 2018), which consists of a series of steps that perform conversions, apply corrections to the structure, generate ionization states and tautomers, and optimize the geometries. The force field chosen was OPLS3.<sup>52</sup>

**Induced Fit Docking** Schrödinger Suite protocol was used (Schrödinger Suite 2018-4 Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2018), taking into account the side chain or backbone movements, or both, upon ligand binding, thus circumventing an inflexible binding site. In the first softened-potential docking step, of the protocol, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018-4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues within 5 Å of any out of the 20 ligand poses. The Prime refinement starts by performing conformational search and by minimizing the side chains of the selected residues. After convergence to a low-energy result, an additional minimization of all selected residues (side chain and backbone) is performed with the Truncated-Newton algorithm using the OPLS3 parameter set<sup>52</sup> and a surface Generalized Born implicit solvent model. The ranking of the obtained complexes is implemented according to Prime calculated energy (molecular mechanics and solvation), and the complexes within 30 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. Finally, the ligands used in the first docking step are re-docked into each of the receptor structures retained from the Prime refinement step. The final ranking of the complexes is performed by a com-

posite score which takes into account the receptor–ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

**Molecular Dynamics (MD).** For the construction of the protein–ligand complex CHARMM-GUI platform was used. Each model was inserted into a heterogeneous fully hydrated bilayer 120 Å × 120 Å × 120 Å, consisting of DPPC lipids and ergosterol. The membrane embedded system was solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. CHARMM36m<sup>53</sup> force field was used for protein and lipids, while the ligand was prepared using Antechamber<sup>54</sup> and the general Amber force field.<sup>55</sup> The protein orientation into the membrane was calculated using the PPM server.<sup>56</sup> The assembled simulation system consisted of ~130,000 atoms. The systems were simulated using GROMACS software.<sup>57</sup> The models were minimized and equilibrated to obtain stable structures. Minimization was carried out for 2,000 steps with a step size of 0.001 kJ/mol applying a steepest descent followed by a conjugate gradient algorithm, and the system was equilibrated for 20 ns by gradually heating and releasing the restraints to expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 300 K for 100 ns, using Nose-Hoover thermostat<sup>58</sup> and Parrinello-Rahman semi-isotropic pressure coupling<sup>59</sup> and compressibility at 4.5e-5 bar-1. The van der Waals and electrostatic interactions were smoothly switched off at 1.2 nm by switching function, while long-range electrostatic interactions were calculated using the particle mesh Ewald method.<sup>60</sup>

## CRedit authorship contribution statement

**Anezia Kourkoulou:** Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing - original draft, Writing - review & editing. **Iliana Zantza:** Data curation, Formal analysis, Investigation, Methodology. **Konstantina Foti:** Investigation. **Emmanuel Mikros:** Formal analysis, Funding acquisition, Resources, Supervision, Validation, Software, Writing - original draft, Writing - review & editing. **George Diallinas:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

## DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## Acknowledgments

This work was supported by a “Stavros S. Niarchos Foundation” grant and by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility -ARIS- under project NCS1\_Mechanism (pr006040).

## Author contributions

A.K. performed genetic, molecular, biochemical cell biology experiments, analyzed and discussed results. K.F. assisted in genetic experiments. I.Z. and E.M. performed the *in silico* analysis, the Molecular Dynamics, and analyzed *in silico* results. E.M. wrote the *in silico* part of the article. G.D. conceived and planned experiments analyzed results and wrote the article.

## Appendix A. Supplementary Data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2021.166814>.

Received 26 October 2020;

Accepted 4 January 2021;

Available online xxxx

### Keywords:

*Aspergillus nidulans*;

Fungi;

UapA;

Genetics;

Nucleobase

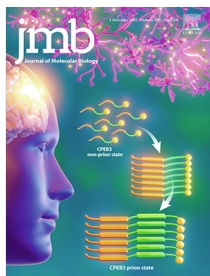
## References

- Algüel, Y., Cameron, A.D., Diallinas, G., Byrne, B., (2016). Transporter oligomerization: form and function. *Biochem. Soc. Trans.*, **44**, 1737–1744.
- Diallinas, G., Martzoukou, O., (2019). Transporter membrane traffic and function: lessons from a mould. *FEBS J.*, **286**, 4861–4875.
- Diallinas, G., Gourmas, C., (2008). Structure-function relationships in the nucleobase-ascorbate transporter (NAT) family: lessons from model microbial genetic systems. *Channels (Austin)*, **2** (5), 363–372.
- Gourmas, C., Papageorgiou, I., Diallinas, G., (2008). The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Mol. BioSyst.*, **4**, 404–416.
- Frillingos, S., (2012). Insights to the evolution of Nucleobase-Ascorbate Transporters (NAT/NCS2 family) from the Cys-scanning analysis of xanthine permease XanQ. *Int. J. Biochem. Mol. Biol.*, **3**, 250–272.
- Niopek-Witz, S., Deppe, J., Lemieux, M.J., Möhlmann, T., (2014). Biochemical characterization and structure-function relationship of two plant NCS2 proteins, the nucleobase transporters NAT3 and NAT12 from *Arabidopsis thaliana*. *BBA*, **1838**, 3025–3035.
- Yamamoto, S., Inoue, K., Murata, T., Kamigaso, S., Yasujima, T., Maeda, J.Y., Yoshida, Y., Ohta, K.Y., Yuasa, H., (2010). Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *J. Biol. Chem.*, **285**, 6522–6531.
- Bürzle, M., Suzuki, Y., Ackermann, D., Miyazaki, H., Maeda, N., Clémenton, B., Burrier, R., Hediger, M.A., (2013). The sodium-dependent ascorbic acid transporter family SLC23. *Mol. Aspects Med.*, **34**, 436–454.
- Kourkoulou, A., Pittis, A.A., Diallinas, G., (2018). Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family. *Microb. Cell.*, **5**, 280–292.
- Maurino, V.G., Grube, E., Zielinski, J., Schild, A., Fischer, K., Flüge, U.I., (2006). Identification and expression analysis of twelve members of the nucleobase-ascorbate transporter (NAT) gene family in *Arabidopsis thaliana*. *Plant Cell Physiol.*, **47**, 1381–1393.
- Chai, W., Peng, X., Liu, B., Wang, J., Zhu, Z., Liu, Y., Zhao, K., Cheng, B., Si, W., Jiang, H., (2018). Comparative Genomics, Whole-Genome Re-sequencing and Expression Profile Analysis of Nucleobase: Cation Symporter 2 (NCS2) Genes in Maize. *Front. Plant Sci.*, **9**, 856.
- Algüel, Y., Amillis, S., Leung, J., Lambrinidis, G., Capaldi, S., Scull, N.J., Craven, G., Iwata, S., Armstrong, A., Mikros, E., Diallinas, G., Cameron, A.D., Byrne, B., (2016). Structure of eukaryotic purine/H(+) symporter UapA suggests a role for homodimerization in transport activity. *Nature Commun.*, **7**, 11336.
- Diallinas, G., (2016). Dissection of Transporter Function: From Genetics to Structure. *Trends Genet.*, **32**, 576–590.
- Diallinas, G., Scazzocchio, C., (1989). A gene coding for the uric acid-xanthine permease of *Aspergillus nidulans*: inactivational cloning, characterization, and sequence of a cis-acting mutation. *Genetics*, **122**, 341–350.
- Gorfinkiel, L., Diallinas, G., Scazzocchio, C., (1993). Sequence and regulation of the uapA gene encoding a uric acid-xanthine permease in the fungus *Aspergillus nidulans*. *J. Biol. Chem.*, **268**, 23376–23381.
- Goudela, S., Karatza, P., Koukaki, M., Frillingos, S., Diallinas, G., (2005). Comparative substrate recognition by bacterial and fungal purine transporters of the NAT/NCS2 family. *Mol. Membr. Biol.*, **22**, 263–275.
- Papageorgiou, I., Gourmas, C., Vlant, A., Amillis, S., Pantazopoulou, A., Diallinas, G., (2008). Specific interdomain synergy in the UapA transporter determines its unique specificity for uric acid among NAT carriers. *J. Mol. Biol.*, **382**, 1121–1135.
- Kosti, V., Papageorgiou, I., Diallinas, G., (2010). Dynamic elements at both cytoplasmically and extracellularly facing sides of the UapA transporter selectively control the accessibility of substrates to their translocation pathway. *J. Mol. Biol.*, **397**, 1132–1143.
- Koukaki, M., Vlant, A., Goudela, S., Pantazopoulou, A., Gioule, H., Tournaviti, S., Diallinas, G., (2005). The

- nucleobase-ascorbate transporter (NAT) signature motif in UapA defines the function of the purine translocation pathway. *J. Mol. Biol.*, **350**, 499–513.
20. Kosti, V., Lambrinidis, G., Myrianthopoulos, V., Diallynas, G., Mikros, E., (2012). Identification of the substrate recognition and transport pathway in a eukaryotic member of the nucleobase-ascorbate transporter (NAT) family. *PLoS ONE*, **7**, e41939
  21. Meintanis, C., Karagouni, A.D., Diallynas, G., (2000). Amino acid residues N450 and Q449 are critical for the uptake capacity and specificity of UapA, a prototype of a nucleobase-ascorbate transporter family. *Mol. Membr. Biol.*, **2000** (17), 47–57.
  22. Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J., Yan, N., (2011). Structure and mechanism of the uracil transporter UraA. *Nature*, **472**, 243–246.
  23. Yu, X., Yang, G., Yan, C., Baylon, J.L., Jiang, J., Fan, H., Lu, G., Hasegawa, K., Okumura, H., Wang, T., Tajkhorshid, E., Li, S., Yan, N., (2017). Dimeric structure of the uracil:proton symporter UraA provides mechanistic insights into the SLC4/23/26 transporters. *Cell Res.*, **27**, 1020–1033.
  24. Karatza, P., Panos, P., Georgopoulou, E., Frillingos, S., (2006). Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential, and Ile-329-Val-339 form an alpha-helix. *J. Biol. Chem.*, **281**, 39881–39890.
  25. Georgopoulou, E., Mermelekas, G., Karna, E., Frillingos, S., (2010). Purine substrate recognition by the nucleobase-ascorbate transporter signature motif in the YgfO xanthine permease: ASN-325 binds and ALA-323 senses substrate. *J. Biol. Chem.*, **285**, 19422–19433.
  26. Kryptou, E., Diallynas, G., (2014). Transport assays in filamentous fungi: kinetic characterization of the UapC purine transporter of *Aspergillus nidulans*. *Fungal Genet. Biol.*, **63**, 1–8.
  27. Kryptou, E., Lambrinidis, G., Evangelidis, T., Mikros, E., Diallynas, G., (2014). Modelling, substrate docking and mutational analysis identify residues essential for function and specificity of the major fungal purine transporter AzgA. *Mol. Microbiol.*, **93**, 129–145.
  28. Kryptou, E., Kosti, V., Amillis, S., Myrianthopoulos, V., Mikros, E., Diallynas, G., (2012). Modeling, substrate docking, and mutational analysis identify residues essential for the function and specificity of a eukaryotic purine-cytosine NCS1 transporter. *J. Biol. Chem.*, **287**, 36792–36803.
  29. Sioupouli, G., Lambrinidis, G., Mikros, E., Amillis, S., Diallynas, G., (2017). Cryptic purine transporters in *Aspergillus nidulans* reveal the role of specific residues in the evolution of specificity in the NCS1 family. *Mol. Microbiol.*, **103**, 319–332.
  30. Amillis, S., Hamari, Z., Roumelioti, K., Scazzocchio, C., Diallynas, G., (2007). Regulation of expression and kinetic modeling of substrate interactions of a uracil transporter in *Aspergillus nidulans*. *Mol. Membr. Biol.*, **24**, 206–214.
  31. Kryptou, E., Evangelidis, T., Bobonis, J., Pittis, A.A., Gabaldón, T., Scazzocchio, C., Mikros, E., Diallynas, G., (2015). Origin, diversification and substrate specificity in the family of NCS1/FUR transporters. *Mol. Microbiol.*, **96**, 927–950.
  32. Hamari, Z., Amillis, S., Drevet, C., Apostolaki, A., Vágvölgyi, C., Diallynas, G., Scazzocchio, C., (2009). Convergent evolution and orphan genes in the Fur4p-like family and characterization of a general nucleoside transporter in *Aspergillus nidulans*. *Mol. Microbiol.*, **73**, 43–57.
  33. Papadaki, G.F., Amillis, S., Diallynas, G., (2017). Substrate Specificity of the FurE Transporter Is Determined by Cytoplasmic Terminal Domain Interactions. *Genetics*, **207**, 1387–1400.
  34. Smirnoff, N., (2018). Ascorbic acid metabolism and functions: A comparison of plants and mammals. *Free Radic. Biol. Med.*, **122**, 116–129.
  35. Sahin-Tóth, M., Akhoun, K.M., Runner, J., Kaback, H.R., (2000). Ligand recognition by the lactose permease of *Escherichia coli*: Specificity and affinity are defined by distinct structural elements of galactopyranosides. *Biochemistry*, **39**, 5097–5103.
  36. Sahin-Tóth, M., Gunawan, P., Lawrence, M.C., Toyokuni, T., Kaback, H.R., (2002). Binding of hydrophobic D-galactopyranosides to the lactose permease of *Escherichia coli*. *Biochemistry*, **41**, 13039–13045.
  37. Papakostas, K., Frillingos, S., (2012). Substrate selectivity of YgfU, a uric acid transporter from *Escherichia coli*. *J. Biol. Chem.*, **287**, 15684–15695.
  38. Botou, M., Yalelis, V., Lazou, P., Zantza, I., Papakostas, K., Charalambous, V., Mikros, E., Fliemetakis, E., Frillingos, S., (2020). Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*. *Mol. Microbiol.*, **114**, 151–171.
  39. Amillis, S., Koukaki, M., Diallynas, G., (2001). Substitution F569S converts UapA, a specific uric acid-xanthine transporter, into a broad specificity transporter for purine-related solutes. *J. Mol. Biol.*, **313**, 765–774.
  40. Kourkoulou, A., Grevias, P., Lambrinidis, G., Pyle, E., Dionysopoulou, M., Politis, A., Mikros, E., Byrne, B., Diallynas, G., (2019). Specific Residues in a Purine Transporter Are Critical for Dimerization, ER Exit, and Function. *Genetics*, **213**, 1357–1372.
  41. Vlanti, A., Amillis, S., Koukaki, M., Diallynas, G., (2006). A novel-type substrate-selectivity filter and ER-exit determinants in the UapA purine transporter. *J. Mol. Biol.*, **357**, 808–819.
  42. Emel'yanenko, V.N., Zaitsau, D.H., Verevkin, S.P., (2017). Thermochemical Properties of Xanthine and Hypoxanthine Revisited. *J. Chem. Eng. Data*, **62**, 2606–2609.
  43. Arakawa, T., Kobayashi-Yurugi, T., Alguel, Y., Iwanari, H., Hatae, H., Iwata, M., Abe, Y., Hino, T., Ikeda-Suno, C., Kuma, H., Kang, D., Murata, T., Hamakubo, T., Cameron, A.D., Kobayashi, T., Hamasaki, N., Iwata, S., (2015). Crystal structure of the anion exchanger domain of human erythrocyte band 3. *Science*, **350**, 680–684.
  44. Boudker, O., Ryan, R.M., Yernool, D., Shimamoto, K., Gouaux, E., (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature*, **445**, 387–393.
  45. Botou, M., Lazou, P., Papakostas, K., Lambrinidis, G., Evangelidis, T., Mikros, E., Frillingos, S., (2018). Insight on specificity of uracil permeases of the NAT/NCS2 family from analysis of the transporter encoded in the pyrimidine utilization operon of *Escherichia coli*. *Mol. Microbiol.*, **2**, 204–219.

46. Sauer, D.B., Trebesch, N., Marden, J.J., Cocco, N., Song, J., Koide, A., Koide, S., Tajkhorshid, E., Wang, D.N., (2020). Structural basis for the reaction cycle of DASS dicarboxylate transporters. *Elife.*, **9**, e61350
47. McCluskey, K., Wiest, A., Plamann, M., (2010). The Fungal Genetics Stock Center: repository for 50 years of fungal genetics research. *J. Biosci.*, **35**, 119–126.
48. Koukaki, M., Giannoutsou, E., Karagouni, A., Diallinas, G., (2003). A novel improved method for *Aspergillus nidulans* transformation. *J. Microbiol. Methods*, **55**, 687–695.
49. Oteri, F., Nadalin, F., Champeimont, R., Carbone, A., (2017). BIS2Analyzer: a server for co-evolution analysis of conserved protein families. *Nucleic Acids Res.*, **W1**, W307–W314.
50. Gournas, C., Amillis, S., Vlanti, A., Diallinas, G., (2010). Transport-dependent endocytosis and turnover of a uric acid-xanthine permease. *Mol. Microbiol.*, **75**, 246–260.
51. Karachaliou, M., Amillis, S., Evangelinos, M., Kokotos, A. C., Yalelis, V., Diallinas, G., (2013). The arrestin-like protein ArtA is essential for ubiquitination and endocytosis of the UapA transporter in response to both broad-range and specific signals. *Mol. Microbiol.*, **88**, 301–317.
52. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., Dahlgren, M.K., Knight, J.L., Kaus, J.W., Cerutti, D.S., Krilov, G., Jorgensen, W.L., Abel, R., Friesner, R.A., (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.*, **12**, 281–296.
53. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., MacKerell Jr., A.D., (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, **14**, 71–73.
54. Wang, J., Wang, W., Kollman, P.A., Case, D.A., (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.*, **25**, 247–260.
55. Wang, J., Wol, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A., (2004). Development and testing of a general amber force field. *J. Comput. Chem.*, **25**, 1157–1174.
56. Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, **40**, D370–D376.
57. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindah, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.
58. Evans, D.J., Holian, B.L., (1985). The Nose-Hoover thermostat. *J. Chem. Phys.*, **83**, 4069–4074.
59. Parrinello, M., Rahman, A., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
60. Darden, T., York, D., Pedersen, L., (1993). Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.





# Identification of New Specificity Determinants in Bacterial Purine Nucleobase Transporters based on an Ancestral Sequence Reconstruction Approach

Ekaterini Tatsaki<sup>1</sup>, Eleni Anagnostopoulou<sup>1</sup>, Iliana Zantza<sup>2</sup>, Panayiota Lazou<sup>1</sup>, Emmanuel Mikros<sup>2</sup> and Stathis Frillingos<sup>1,3\*</sup>

**1 - Laboratory of Biological Chemistry, Department of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece**

**2 - Division of Pharmaceutical Chemistry, Department of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, Athens, Greece**

**3 - Institute of Biosciences, University Research Center of Ioannina, Ioannina, Greece**

**Correspondence to Stathis Frillingos:** Laboratory of Biological Chemistry, Department of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece. [efrilligo@uoi.gr](mailto:efrilligo@uoi.gr) (S. Frillingos)

[@EleniAnagn](https://twitter.com/EleniAnagn) (E. Anagnostopoulou)

<https://doi.org/10.1016/j.jmb.2021.167329>

Edited by Daniel L. Minor

## Abstract

The relation of sequence with specificity in membrane transporters is challenging to explore. Most relevant studies until now rely on comparisons of present-day homologs. In this work, we study a set of closely related transporters by employing an evolutionary, ancestral-reconstruction approach and reveal unexpected new specificity determinants. We analyze a monophyletic group represented by the xanthine-specific XanQ of *Escherichia coli* in the Nucleobase-Ascorbate Transporter/Nucleobase-Cation Symporter-2 (NAT/NCS2) family. We reconstructed AncXanQ, the putative common ancestor of this clade, expressed it in *E. coli* K-12, and found that, in contrast to XanQ, it encodes a high-affinity permease for both xanthine and guanine, which also recognizes adenine, hypoxanthine, and a range of analogs. AncXanQ conserves all binding-site residues of XanQ and differs substantially in only five intramembrane residues outside the binding site. We subjected both homologs to rationally designed mutagenesis and present evidence that these five residues are linked with the specificity change. In particular, we reveal Ser377 of XanQ (Gly in AncXanQ) as a major determinant. Replacement of this Ser with Gly enlarges the specificity of XanQ towards an AncXanQ-phenotype. The ortholog from *Neisseria meningitidis* retaining Gly at this position is also a xanthine/guanine transporter with extended substrate profile like AncXanQ. Molecular Dynamics shows that the S377G replacement tilts transmembrane helix 12 resulting in rearrangement of Phe376 relative to Phe94 in the XanQ binding pocket. This effect may rationalize the enlarged specificity. On the other hand, the specificity effect of S377G can be masked by G27S or other mutations through epistatic interactions.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Nucleobases are essential to cellular organisms, as they provide both the coding building blocks of DNA and/or RNA and core molecular moieties of

key metabolites, energy carriers, coenzymes and signaling molecules. Transmembrane nucleobase/nucleoside transporters are needed to import nucleobases from exogenous sources when *de novo* synthesis and recycling from nucleic acids

are not sufficient. In addition, the imported nucleobases can be used catabolically as N sources and, less commonly,<sup>28,73</sup> C sources by various microorganisms when other sources are scarce or missing.

The most widespread nucleobase transporter family in microorganisms is the NAT/NCS2 (Nucleobase-Ascorbate Transporter/Nucleobase-Cation Symporter-2) family. This family is present in all domains of life and in almost all major phyla of organisms,<sup>21,10,39</sup> but does not include any nucleobase-transporting homolog in primates where the group is represented by two L-ascorbate-transporting homologs (SVCT1, SVCT2), a functionally unknown member (SLC23A3) and a nucleobase-transporter related pseudogene (SNBT1/Slc23a4).<sup>63,9</sup> In terms of functional characterization, few homologs are known in detail. An interesting specificity-related feature is that most known bacterial homologs display distinct and highly specific substrate profiles,<sup>21,51,7,8</sup> whereas all known nucleobase-related eukaryotic homologs are of broader specificity.<sup>25,41,42,47,48,68</sup>

The structural fold of the family is based on the structurally solved uracil permease UraA from *Escherichia coli* (in inward-open and occluded conformations)<sup>46,67</sup> and xanthine/uric acid permease UapA from *Aspergillus nidulans* (in inward-open conformation),<sup>2</sup> which also share structural homology with related transporters of at least two additional families (AE/SLC4, SulP/SLC26)<sup>4,23,62,13</sup> and appear to function as homodimers.

The mechanism proposed for these transporters, as well as several other transporters from diverse other families,<sup>45,22,55</sup> is the so-called elevator-type mechanism, in which a core (or transport, or elevator) domain carrying the binding site slides along a relatively immobile gate (or scaffold, or oligomerization) domain with an elevator-like motion alternatively exposing the binding site to either side of the membrane. It has been proposed that, apart from changes in the binding pocket *per se*, certain side chain alterations at the interface between the gate and the core domains may affect specificity by loosening constraints for the access of non-canonical substrates to the binding site<sup>17</sup> and/or uncoupling the elevator from high-affinity binding of substrates.<sup>19</sup>

With regard to understanding the diversification in substrate profiles and the molecular underpinnings of the different NAT/NCS2 specificities, our progress is rather limited to date and is based only on the rigorous mutational analysis of two xanthine and/or uric acid-transporting homologs, the fungal UapA,<sup>18,40</sup> and the bacterial XanQ,<sup>21,37</sup> and comparative analysis of these homologs with other related present-day transporters.

In this report, we explore the specificity profiles of bacterial NAT/NCS2 transporters by using a different, evolutionary strategy, based on ancestral

sequence reconstruction. Since the introduction of this method,<sup>11,61,60</sup> ancestral sequence reconstruction has been used for analyzing diverse groups of proteins and yielded important conclusions on the molecular basis of the emergence of different functions and specificities in extant proteins. Such studies have shown that new functions often derive from few sequence changes and rather simple mechanisms which, however, remain cryptic in present-day homologs due to confounding epistatic effects from other mutations.<sup>27,31,65</sup> Application of this experimental strategy to membrane solute transporters is limited to date,<sup>16,56</sup> due in part to lack of structural and functional knowledge on an adequate number of homologs in most cases. The expansion of available atomic-resolution structures of membrane transport proteins through current advances in crystallography and single-particle electron cryo-microscopy,<sup>64,57,66</sup> will certainly offer a basis for more systematic application of such evolutionary strategies to membrane transporters.

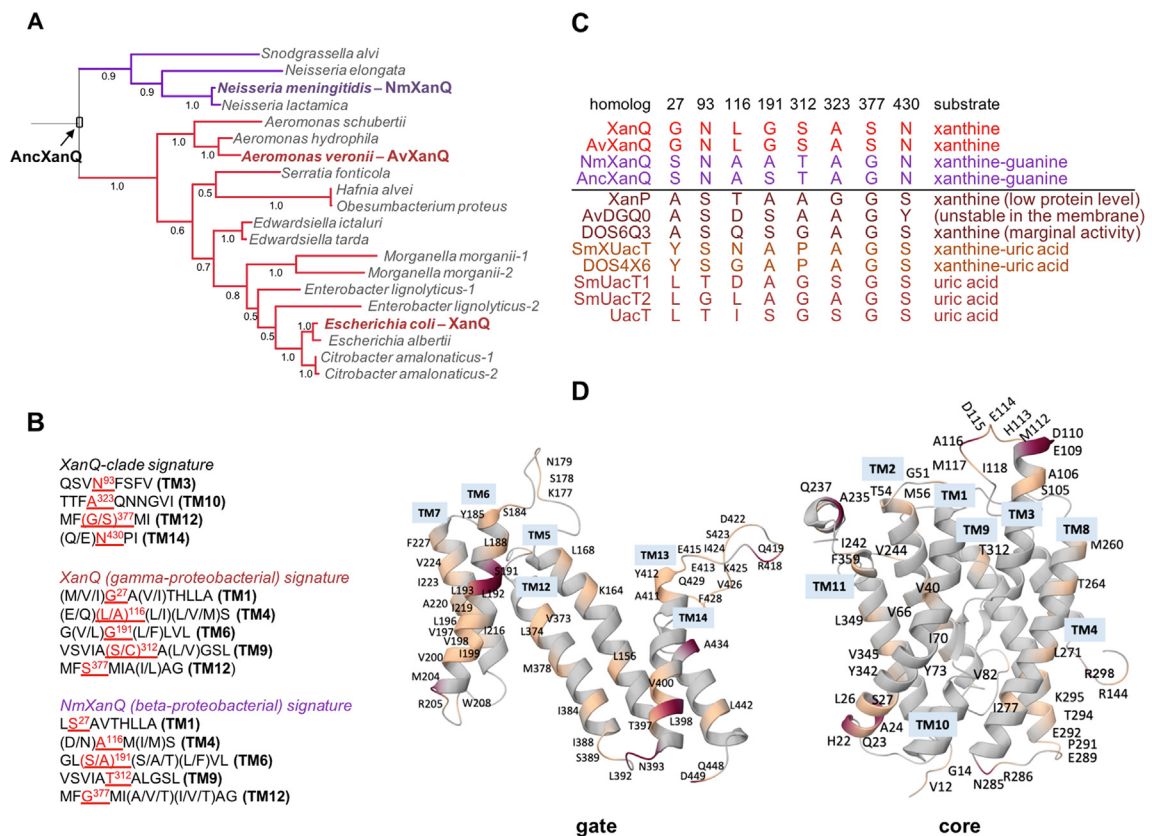
In our current work, we use ancestral protein reconstruction to study the evolution of bacterial xanthine permeases of the NAT/NCS2 family. Our results provide evidence that the modern xanthine-specific permeases have derived from xanthine-guanine transporters of broader specificity and the shift from polyspecific to xanthine-specific transporters is linked with few changes outside the binding-site region.

## Results

### Elucidation of the XanQ clade and reconstruction of an ancestral XanQ transporter

The xanthine permease XanQ of *E. coli* K-12 is unique among the functionally known permeases of NAT/NCS2 family as it is specific for xanthine and not any other purine or pyrimidine nucleobase. XanQ is classified phylogenetically in a major cluster of subfamily NAT/COG2233, namely the XanQ/UapA cluster (or Cluster C1\_Xanthine-Uric Acid).<sup>10</sup> The XanQ/UapA cluster contains several well characterized permeases which can transport both xanthine and uric acid, including the structurally known fungal UapA<sup>2</sup> and the rhizobial SmXUacT,<sup>8</sup> as well as permeases specific for uric acid like the *E. coli* UacT,<sup>50</sup> and the rhizobial SmUacT1 and SmUacT2.<sup>8</sup> Although highly similar in sequence and in all key binding-site residues, the xanthine-specific XanQ differs from all the above permeases in few characteristic amino acid residues that are functionally relevant (see below).

We performed a detailed phylogenetic analysis of all homologs of Proteobacteria that belong to the XanQ/UapA cluster and specified that it comprises seven major monophyletic subgroups (Supplement, Figure S1), one of which contains XanQ and was given the name *XanQ clade*. All



**Figure 1.** Phylogenetic and sequence analysis of AncXanQ. (A) Phylogenetic analysis of the 20 XanQ orthologs representing one genome per *xanQ*-containing species derived from the analysis of fully sequenced genomes retrieved from the IGM/M database at JGI. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jones-Taylor-Thornton matrix-based model as implemented in MEGA7. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown (as decimal) next to each branch. Orthologs are denoted by the name of the corresponding species; their GenBank accession numbers are given in Supplement, Figure S4. For further details, see text and Supplementary Figure S1. (B) Characteristic amino acid residues found in all (top), all gamma-proteobacterial (middle) or all beta-proteobacterial XanQ orthologs (bottom) and their flanking sequence context. The residues shown in red and underlined (and numbered according to XanQ) are either important for function in XanQ and not conserved in other NAT/NCS2 groups (XanQ-clade signature) or distinguish between the gamma- and beta-proteobacterial orthologs and linked with specificity, as shown in the current study. (C) Amino acid occupation of the positions highlighted in B in the functionally known proteobacterial homologs of the XanQ/UapA cluster. (D) Structural homology model of AncXanQ based on the inferred amino acid ancestral sequence showing separately the region of the core domain (residues 22–135 and 259–352) and the gate domain (residues 151–229 and 368–445). Positions of amino acids predicted with a posterior probability of <90% (but >50%) or <50% are named and shown in orange or red, respectively. The AncXanQ sequence (shown in Supplement, Figure S2) was inferred using Maximum Likelihood statistics in MEGA7.

homologs in the XanQ clade (Figure 1(A)) are closely related to each other (>65% pairwise sequence identity) and share a combination signature of four important XanQ residues not found in any other NAT/NCS2 (Figure 1(B)). These are: Asn93 (numbering according to XanQ), which participates to a H-bond network peripheral to the binding site and has been linked with the specificity preference for xanthine over uric acid<sup>36,21</sup>; Ala323, which is in the substrate binding site and may also confer to specificity<sup>24,50</sup>; Ser/Gly377, next to a conserved Phe/Tyr that contributes to substrate coordination at the periphery

of the binding site<sup>37</sup>; and Asn430, which is distant from the binding site but may influence specificity indirectly.<sup>52,38</sup> Of these residues, Asn93 and Asn430 are invariably present in the XanQ clade but absent in any other NAT/NCS2 homolog, Ala323 is invariable in the XanQ clade and inconsistently present in other NAT/NCS2s (having Ala, Gly or Ser at this position), and Ser377 is uniquely present in a subset of the XanQ clade whereas all other bacterial NAT/NCS2s have a Gly at this position (Figure 1(C)).

Our previous mutagenesis studies of XanQ<sup>21,37</sup> and other homologs in the XanQ/UapA cluster,<sup>50,8</sup>



which involved over 1000 site-directed mutants at more than 80% of the permease residues, did not reveal any XanQ mutant which deviates substantially from the xanthine-specific profile of XanQ. To address the evolutionary basis of this resilient substrate profile of XanQ, we set to explore the putative ancestral protein sequence of the XanQ clade and its substrate specificity properties.

We inferred the putative protein sequence of the ancestral node of the XanQ clade using Maximum Likelihood statistics in MEGA7. The ancestral sequence was reconstructed with high confidence (Supplement, Figure S2). The posterior probabilities (PP) for most residues (>80% for residues predicted to be in transmembrane segments, and 100% for binding-site residues) were >90% whereas the few less-certain amino acid residues fall mostly at tips of transmembrane helices or at hydrophilic loops of the homology-modeled protein (Figure 1(D)). The reconstructed ancestral protein, which was termed AncXanQ, differs from XanQ at 24% of residues of the overall sequence and conserves all key residues predicted to be at the binding site region (Supplement, Figure S3). A DNA sequence that codes for this ancestral protein was designed, codon-optimized for expression in *E. coli* (Supplement, Table S1) and produced by gene synthesis.

### The functional profile of AncXanQ

The synthesized coding sequence of AncXanQ was tagged with a biotin-acceptor domain at C terminus and transferred to pT7-5 for expression in *E. coli* K-12. Western blot analysis of the protein product of the synthetic gene in the host cytoplasmic membrane shows that AncXanQ is stably present in the membrane at levels comparable to the ones of XanQ control (Supplement, Figure S5). Transport assay analysis shows that AncXanQ displays lower xanthine uptake activity but broader substrate profile than XanQ (Figure 2(A) and (B)). Of all putative purine and pyrimidine substrates tested (adenine, guanine, hypoxanthine, xanthine, uric acid, uracil thymine, cytosine), AncXanQ transports guanine and xanthine, in contrast to XanQ which transports only xanthine. Similar to the case of XanQ,<sup>34</sup> the transport activities of AncXanQ are inhibited by carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), implying that they are proton-gradient dependent (Supplement, Figure S6). AncXanQ transports xanthine with comparable affinity ( $K_M$ ) but 5-fold lower efficiency ( $V_{max}/K_M$ ) relative to XanQ (Table 1). With respect to the guanine-transport activity, AncXanQ transports guanine with high affinity ( $K_M$  0.2  $\mu$ M), which is higher than the one of GhxP, the major guanine-hypoxanthine transporter of *E. coli* K-12,<sup>51</sup> but lower capacity ( $V_{max}$ ), resulting in comparable efficiency

( $V_{max}/K_M$ ) relative to the modern guanine transporter (Table 1).

The xanthine transport activity of AncXanQ is inhibited with high affinity by a range of analogs, including 7-methylxanthine, 8-methylxanthine, as well as guanine, which are not ligands for XanQ (Figure 2(C); Table 2). The guanine transport activity of AncXanQ is inhibited with high affinity by numerous purine nucleobases and analogs, including hypoxanthine (HX), 6-mercaptopurine (6-MP), 6-thioguanine (6-SG), 1-methylguanine (1-MG), and 8-azaguanine (8-azaG), which are high-affinity ligands for GhxP,<sup>51</sup> but also adenine (A), 2,6-diaminopurine (2,6-DAP), 7-methylguanine (7-MG), and xanthine which are not ligands for GhxP (Figure 2(C), Table 2). Overall, AncXanQ recognizes as ligands not only guanine and xanthine, but also adenine and hypoxanthine, and all of their analogs tested, except allopurinol and uric acid. This profile is different than any modern NAT/NCS2 profile recorded thus far for either a bacterial or a eukaryotic transporter (see Discussion).

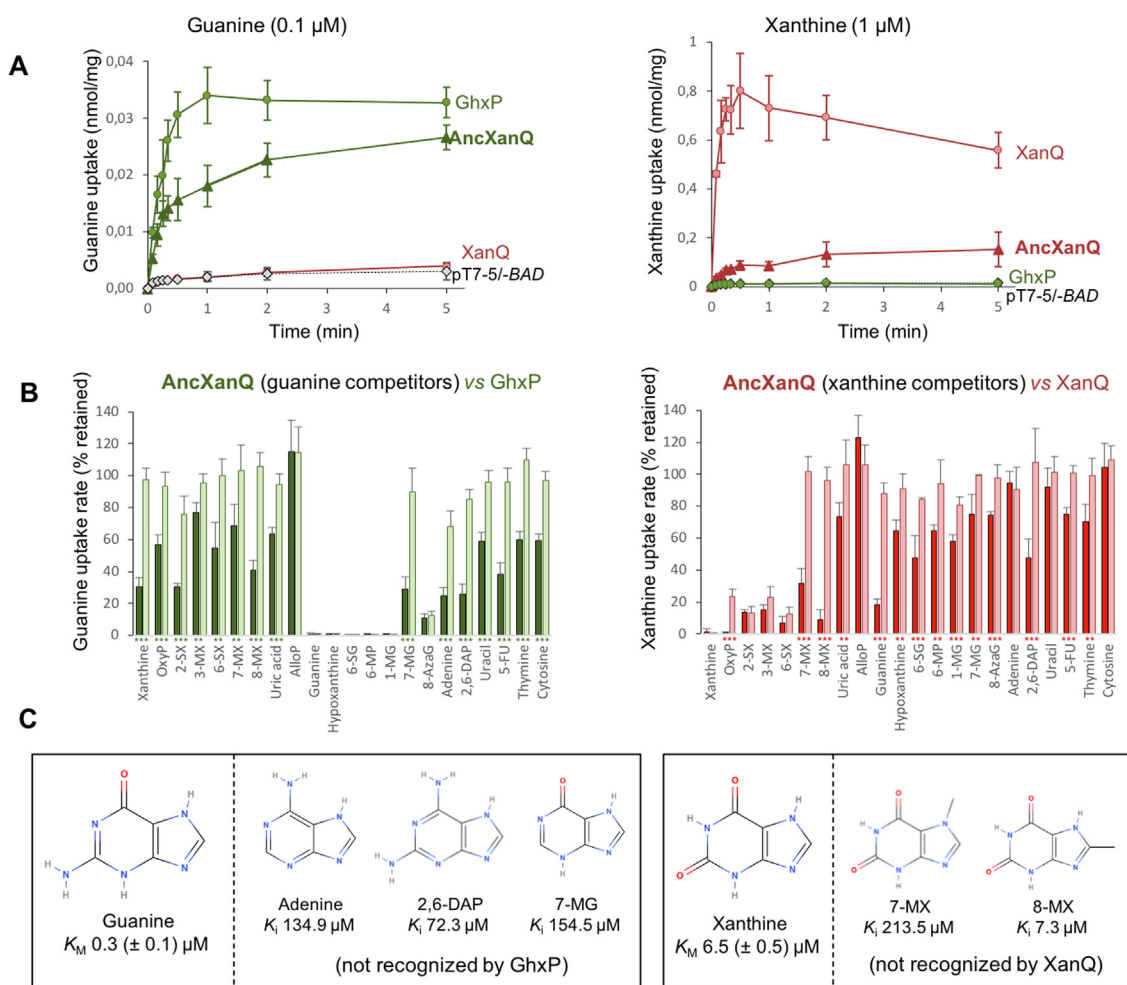
### Structural model of AncXanQ

In order to address the mode of interaction of xanthine, guanine and their analogs in the AncXanQ binding pocket, we constructed a homology structural model of AncXanQ. This model was constructed on the template of the XanQ model which had been reported in an inward-open conformation<sup>8</sup> based on the UapA (PDB 516C) crystal structure.<sup>2</sup> AncXanQ model comprises 14 transmembrane segments (TMs), of which TMs 1–4 and 8–11 form the core domain and TMs 5–7 and 12–14 the gate domain (Figure 3(A)). All transmembrane helices and interconnecting loops display the same length as in XanQ, as a consequence of the high (76%) identity shared between the two transporter sequences.

The initial AncXanQ model was subjected to 50 ns molecular dynamics simulation in the *apo* form embedded in DPPC lipid bilayer solvated with explicit water. The C $\alpha$ -carbons RMSD calculated did not exceed 1 Å suggesting that the model structure is stable.

### Binding mode of xanthine and guanine on AncXanQ

Since AncXanQ transports both xanthine and guanine, induced fit docking calculations (IFD) was performed to characterize the binding mode of these ligands in the binding pocket of the transporter. Xanthine is accommodated in the binding pocket of AncXanQ exhibiting a remarkable similarity with the binding mode in UapA<sup>2</sup> and in XanQ<sup>8</sup> (Figure 3(A)). Considering that there are no related structural data for the interaction of guanine in any other NAT/NCS2 transporter, we decided to retain only the theoretically derived poses exhibiting similar orientation in the binding



**Figure 2.** AncXanQ is a broad-specificity guanine/xanthine transporter. (A) *E. coli* JW4025 and T184 expressing the indicated permeases through pT7-5/-BAD were assayed for uptake of [ $^3\text{H}$ ]-guanine (0.1  $\mu\text{M}$ ) and [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ), respectively, at 25  $^\circ\text{C}$ , as shown. Each data point represents the mean and standard deviation (SD) from triplicate measurements. (B) *E. coli* JW4025 expressing AncXanQ (dark green) or GhxP (light green) were assayed for initial rates of [ $^3\text{H}$ ]-guanine (0.1  $\mu\text{M}$ ) uptake and T184 expressing AncXanQ (dark red) or XanQ (light red) were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) uptake, in the absence or presence of the indicated non-radiolabeled nucleobases and analogs (1 mM). Transport rates are expressed as percentages of the rate obtained in the absence of competitor. Each bar represents the means of five to six determinations with standard deviation (SD) shown. The values obtained for the ancestral and the extant transporters were compared for each ligand using unpaired two-tailed t-test (as implemented in GraphPad Prism 8.0) and statistically significant differences are indicated with one ( $p < 0.05$ ), two ( $p < 0.01$ ) or three asterisks ( $p < 0.001$ ) under each pair of histogram bars. Values obtained with cells harboring pT7-5/-BAD alone have been subtracted from the sample measurements in all cases. (C) AncXanQ ligands that are not recognized by GhxP (on the left) or XanQ (on the right). See Tables 1 and 2 for further details. Abbreviations of the analogs are as in Table 2.

cavity for guanine and all guanine derivatives identified experimentally as high-affinity ligands (1-MG, 6-MP, 6-SG, HX, 8-AzaG, 2,6-DAP, 7-MG, A) (see Table 2). The selected binding poses showed to be the most energetically favorable and are presented for guanine and 1-MG in Figure 4. The lowest energy pose (Figure 4(A)) suggests that the AncXanQ-guanine structure is stabilized by H-bond formation of N1H and C2-NH<sub>2</sub> with E272 carboxylate as well as N3 with Q324, N7 with F94 backbone carbonyl group and p-p and T stacking interactions with F322 and F94, respectively. In a

similar lowest energy structure (Figure 4(C)), 1-MG is stabilized in the binding cavity by H-bond formation of C2-NH<sub>2</sub> with E272, N7 with F94 backbone and p-p and T stacking interactions with F322 and F94, respectively. However, in the 1-MG complex, the steric interaction between the methyl group and the E272 carboxylate implies a slight shift compared to guanine resulting in loss of the interaction with Q324 (Figure 4(C)). In the second lowest energy pose (Figure 4(B)), a H-bond is formed between guanine N1H and E272, C6 = O and A323 backbone, N7 and Q324, as well as p-p and

Table 1  $K_M$  and  $V_{max}$  values of AncXanQ, XanQ, NmXanQ and mutants.

Permease	Xanthine uptake			Guanine uptake		
	$K_M$ ( $\mu\text{M}$ $\text{mg}^{-1}$ )	$V_{max}$ ( $\text{nmol min}^{-1}$ $\text{mg}^{-1}$ )	$V_{max}/K_M$ ( $\text{ml min}^{-1}$ $\text{mg}^{-1}$ )	$K_M$ ( $\mu\text{M}$ $\text{mg}^{-1}$ )	$V_{max}$ ( $\text{nmol min}^{-1}$ )	$V_{max}/K_M$ ( $\text{ml min}^{-1}$ )
XanQ	4.1 ( $\pm 0.5$ )	7.5 ( $\pm 1.1$ )	1.83	n.a.	n.a.	
AncXanQ	6.5 ( $\pm 0.9$ )	2.4 ( $\pm 0.1$ )	0.37	0.24 ( $\pm 0.05$ )	0.25 ( $\pm 0.02$ )	1.04
AvpXanQ	8.9 ( $\pm 2.8$ )	20.9 ( $\pm 2.4$ )	2.35	n.a.	n.a.	
NmXanQ	11.8 ( $\pm 3.5$ )	2.5 ( $\pm 0.2$ )	0.21	1.56 ( $\pm 0.07$ )	0.17 ( $\pm 0.01$ )	0.11
GhxP	n.a.	n.a.	1.10 ( $\pm 0.12$ )	1.02 ( $\pm 0.09$ )	0.93	
XanQ	4.1 ( $\pm 0.5$ )	7.5 ( $\pm 1.1$ )	1.83	n.a.	n.a.	
XanQ-S377G	3.6 ( $\pm 1.3$ )	11.8 ( $\pm 1.2$ )	2.88	0.57 ( $\pm 0.18$ )	0.16 ( $\pm 0.02$ )	0.28
XanQ-G27S	12.6 ( $\pm 2.2$ )	32.4 ( $\pm 2.1$ )	2.57	n.a.	n.a.	
XanQ-L116A	6.4 ( $\pm 0.7$ )	77.1 ( $\pm 3.1$ )	12.04	n.a.	n.a.	
XanQ-G191S	5.6 ( $\pm 0.6$ )	55.8 ( $\pm 2.5$ )	9.96	n.a.	n.a.	
XanQ-S312T	2.7 ( $\pm 0.6$ )	24.9 ( $\pm 1.7$ )	9.22	n.a.	n.a.	
XanQ-2M (27)	4.1 ( $\pm 0.5$ )	28.9 ( $\pm 1.1$ )	7.05	n.a.	n.a.	
XanQ-2M (191)	4.3 ( $\pm 0.5$ )	50.8 ( $\pm 2.8$ )	12.09	0.68 ( $\pm 0.05$ )	0.39 ( $\pm 0.01$ )	0.57
XanQ-2M (312)	3.1 ( $\pm 0.4$ )	35.1 ( $\pm 1.2$ )	11.32	n.a.	n.a.	
XanQ-4M	7.6 ( $\pm 0.4$ )	92.7 ( $\pm 1.2$ )	12.20	n.a.	n.a.	
XanQ-5M	3.5 ( $\pm 1.2$ )	45.6 ( $\pm 4.7$ )	13.03	n.a.	n.a.	
AncXanQ	6.5 ( $\pm 0.9$ )	2.4 ( $\pm 0.1$ )	0.37	0.24 ( $\pm 0.05$ )	0.25 ( $\pm 0.02$ )	1.04
AncXanQ-G377S	3.4 ( $\pm 0.7$ )	2.6 ( $\pm 0.2$ )	0.75	0.50 ( $\pm 0.08$ )	0.46 ( $\pm 0.05$ )	0.92
AncXanQ-S27G	5.8 ( $\pm 1.1$ )	4.2 ( $\pm 0.2$ )	0.72	1.02 ( $\pm 0.09$ )	0.18 ( $\pm 0.01$ )	0.18
AncXanQ-2M	6.7 ( $\pm 0.6$ )	3.5 ( $\pm 0.1$ )	0.52	0.85 ( $\pm 0.06$ )	0.42 ( $\pm 0.01$ )	0.49
AncXanQ-4M	9.5 ( $\pm 0.8$ )	4.9 ( $\pm 0.2$ )	0.52	n.a.	n.a.	
AncXanQ-5M	4.4 ( $\pm 1.5$ )	3.9 ( $\pm 0.3$ )	0.87	n.a.	n.a.	
NmXanQ	11.8 ( $\pm 3.5$ )	2.5 ( $\pm 0.2$ )	0.21	1.56 ( $\pm 0.07$ )	0.17 ( $\pm 0.01$ )	0.11
NmXanQ-G377S	6.8 ( $\pm 2.0$ )	4.2 ( $\pm 0.7$ )	0.62	5.01 ( $\pm 0.75$ )	0.35 ( $\pm 0.02$ )	0.07
NmXanQ-A191S	13.8 ( $\pm 2.9$ )	3.1 ( $\pm 0.2$ )	0.22	0.76 ( $\pm 0.13$ )	0.17 ( $\pm 0.01$ )	0.22

*E. coli* T184 or JW4025 expressing the corresponding constructs were assayed for initial rates of [ $^3\text{H}$ ]xanthine or [ $^3\text{H}$ ]guanine uptake, respectively, at eight concentrations in the range of 0.01  $\mu\text{M}$  to 0.04 mM, at 25 °C. Transport rates were measured at 5–15 sec. Kinetic parameters were determined from non-linear regression fitting to the Michaelis-Menten equation. The values are the means of three determinations with standard deviations (SD) shown in parentheses. n.a. not applied (due to negligible transport rates at any concentration tested). Abbreviations for multiple-site mutants: XanQ-2M (27), XanQ-G27S/S377G; XanQ-2M (191), XanQ-G191S/S377G; XanQ-2M (312), XanQ-S312T/S377G; XanQ-4M, XanQ-G27S/L116A/G191S/S312T; XanQ-5M, XanQ-G27S/L116A/G191S/S312T/S377G; AncXanQ-2M, AncXanQ-S27G/G377S; AncXanQ-4M, AncXanQ-S27G/A116L/S191G/T312S; AncXanQ-5M, AncXanQ-S27G/A116L/S191G/T312S/G377S.



Table 2 Specificity profiles of AncXanQ, XanQ, NmXanQ and mutants

Permease	$K_i$ ( $\mu\text{M}$ ) for inhibition of xanthine/guanine uptake by other nucleobases or analogs						
	Xanthine uptake	2-SX	3-MX	6-SX	7-MX	8-MX	OxyP
	G						
XanQ	n.a.	91.1 ( $\pm 10.1$ )	71.9 ( $\pm 9.2$ )	40.6 ( $\pm 8.3$ )	n.a.	n.a.	29.1 ( $\pm 3.8$ )
AncXanQ	51.6 ( $\pm 10.8$ )	111.5 ( $\pm 14.6$ )	49.6 ( $\pm 13.3$ )	31.1 ( $\pm 14.3$ )	213.5 ( $\pm 27.2$ )	7.3 ( $\pm 0.5$ )	7.7 ( $\pm 2.2$ )
AvpXanQ	n.a.	10.7 ( $\pm 2.2$ )	40.4 ( $\pm 11.0$ )	7.7 ( $\pm 1.8$ )	n.a.	n.a.	73.1 ( $\pm 20.4$ )
NmXanQ	29.0 ( $\pm 6.0$ )	38.4 ( $\pm 5.8$ )	62.8 ( $\pm 11.8$ )	41.3 ( $\pm 8.5$ )	n.a.	n.a.	2.7 ( $\pm 0.8$ )
XanQ	n.a.	91.1 ( $\pm 10.1$ )	71.9 ( $\pm 9.2$ )	40.6 ( $\pm 8.3$ )	n.a.	n.a.	29.1 ( $\pm 3.8$ )
XanQ-S377G	42.3 ( $\pm 11.9$ )	71.5 ( $\pm 17.4$ )	n.a.	3.0 ( $\pm 0.8$ )	n.a.	n.a.	3.0 ( $\pm 0.2$ )
XanQ-G27S	n.a.	30.5 ( $\pm 3.0$ )	35.1 ( $\pm 19.4$ )	12.1 ( $\pm 3.8$ )	n.a.	n.a.	17.1 ( $\pm 2.5$ )
XanQ-L116A	n.a.	16.5 ( $\pm 2.7$ )	34.1 ( $\pm 8.4$ )	33.2 ( $\pm 6.1$ )	n.a.	n.a.	17.8 ( $\pm 5.3$ )
XanQ-G191S	n.a.	32.8 ( $\pm 6.9$ )	22.6 ( $\pm 6.1$ )	13.6 ( $\pm 5.0$ )	n.a.	n.a.	15.0 ( $\pm 3.9$ )
XanQ-S312T	n.a.	21.2 ( $\pm 7.3$ )	23.7 ( $\pm 8.1$ )	24.9 ( $\pm 3.8$ )	74.5 ( $\pm 15.0$ )	66.0 ( $\pm 17.5$ )	23.2 ( $\pm 4.5$ )
XanQ-2M (27)	n.a.	21.1 ( $\pm 5.6$ )	56.1 ( $\pm 10.6$ )	16.1 ( $\pm 2.0$ )	n.a.	n.a.	15.6 ( $\pm 4.1$ )
XanQ-2M (191)	34.2 ( $\pm 9.4$ )	22.8 ( $\pm 5.8$ )	30.5 ( $\pm 5.0$ )	18.2 ( $\pm 2.5$ )	n.a.	n.a.	19.5 ( $\pm 2.5$ )
XanQ-2M (312)	n.a.	16.4 ( $\pm 3.6$ )	21.0 ( $\pm 6.8$ )	14.7 ( $\pm 6.5$ )	n.a.	n.a.	6.2 ( $\pm 2.3$ )
XanQ-4M	n.a.	82.7 ( $\pm 14.1$ )	32.3 ( $\pm 7.9$ )	n.a.	n.a.	n.a.	49.5 ( $\pm 12.0$ )
XanQ-5M	n.a.	48.6 ( $\pm 7.3$ )	32.1 ( $\pm 8.0$ )	35.6 ( $\pm 6.7$ )	n.a.	n.a.	25.1 ( $\pm 5.6$ )
AncXanQ	51.6 ( $\pm 10.8$ )	111.5 ( $\pm 14.6$ )	49.6 ( $\pm 13.3$ )	31.1 ( $\pm 14.3$ )	213.5 ( $\pm 27.2$ )	7.3 ( $\pm 0.5$ )	7.7 ( $\pm 2.2$ )
AncXanQ-G377S	33.8 ( $\pm 5.6$ )	38.1 ( $\pm 3.0$ )	46.9 ( $\pm 1.8$ )	12.2 ( $\pm 1.8$ )	n.a.	n.a.	4.2 ( $\pm 0.7$ )
AncXanQ-S27G	27.0 ( $\pm 6.2$ )	23.2 ( $\pm 14.6$ )	33.0 ( $\pm 6.1$ )	50.3 ( $\pm 17.7$ )	n.a.	n.a.	24.5 ( $\pm 8.5$ )
AncXanQ-2M	16.6 ( $\pm 3.1$ )	22.9 ( $\pm 2.5$ )	47.8 ( $\pm 8.6$ )	26.4 ( $\pm 4.9$ )	n.a.	n.a.	6.4 ( $\pm 2.8$ )
AncXanQ-4M	n.a.	49.6 ( $\pm 11.1$ )	41.2 ( $\pm 11.8$ )	2.7 ( $\pm 0.4$ )	n.a.	n.a.	3.1 ( $\pm 0.2$ )
AncXanQ-5M	n.a.	67.4 ( $\pm 15.9$ )	61.3 ( $\pm 16.9$ )	46.4 ( $\pm 12.6$ )	38.6 ( $\pm 9.8$ )	n.a.	27.9 ( $\pm 3.7$ )
NmXanQ	29.0 ( $\pm 6.0$ )	38.4 ( $\pm 5.8$ )	62.8 ( $\pm 11.8$ )	41.3 ( $\pm 8.5$ )	n.a.	n.a.	2.7 ( $\pm 0.8$ )
NmXanQ-G377S	27.6 ( $\pm 7.0$ )	62.2 ( $\pm 12.7$ )	40.2 ( $\pm 7.7$ )	35.6 ( $\pm 8.7$ )	n.a.	n.a.	26.0 ( $\pm 6.1$ )
NmXanQ-A191S	24.8 ( $\pm 6.9$ )	35.3 ( $\pm 4.5$ )	61.1 ( $\pm 7.6$ )	39.5 ( $\pm 6.5$ )	n.a.	n.a.	9.8 ( $\pm 3.1$ )
	Guanine uptake						
	X	HX	6-SG	6-MP	1-MG	7-MG	8-AzaG
AncXanQ	54.9 ( $\pm 4.0$ )	5.2 ( $\pm 0.2$ )	2.4 ( $\pm 0.3$ )	1.9 ( $\pm 0.2$ )	2.2 ( $\pm 0.5$ )	154.5 ( $\pm 48.5$ )	40.2 ( $\pm 10.1$ )
NmXanQ	50.6 ( $\pm 8.5$ )	21.1 ( $\pm 5.2$ )	6.4 ( $\pm 2.5$ )	5.6 ( $\pm 1.0$ )	22.1 ( $\pm 5.1$ )	n.a.	n.a.
GhxP	n.a.	5.7 ( $\pm 1.1$ )	3.1 ( $\pm 0.3$ )	8.4 ( $\pm 1.5$ )	55.1 ( $\pm 7.4$ )	n.a.	50.5 ( $\pm 6.9$ )
XanQ-S377G	22.3 ( $\pm 5.1$ )	18.7 ( $\pm 5.3$ )	4.5 ( $\pm 2.1$ )	15.0 ( $\pm 2.7$ )	6.6 ( $\pm 0.3$ )	3.6 ( $\pm 1.5$ )	n.a.
XanQ-2M (191)	32.0 ( $\pm 5.0$ )	15.2 ( $\pm 1.9$ )	4.1 ( $\pm 1.5$ )	9.4 ( $\pm 2.6$ )	5.7 ( $\pm 0.9$ )	9.9 ( $\pm 1.6$ )	n.a.
AncXanQ	54.9 ( $\pm 4.0$ )	5.2 ( $\pm 0.2$ )	2.4 ( $\pm 0.3$ )	1.9 ( $\pm 0.2$ )	2.2 ( $\pm 0.5$ )	154.5 ( $\pm 48.5$ )	40.2 ( $\pm 10.1$ )
AncXanQ-G377S	35.0 ( $\pm 7.4$ )	18.2 ( $\pm 2.4$ )	7.1 ( $\pm 3.9$ )	8.0 ( $\pm 1.9$ )	25.1 ( $\pm 6.8$ )	5.1 ( $\pm 1.5$ )	n.a.
AncXanQ-S27G	36.0 ( $\pm 9.2$ )	7.5 ( $\pm 4.3$ )	22.6 ( $\pm 7.0$ )	13.9 ( $\pm 3.0$ )	17.4 ( $\pm 3.4$ )	30.1 ( $\pm 10.0$ )	n.a.
AncXanQ-2M	28.7 ( $\pm 7.3$ )	12.8 ( $\pm 2.2$ )	8.6 ( $\pm 1.6$ )	9.9 ( $\pm 2.2$ )	14.2 ( $\pm 2.9$ )	32.4 ( $\pm 10.4$ )	n.a.
NmXanQ	50.6 ( $\pm 8.5$ )	21.1 ( $\pm 5.2$ )	6.4 ( $\pm 2.5$ )	5.6 ( $\pm 1.0$ )	22.1 ( $\pm 5.1$ )	n.a.	n.a.
NmXanQ-G377S	30.6 ( $\pm 4.8$ )	22.8 ( $\pm 3.4$ )	28.4 ( $\pm 1.3$ )	6.1 ( $\pm 8.6$ )	14.6 ( $\pm 2.7$ )	n.a.	n.a.
NmXanQ-A191S	44.8 ( $\pm 7.0$ )	15.2 ( $\pm 4.6$ )	6.1 ( $\pm 1.3$ )	5.9 ( $\pm 0.8$ )	19.8 ( $\pm 4.6$ )	n.a.	n.a.

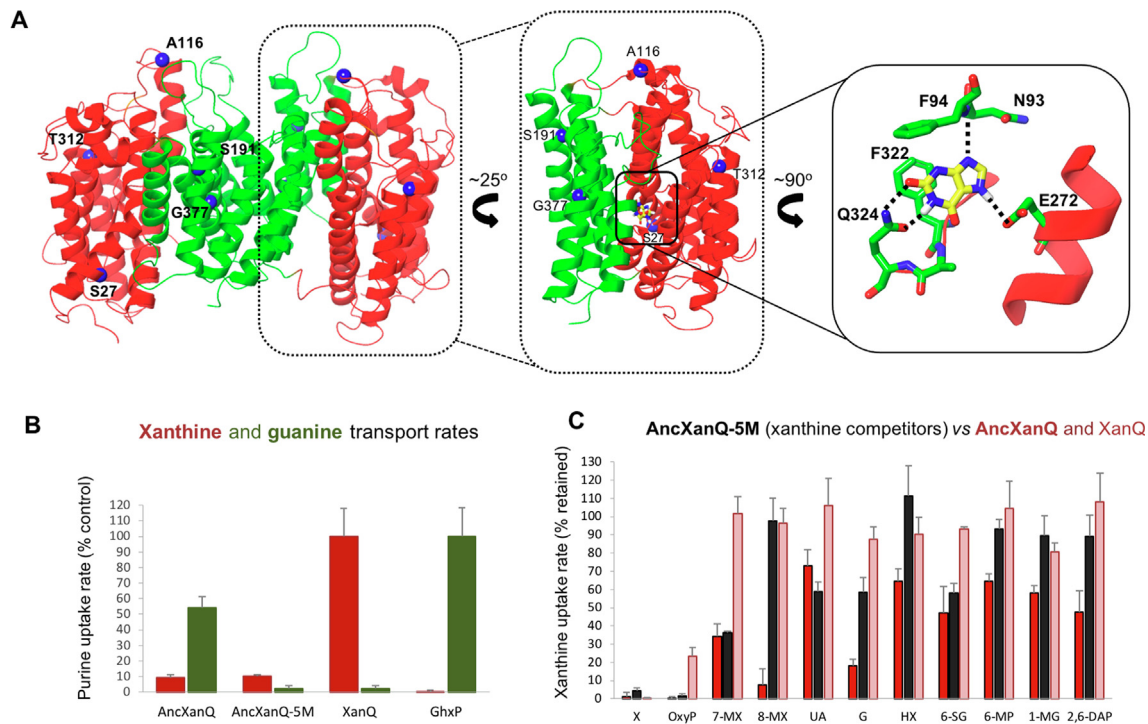
*E. coli* T184 or JW4025 expressing the corresponding constructs were assayed for initial rates of [ $^3\text{H}$ ]xanthine (1  $\mu\text{M}$ ) or [ $^3\text{H}$ ]guanine uptake (0.1  $\mu\text{M}$ ), respectively, at 25  $^\circ\text{C}$ . Transport rates were measured at 5–15 sec, in the absence or presence of the indicated unlabeled competitors. Competitors were applied at eight concentrations in the range of 0.01  $\mu\text{M}$  to 1 mM. The data were analyzed with *Prism7* to deduce  $\text{IC}_{50}$  and thereof  $K_i$  values. The values are the means of three determinations with standard deviations (SD) shown in parentheses. n.a. not applied (<40% inhibition of transport activity in any ligand concentration tested). Abbreviations of nucleobases and analogs: G, guanine; 2-SX, 2-thioxanthine; 3-MX, 3-methylxanthine; 6-SX, 6-thioxanthine; 7-MX, 7-methylxanthine; 8-MX, 8-methylxanthine; OxyP, oxypurinol; X, xanthine; HX, hypoxanthine; 6-SG, 6-thioguanine; 6-MP, 6-mercaptopurine; 1-MG, 1-methylguanine; 7-MG, 7-methylguanine; 8-azaG, 8-azaguanine. Although not shown, adenine and 2,6-diaminopurine inhibit the guanine uptake activity of AncXanQ with deduced  $K_i$  values of 134.9 ( $\pm 42.2$ )  $\mu\text{M}$  and 72.3 ( $\pm 22.4$ )  $\mu\text{M}$ , respectively, but not of any other permease or mutant presented in the table (see also Figure 2). Abbreviations for multiple-site mutants are as in Table 1.

T stacking interactions with F322 and F94, respectively. In the similar orientation, 1-MG is shifted compared to guanine (Figure 4(D)) forming H-bonds between N3 and F94 backbone, C6 = O and Q324 and p-p stacking interactions with both F322 and F94. The lowest-energy AncXanQ-ligand complexes with the other guanine analogs are shown in Supplementary Figure S7. Hypoxanthine, 6-SG, 7-MG, and adenine adopt the same two lowest-energy orientations and similar interac-

tions in the binding pocket as guanine and 1-MG, whereas 2,6-DAP and 6-MP favor a binding orientation and binding interactions similar to xanthine.

### Combination of 5 mutations outside the binding site restricts specificity of AncXanQ to recognition of xanthine

Other than conserved hydrophobic-residue variations, AncXanQ differs from XanQ at only five



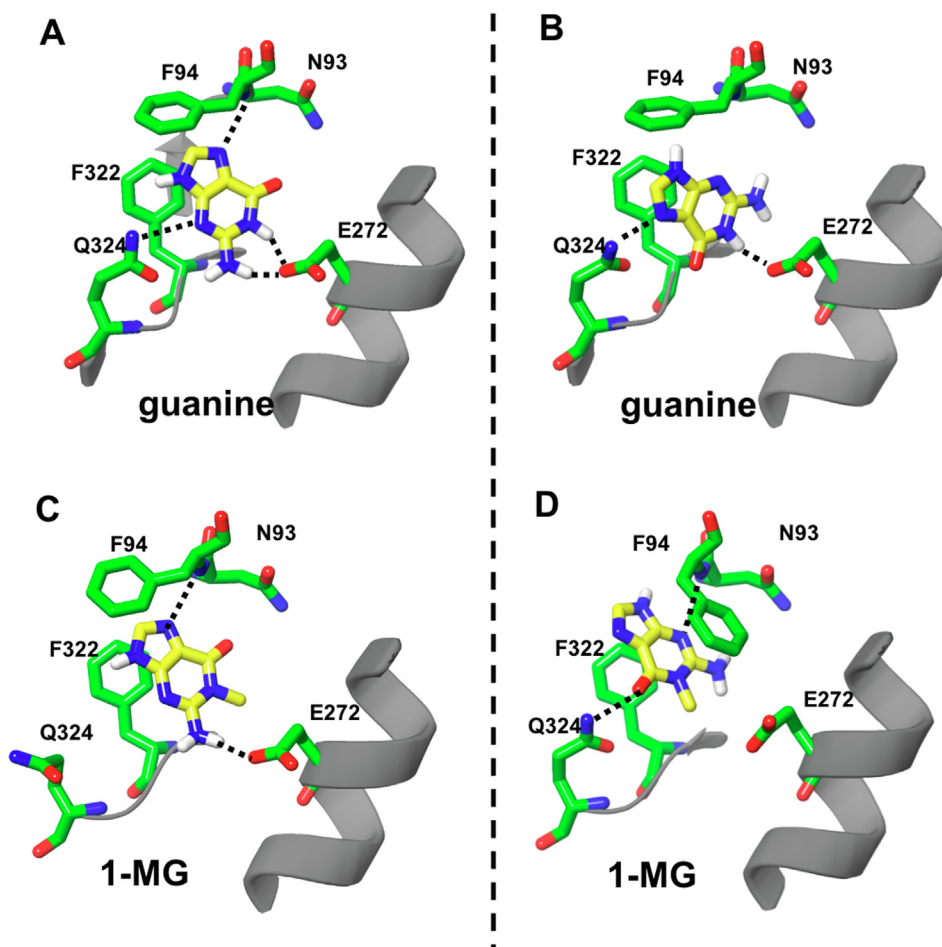
**Figure 3.** AncXanQ-5M approximates the profile of the xanthine-specific XanQ. (A) AncXanQ dimer derived from homology modeling. The core domain is shown in red and the gate domain in green. The positions of residues S27, A116, T312, S191 and G377 are presented as blue spheres, while the computational predicted mode of xanthine in the binding site is highlighted. (B) Xanthine and guanine transport rates for AncXanQ and the quintuple mutant AncXanQ-5M (S27G/A116L/T312S/S191G/G377S) relative to the xanthine transporter XanQ and the guanine transporter GhxP. Initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) or [ $^3\text{H}$ ]-guanine (0.1  $\mu\text{M}$ ) uptake were measured with *E. coli* T184 or JW4025, respectively, expressing the indicated permeases. The results are expressed as percentages of the value obtained for XanQ (xanthine uptake, red bars) or GhxP (guanine uptake, green bars). Values obtained with pT7-5/-BAD alone were subtracted from the sample measurements in all cases. Each bar represents the means of three determinations with standard deviation (SD) shown. (C) Comparison of the ligand inhibition profile of AncXanQ-5M with the profiles of AncXanQ and XanQ, highlighting differences in certain analogs. *E. coli* T184 expressing AncXanQ (dark red), AncXanQ-5M (black) or XanQ (light red) were assayed for initial rates of [ $^3\text{H}$ ]-xanthine (1  $\mu\text{M}$ ) uptake, in the absence or presence of non-radiolabeled analogs (1 mM). All other experimental and presentation details are as in Figure 2(B). The abbreviations are as in Table 2.

amino acid positions predicted to be in transmembrane segments (Figure 3(A) and Supplementary Figure S3). The deviating amino acids of AncXanQ are: Ser27 (Gly in XanQ); Ala116 (Leu in XanQ); Ser191 (Gly in XanQ); Thr312 (Ser in XanQ); and Gly377 (Ser in XanQ). All these sites are outside the binding cavity (Figure 3(A)). We generated a mutant including all relevant changes (S27G/A116L/S191G/T312S/G377S) (which we named **AncXanQ-5M**) and found that this variant is expressed in the *E. coli* host membrane normally (Supplement, Figure S5), transports xanthine with similar kinetic properties as wild-type AncXanQ, but cannot transport guanine (Figure 3(B); Table 1). The ligand inhibition profile of this mutant shows also that it does not recognize 8-methylxanthine, although it can still recognize 7-methylxanthine with high affinity (Figure 3(C); Table 2). Thus, AncXanQ-5M

is a xanthine-specific transporter emulating the profile of XanQ.

### A modern permease of the XanQ clade recapitulates the profile of AncXanQ

If we consider all modern homologs in the XanQ clade, the combination of the aforementioned five characteristic residues of AncXanQ is retained only in a small set of orthologs which have Ser, Ala/Thr, Ser/Ala, Thr, and Gly at the relevant positions (Supplement, Figure S4). These orthologs come from beta-proteobacteria, primarily of the genus *Neisseria*, and constitute a separate subclade in the XanQ clade (Figure 1(A)). All other (gamma-proteobacterial) homologs in the XanQ clade share a combination of Gly, Leu/Ala/Gln, Gly, Ser/Cys/Ala, and Ser at the relevant positions, reminiscent of the Gly27/Leu116/Gl



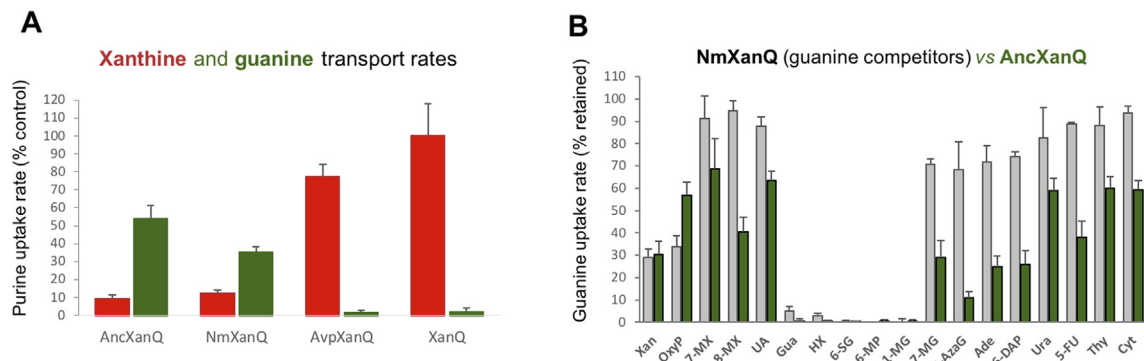
**Figure 4.** Molecular Docking Simulations of the AncXanQ-guanine interactions. Common lowest-energy binding modes of guanine and 1-MG are presented. In the first binding mode, guanine (A) and 1-MG (C) are oriented so that their C2-NH<sub>2</sub> forms a H bond (dashed line) with E272 and their N7 forms a H bond with the F94 backbone. In the second binding mode, guanine (B) and 1-MG (D) are oriented so that their C6 = O forms a H bond with the A323 backbone (guanine) or Q324 (1-MG) while N7 in guanine forms a H bond with Q324 and N3 in 1-MG forms a H bond with the F94 backbone. For further details, see text.

y191/Ser312/Ser377 signature of XanQ. Based on these observations, we selected to study experimentally two XanQ-homologs, the one from *Aeromonas veronii* (AvXanQ) and the other from *Neisseria meningitidis* (NmXanQ), which conserve the five-residue signature of XanQ and AncXanQ, respectively. To be precise, NmXanQ contains a combination of Ser27/Ala116/Ala191/Thr312/Gly377 which differs in one residue (Ala191) from AncXanQ (Ser191). In that case, we engineered also the NmXanQ mutant A191S (see below).

The corresponding genes were mobilized from genomic DNA of *A. veronii* pamvotica (AvpxanQ) and *N. meningitidis* isolate W-5295 (NmXanQ) and expressed through pT7-5 in *E. coli* K-12 yielding high protein levels in the host membrane (Figure 5 (A)). AvpXanQ can transport xanthine with affinity and efficiency comparable to the ones of XanQ (Figure 5(B); Table 1) and does not transport any other nucleobase. In contrast, NmXanQ transports

both xanthine and guanine; it transports xanthine with similar affinity and efficiency as AncXanQ (8-fold lower efficiency relative to XanQ) and guanine with 10-fold lower affinity and efficiency than AncXanQ (similar affinity but 10-fold lower efficiency relative to the *E. coli* transporter for guanine, GhxP) (Figure 5(B); Table 1). The guanine and xanthine transport activities of NmXanQ are inhibited by CCCP, implying that they are proton-gradient dependent (Supplement, Figure S6). Exactly the same functional properties are displayed by NmXanQ mutant A191S, implying that the Ala/Ser191 site is not important for specificity (Tables 1 and 2).

The xanthine uptake activity of NmXanQ is inhibited with high affinity by guanine ( $K_i$  29.0  $\mu$ M) and by xanthine analogs that are ligands of XanQ but not by 7-methylxanthine, 8-methylxanthine or uric acid. The guanine uptake activity of NmXanQ is inhibited with high affinity by xanthine,



**Figure 5.** NmXanQ is a guanine/xanthine transporter mimicking AncXanQ. (A) Xanthine and guanine transport rates for NmXanQ and AvpXanQ in comparison to XanQ and AncXanQ. The initial rates of [ $^3$ H]-xanthine (1  $\mu$ M) or [ $^3$ H]-guanine (0.1  $\mu$ M) uptake were measured with *E. coli* T184 or JW4025, respectively, expressing the permeases through pT7-5/-BAD. The results are expressed as percentages of the value obtained for XanQ (xanthine uptake) (red bars) or GhxP (guanine uptake; not shown in the figure) (green bars). All other experimental and presentation details are as in Figure 3(B). (B) Comparison of the ligand inhibition profiles of NmXanQ and AncXanQ, highlighting part of the analogs assayed. *E. coli* JW4025 expressing NmXanQ (gray) or AncXanQ (green) were assayed for initial rates of [ $^3$ H]-guanine (0.1  $\mu$ M) uptake, in the absence or presence of non-radiolabeled analogs (1 mM). All other experimental and presentation details are as in Figure 2(B). Abbreviations of analogs are as in Table 2.

hypoxanthine, 6-thioguanine, 6-mercaptopurine, and 1-methylguanine, but not by adenine, 2,6-diaminopurine, 7-methylguanine or 8-azaguanine (Figure 5(C); Table 2). These data show that the ligand profile of NmXanQ is broad but less promiscuous than the one of AncXanQ. In contrast, AvpXanQ is a strict xanthine-specific transporter with a ligand profile very similar to XanQ (Table 2).

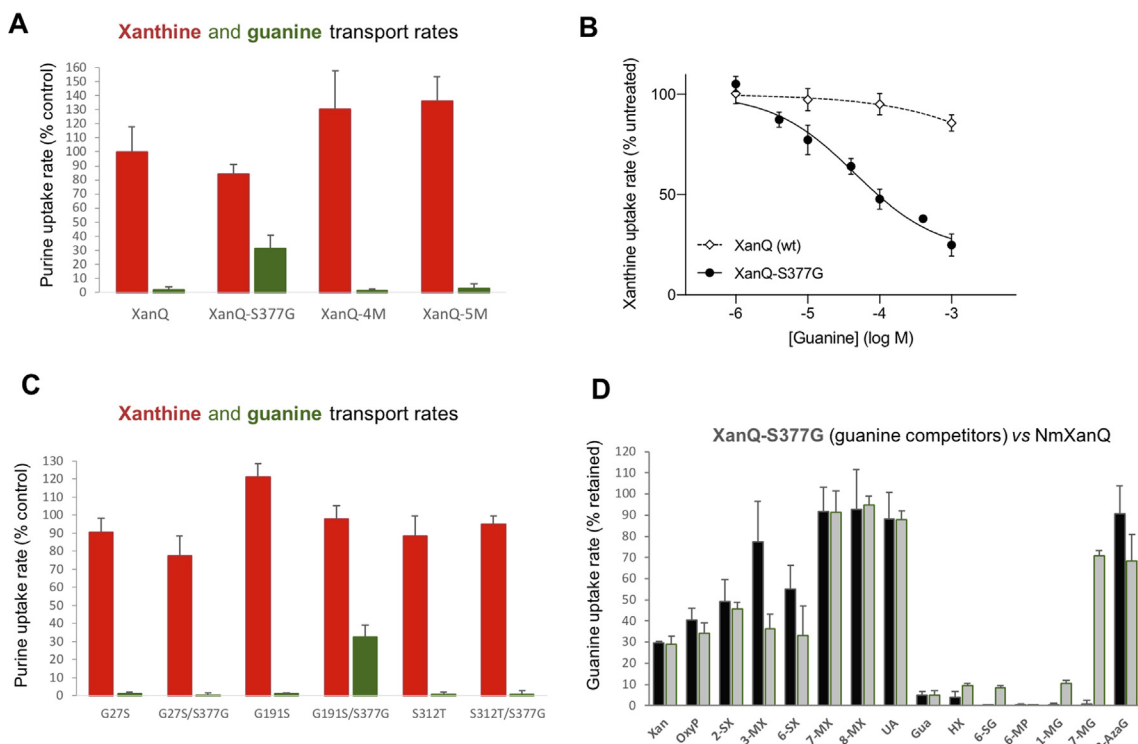
### The role of Ser/Gly377 in relevance to the xanthine specificity of XanQ

Based on the evidence that both AncXanQ and NmXanQ are broad-specificity xanthine/guanine permeases whereas XanQ, AvpXanQ and the mutant AncXanQ-5M are xanthine-specific, it appears that the five-residue combination Ser27/Ala116/Ser(or Ala)191/Thr312/Gly377 in the XanQ clade signifies a permease homolog with broad specificity for purines whereas the combination Gly27/Leu116/Gly191/Ser312/Ser377 signifies a xanthine-specific homolog. Of the five variations, the reconstructed ancestral sequence of the clade retains > 90% posterior probability (PP) for Gly377 but lower PP for Ser27 and Thr312 (50%-90%) or Ala116 and Ser191 (<50%) (Supplement, Figure S2). Ser/Gly377 is one of the four characteristic residues of the XanQ clade (Figure 1(B)) and is next to a conserved Phe/Tyr (Phe376 in XanQ) which is expected to contribute to substrate coordination at the inward-facing permease conformation based on the structural evidence from UraA,<sup>46,67</sup> and UapA.<sup>2</sup> To explore the role of Ser/Gly377 in more detail, we engineered and studied a series of mutants in the AncXanQ, NmXanQ, and XanQ backgrounds.

At first, we asked whether introduction of Gly *in lieu* of Ser377 in a xanthine-specific permease background is sufficient to broaden the specificity. Indeed, mutant XanQ-S377G was found to yield a guanine/xanthine-transporting variant which transports xanthine with similar kinetic properties as wild-type XanQ but, in addition, displays a high-affinity, albeit low-efficiency, transport for guanine (Figure 6(A); Table 1). XanQ-S377G displays a ligand profile reminiscent of NmXanQ, as its xanthine uptake activity is inhibited by guanine ( $K_i$  42.3  $\mu$ M) (Figure 6(B)) but not by 7-methylxanthine, 8-methylxanthine and uric acid, and its guanine uptake activity is inhibited by xanthine, hypoxanthine, 6-thioguanine, 6-mercaptopurine, 1-methylguanine, and 7-methylguanine, but not by adenine, 2,6-diaminopurine, or 8-azaguanine (Figure 6(D); Table 2). However, XanQ-S377G differs from NmXanQ (and AncXanQ), but also from wild-type XanQ, by its inability to recognize 3-methylxanthine. On the other hand, replacement of Ser377 with Gly in the xanthine-specific mutant AncXanQ-5M does not restore the broad specificity profile of AncXanQ; the resulting mutant (AncXanQ-4M) transports xanthine with similar kinetic properties as AncXanQ (Table 1) but does not transport guanine and its xanthine uptake inhibition profile is very similar to that of XanQ (Table 2).

Apart from S377G, no other single-replacement mutant at a position differing from AncXanQ can broaden the xanthine-specific profile. The engineered XanQ mutants G27S, L116A, G191S and S312T were found to transport xanthine with similar affinity and increased efficiency relative to wild type but do not transport guanine or any other

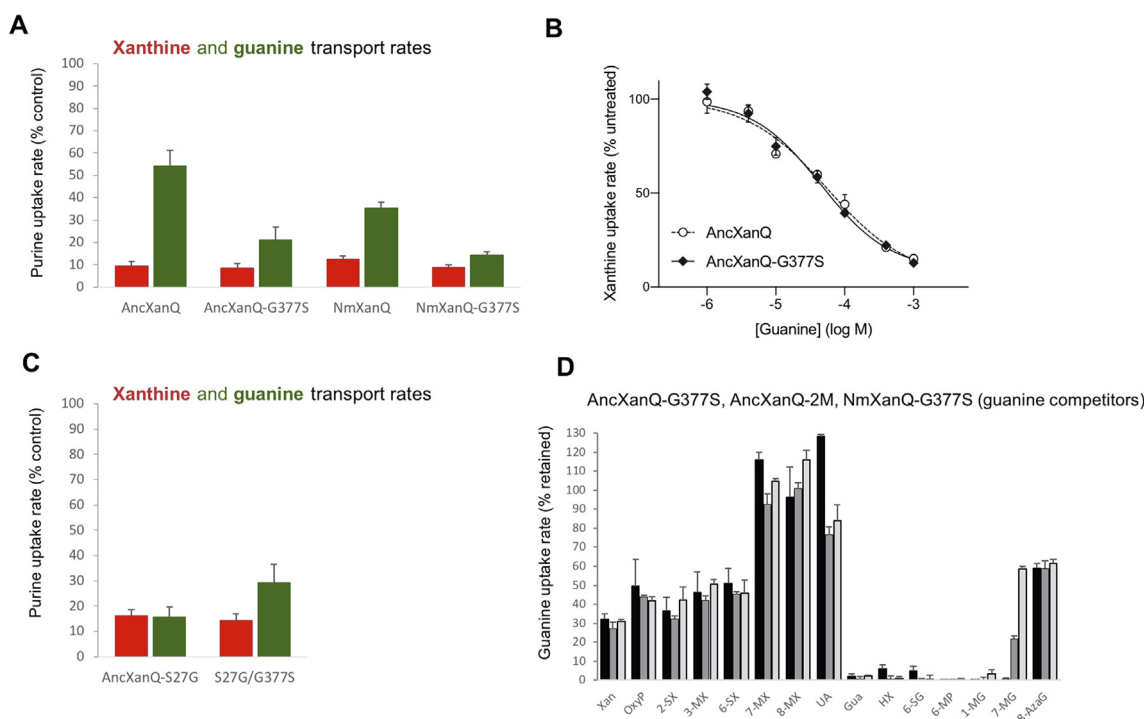




**Figure 6.** The S377G replacement changes XanQ to a guanine/xanthine transporter. (A) and (C). Xanthine and guanine transport rates for XanQ and the indicated mutants. All experimental and presentation details are as in Figure 5(A). (B) *E. coli* T184 expressing XanQ (wild-type) or XanQ-S377G were analyzed for initial rates of [<sup>3</sup>H]-xanthine (1  $\mu$ M) in the presence of guanidine, as indicated, and compared with control cells assayed in the absence of guanidine. Each data point represents the mean and standard deviation (SD) of triplicate measurements (after subtraction of values obtained with pT7-5/-BAD alone, in all cases). Data were analyzed with *Prism7* to obtain  $IC_{50}$  and thereof  $K_i$  values. The guanidine  $K_i$  value deduced from the above experiment (mean  $\pm$  standard error) for XanQ-S377G is  $35.4 \pm 10.1 \mu$ M. (D) Comparison of the ligand inhibition profiles of XanQ-S377G (black) and NmXanQ (gray), highlighting part of the analogs assayed. Experimental and presentation details are as in Figure 5(B). Abbreviations of analogs and mutants are as in Table 2.

nucleobase (Table 1); their xanthine transport activity is inhibited only by xanthine analogs that are high-affinity ligands of wild-type XanQ, except for S312T which is also inhibited by 7-methylxanthine and 8-methylxanthine (Table 2). In addition, the engineered quadruple mutant G27S/L116A/G191S/S312T was found to be specific for xanthine transport and display a ligand profile similar to wild-type XanQ (Figure 6(A); Tables 1 and 2). When G27S or G27S/L116A/G191S/S312T is combined with S377G, the resulting mutants are also xanthine-specific (Figure 6(A) and (C); Table 1), showing a substrate/ligand-profile indistinguishable from wild-type XanQ (Table 2). We also tested if the combination of S377G with G191S or S312T alone affects the specificity and found that G191S/S377G is indistinguishable from S377G in having a similarly broadened guanine/xanthine profile whereas S312T/S377G is xanthine-specific like wild-type XanQ (Figure 6(C); Tables 1 and 2). Thus, it appears that the effect of S377G in broadening the specificity of XanQ is suppressed by epistatic interactions with G27S and/or S312T.

We then tested whether introduction of Ser *in lieu* of Gly377 in a xanthine/guanine permease background is sufficient to establish specificity for xanthine. We engineered the G377S replacement in AncXanQ and NmXanQ and found that both AncXanQ-G377S and NmXanQ-G377S yield broad-specificity guanine/xanthine permeases mimicking the wild-type controls. NmXanQ-G377S is essentially indistinguishable from wild-type NmXanQ in either the kinetic properties for xanthine and guanine transport (Figure 7(A); Table 1) or the ligand-recognition profile (Figure 7(B); Table 2). AncXanQ-G377S is also comparable with wild type in xanthine and guanine transport kinetics but recognizes fewer xanthine and guanine analogs than wild type in ligand inhibition analysis; it recognizes guanine and most of the guanine-related ligands of AncXanQ but does not recognize 7-methylxanthine, 8-methylxanthine, adenine, 2,6-aminopurine and 8-azaguanine which are high-affinity ligands for wild-type AncXanQ (Figure 7(B) and (D); Table 2), yielding a profile that is more similar to those of NmXanQ and NmXanQ-G377S.



**Figure 7.** The G377S replacement does not restrict specificity of AncXanQ. (A) and (C). Xanthine and guanine transport rates for AncXanQ, NmXanQ, and the indicated mutants. All experimental and presentation details are as in Figure 5(A). (B) *E. coli* T184 expressing AncXanQ or AncXanQ-G377S were analyzed for initial rates of [<sup>3</sup>H]-xanthine (1  $\mu$ M) in the presence of guanine, as indicated, and compared with control cells assayed in the absence of guanine. Each data point represents the mean and standard deviation (SD) of triplicate measurements (after subtraction of values obtained with pT7-5/-BAD alone, in all cases). Data were analyzed with Prism7 to obtain  $IC_{50}$  and thereof  $K_i$  values. The guanine  $K_i$  values deduced from the above experiment (mean  $\pm$  standard error) are  $48.4 \pm 9.2 \mu$ M (AncXanQ) and  $34.8 \pm 8.3 \mu$ M (AncXanQ-G377S). (D) Comparison of the ligand inhibition profiles of AncXanQ-G377S (black), AncXanQ-2M (S27G/G377S) (gray) and NmXanQ-G377S (light gray), highlighting part of the analogs assayed. Experimental and presentation details are as in Figure 5(B). Abbreviations of analogs and mutants are as in Table 2.

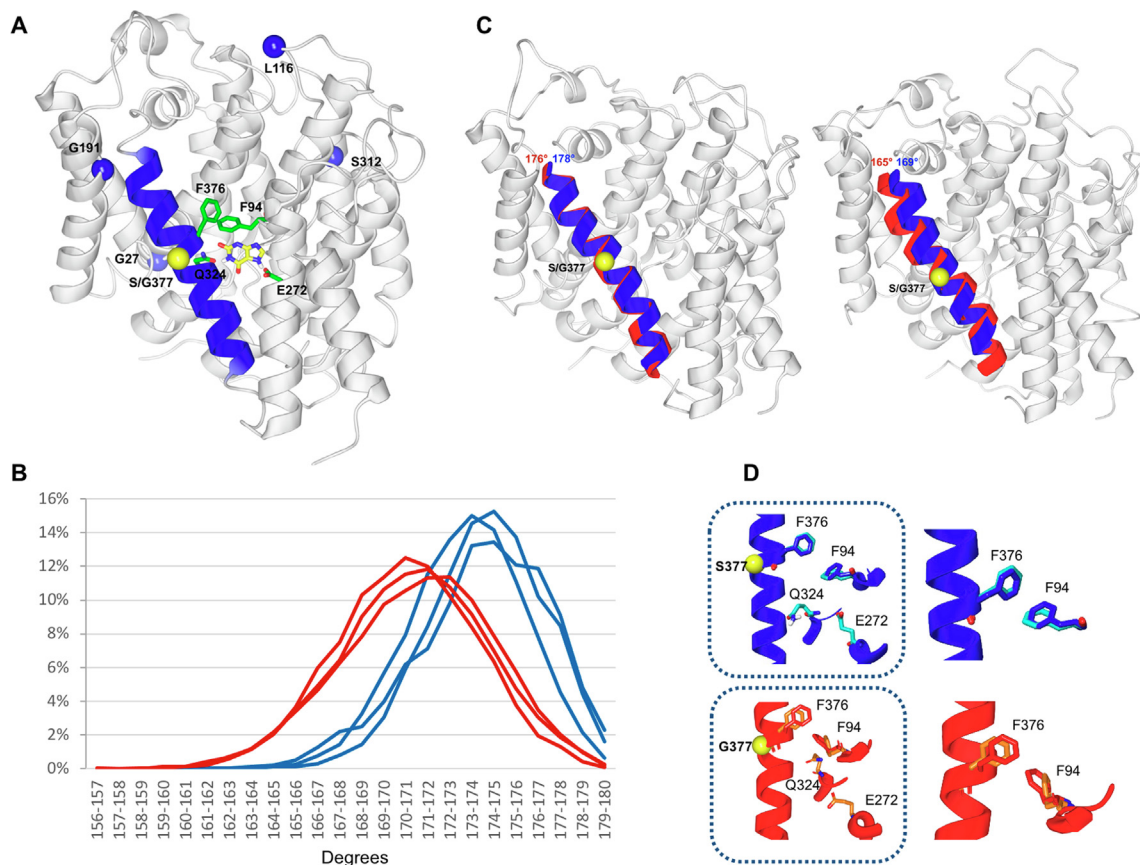
Prompted by the data that AncXanQ-G377S retains a broad-specificity guanine/xanthine profile whereas AncXanQ-5M (S27G/A116L/S191G/T312S/G377S, which includes G377S) and AncXanQ-S27G/A116L/S191G/T312S are xanthine-specific, we asked whether the xanthine-specificity shift is related to S27G or an epistatic interaction of G377S with S27G, as in the XanQ case. However, we engineered AncXanQ-S27G and AncXanQ-S27G/G377S and found that both mutants yield broad-specificity guanine/xanthine permeases with transport kinetics (Figure 7(C); Table 1) and ligand profiles (Figure 7(D); Table 2) similar to AncXanQ-G377S.

### Structural alterations in the XanQ S377G mutant

To better understand the specificity alterations of XanQ observed due to the S377G mutation, 500 ns MD simulations were performed on both the wild type and the mutated 377G-XanQ. It appears that

in the mutated structure Gly377 serves in helix tilting as the 365–377 part of TM12 helix in 377G-XanQ appears to bend in a greater percentage compared to wild type XanQ, as shown by monitoring the angle formed between the  $C\alpha$  atoms of residues 366, 377, and 384 (Supplement, Figure S8). The distribution of angles during the MD simulations shows that TM12 in 377G-XanQ bends more frequently towards 160 degrees compared to wild type and the 5th and 95th percentiles of TM12 tilting angle values are 165 and 176 in the mutant compared to 169 and 178 in wild type (Figure 8(B)). Another observation of the MD simulations is related to the interaction between F376 and F94. F94 lays in the middle of the translocation pathway and interacts directly with the ligand (Figure 8(A)) playing probably a gating role and appears to interact differently with F376 in the mutant compared to the wild type (Figure 8(D)). Specifically, the F376 phenyl group is relatively rigid in both wild type and mutant but the F94 phenyl fluctuates between





**Figure 8.** Molecular Dynamics Simulations of the effect of S377G on XanQ. (A) Overall structural model of wild-type XanQ complexed with xanthine highlighting the key binding residues Q324 and E272, the interacting pair F376-F94 and TM12 (in blue). The position of S/G377 is indicated with a yellow sphere and the positions of G27, L116, G191 and S312 are indicated with blue spheres. (B) Value distribution of TM12 tilting angle formed between the C $\alpha$  atoms of residues 366, 377 and 384 in TM12 in the wild-type XanQ (blue) and S377G XanQ (red), represented as population percentages per one-degree interval, as resulted from the 500 ns MD calculations after trajectory structure sampling every 60 ps (all data shown in Figure S8). Each one of the three lines of each color corresponds to the data from one of the three MD simulations for the relevant permease. (C) Structure snapshots of TM12 presenting the 95th percentile (left) and the 5th percentile (right) of TM12 tilting angle value distribution in the wild-type (blue) and S377G XanQ (red). (D) Interaction between F376 (TM12) and F94 (TM3) in the binding site of wild-type XanQ (blue/cyan) and S377G XanQ (red/orange). In wild type, a T-stacking interaction is observed between the F376-F94 phenyls in the 80% (blue) of the trajectory, while a second conformation exists covering the 3% of the trajectory where no interaction is observed (cyan). In the S377G mutant, the F94 dihedral monitored leads to phenyl orientations not suitable for interaction with F376, in the 45% of the trajectory (red), or exhibiting a weak interaction, in 43% of the trajectory (orange).

different conformations leading to different interactions. In wild type XanQ, a well-preserved T-stacking interaction is observed in the 80% of the trajectory, while the second most frequent phenyl conformation which is not valid for any interaction with F376 accounts for only 3% (Figure 8(D), upper panel). In contrast, in 377G-XanQ, 88% of the obtained dihedrals correspond to a non-interacting or weakly interacting setup of the two phenyl groups (Figure 8(D), lower panel) although the distance between the two phenyl groups appears almost the same in the mutant and the wild type (Supplement, Figure S9).

Regarding the potential epistatic effect of the mutation G27S on the S377G phenotype, our MD analysis indicates that the introduction of Ser in place of Gly27 at TM1 adds two key H-bond interactions, with T332 (TM10) and His31 (TM1), on an extensive network of H bonds that runs the core domain of XanQ including interactions between residues at TM9 (Asp304), TM3 (Gln90, Ser88), TM2 (Gln75), TM1 (His31) and TM10 (Asn325, Thr321) (Supplement, Figure S10). Interactions in this network have been proposed to be crucial for the binding and transport of substrate, based on the properties of the XanQ

mutants at His31-Asn325,<sup>35,21</sup> and Gln75-Asp304.<sup>37</sup> The additional H-bonds between Ser27-Thr332 and between Ser27-His31 can affect the positioning of Phe322 and Gln324 in the binding site resulting in restoration of constraints for the binding and transport of guanine in the context of the S377G mutant.

## Discussion

The work presented here promotes our knowledge on the NAT/NCS2 purine permeases in three directions. First, it presents evidence that bacterial NAT/NCS2s which are strictly specific for xanthine may have arisen from the evolution of purine transporters with more versatile, broad substrate profile. Second, it demonstrates that, apart from transporters specializing in xanthine (and/or uric acid) and transporters specializing in guanine (and hypoxanthine and/or adenine) that belong to separate subfamilies (NAT/COG2233 and AzgA-like/COG2252, respectively), NAT/NCS2 contains also transporters specific for both xanthine and guanine, like the *N. meningitidis* XanQ. Third, it demonstrates that changes outside the binding region of XanQ can affect specificity drastically, as shown with the S377G mutant, and that these effects are sensitive to small additional changes that might have epistatic consequences on them, depending on the molecular context, as shown with G27S. The above findings were based on a line of experiments that stemmed from application of an ancestral resurrection approach on a set of closely related transporter homologs, indicating the power of such an experimental strategy for specificity-related studies in transporters.

Our results suggest that a potential ancestral form of the xanthine-specific XanQ was actually a broad-specificity purine permease for both xanthine and guanine which could recognize with high affinity a range of purine bases and analogs. From an evolutionary point of view, this implies loss of the ability to bind and transport guanine and guanine-related analogs in the lineage of the orthologs of gamma-proteobacteria. In parallel, these xanthine-specific orthologs appear to have increased efficiency for xanthine transport relative to AncXanQ, as exemplified by the orthologs from *E. coli* and *A. veronii*. On the other hand, the orthologs from beta-proteobacteria (Neisseriales) seem to have retained efficiency for both xanthine and guanine, with transport efficiency and affinity similar to AncXanQ for xanthine but lower than AncXanQ for guanine, as exemplified by the ortholog from *N. meningitidis*. Interestingly, *E. coli* and *A. veronii* retain a separate homolog, in a distantly related group of the family (the AzgA-like/COG2252 subfamily), which is specific for guanine (and hypoxanthine) uptake,<sup>51</sup> whereas Neisseriales

do not contain such a homolog (Supplement, Figure S11).

Another interesting feature is the sporadic presence of XanQ orthologs in proteobacteria. As shown in Figure 1A, the XanQ clade is confined to few proteobacterial orthologs. A survey of representative genomes of Enterobacterales shows that XanQ is absent in most of them (Supplement, Figure S12). In the *Escherichial/Shigella* group, XanQ is retained with strong sequence conservation in all strains of *E. coli* but is absent in *E. fergusonii* and all strains of *Shigella* (Supplement, Source File S4). In *Aeromonas*, XanQ is also absent in some strains including all strains of *A. salmonicida* and *A. caviae* (Supplement, Source File S4). In *Neisseria*, XanQ is absent from *N. gonorrhoeae* but well conserved in sequence in all other strains (Supplement, Source File S4). Despite the absence of *xanQ* in several related genomes, the amino acid sequence of XanQ remains relatively unchanged in all other genomes of the same phylogenetic group. It is evident that the genomes lacking *xanQ* are primarily of pathogenic strains that are host-restricted and display reduced effective population sizes relative to their versatile-host counterparts, a condition that leads to global gene losses due to neutral drift rather than selection,<sup>49</sup> applied mostly to less conserved/constrained genes. Such effects have been shown clearly with *Shigella*<sup>30</sup> and several other bacteria.<sup>6</sup> However, given that the sequence of XanQ orthologs is well conserved, the absence/loss of *xanQ* in *Shigella*, *A. salmonicida* and the other relevant species/strains may point to a more specific, functionally constrained role of XanQ.

In *E. coli* K-12, *xanQ* is part of a gene cluster related with purine catabolism<sup>72</sup> which also includes genes for transporters of hypoxanthine/guanine (GhxQ)<sup>51</sup> and of uric acid (UacT).<sup>50</sup> This genetic locus is conserved in all genomes of the *Escherichial/Shigella* group. In *Shigella*, *xanQ* (and *ghxQ*) are missing but *uacT* (in most of the strains) and the catabolic genes are conserved. The absence of XanQ can be functionally replaced by its paralog XanP which has also been shown to transport xanthine<sup>34</sup> and is present in all *Escherichial/Shigella* genomes. Similarly, the absence of GhxQ is replaceable by its conserved paralog GhxP.<sup>51</sup> A similar case is seen in *Aeromonas*, where *xanQ* (together with *ghxQ* and *uacT*) is part of a homologous purine catabolic gene cluster which is present even in the genomes lacking the corresponding transporter genes. Overall, the function of XanQ appears to be redundant in several genomes, but important in other related ones. In *E. coli* K-12 or *A. veronii* panvotica, the only other NAT/NCS2 that can be linked with transport of xanthine is the XanQ-paralog XanP, but we have found that this paralog attains very low (*E. coli* XanP)<sup>34</sup> or negligible protein levels (*A. veronii* XanP) (Lazou and Frillingos,

unpublished) in the membrane of *E. coli* host. No other nucleobase transporter from the NAT/NCS2 or the NCS1<sup>54</sup> family can complement for xanthine uptake in these two bacteria (Botou, Lazou and Frilingos, unpublished). Thus, XanQ is probably needed as a xanthine transporter at least in these microorganisms. It is also interesting to note that XanP (and GhxP/GhxQ) are absent from the genomes of Neisseriales that have (a guanine-xanthine transporting) XanQ, whereas other beta-proteobacteria (Burkholderiales) that lack XanQ retain both a XanP-related homolog (Figure S1) and a GhxP/GhxQ-related homolog in the AzgA-like/COG2252 clade (Figure S11).

On the structure-functional front, our results show for the first time a guanine/xanthine substrate profile for a bacterial NAT/NCS2 transporter, i.e. the XanQ of *N. meningitidis*. This homolog is shown to transport xanthine and guanine but also recognize with high affinity hypoxanthine and several analogs (6-SG, 6-MP, 1-MP) in addition to the ligands recognized by xanthine-specific XanQs (oxypurinol, 3-MX, 2-SX, 6-SX). This *Neisseria* XanQ combines properties of two separate transporters that are used in *E. coli* for transport of xanthine (XanQ) and guanine (GhxP). It cumulatively accepts practically all XanQ and GhxP ligands and has similar  $K_M$  values for xanthine and guanine with XanQ and GhxP, respectively, differing only in having 10-fold lower transport efficiencies (Table 1). These properties are reminiscent of but not identical to the profile of its putative evolutionary precursor AncXanQ, which has a broader range of ligands (including analogs that are not ligands for XanQ or GhxP; see Figure 2(C)) and is a more efficient transporter for guanine. It is important to point out that the specificity profile of *Neisseria* XanQ would have been impossible to hypothesize without the knowledge on AncXanQ which prompted us to search for the properties of XanQ in *N. meningitidis*. This result highlights the inadequacy of phylogenetic evidence for predicting the substrate profile of even very closely related transporter homologs if “dense” functional data are unavailable and, also, the importance of the ancestral-reconstruction approach in this respect.

From our MD studies with AncXanQ, it appears that guanine (but also hypoxanthine or adenine) is oriented differently than xanthine in the substrate binding pocket even though the key binding interactions are conferred by the same residues (compare Figures 3(A) and 4(A),(B)). This difference may explain the finding that guanine and xanthine inhibit the transport of one another with one to two orders of magnitude lower apparent affinity ( $K_i$  about 50  $\mu$ M) than the affinity values deduced from the transport kinetics assays ( $K_M < 0.5 \mu$ M for guanine and 6.5  $\mu$ M for xanthine). The same trend is seen with all guanine/xanthine permeases and mutants in the

XanQ clade (Tables 1 and 2). Regarding the binding mode of guanine, there is no previous relevant evidence from any structurally solved transporter, either in NAT/NCS2 or in any other family. Crystal structures of *E. coli* enzymes associated with guanine typically show bidentate H-bonds of N1H and C2-NH<sub>2</sub> or C2-NH<sub>2</sub> and N3 with a Glu or a backbone carboxylate, as in GuaD,<sup>58</sup> XapA<sup>14</sup> and XGPRTase,<sup>69</sup> reminiscent of the binding mode in Figure 4(A). However, enzymes that utilize both guanine and xanthine like XapA or XGPRTase bind guanine and xanthine with similar interactions and orientations, unlike the case with transporters of guanine and xanthine presented here.

All orthologs in the XanQ clade share high pairwise sequence identity (65–99%), an identical set of binding site residues, and some characteristic conserved residues that are functionally relevant, based on the mutational analysis of *E. coli* XanQ.<sup>21,37</sup> Strikingly, as well, they all share a rarely appearing Asn93 (TM3) in the binding site region which is exclusive to the XanQ clade and has been linked with the specificity preference for xanthine over uric acid, based on numerous mutagenesis studies.<sup>35,36,3,50,8</sup> All permeases and mutants of the XanQ clade analyzed in the current study retain this Asn93 residue and, independent of their particular substrate profile shown, are high-affinity transporters for xanthine but do not recognize uric acid. Thus, our new data strengthen the evidence that Asn93 is important for the specificity for xanthine. Furthermore, our data reveal Ser377 (TM12) as an additional important site controlling the xanthine specificity of XanQ transporters.

Ser377 had been changed to Cys, Ala, or Thr in the context of our Cys-scanning mutagenesis studies of XanQ,<sup>37</sup> but those replacements did not reveal any effect on specificity. S377C and S377A do not differ significantly from wild type in transport activity, kinetics, or specificity profile, whereas S377T is inactive for xanthine transport although expressed normally and stably in the membrane and S377C activity is inhibited by 65% upon derivatization of the Cys with *N*-ethylmaleimide. S377C and S377A are not inhibited by guanine and S377A has been tested and found unable to transport guanine to any significant extent whatsoever.<sup>37</sup> In the light of those findings, our current evidence that S377G enlarges the profile of XanQ, allowing transport of guanine and high-affinity binding of many guanine-related analogs (Figure 5-Table 2), points to a distinctive effect of the Gly replacement, consistent with the MD data. Gly377 in S377G is situated in the middle of transmembrane helix 12 and the tilting of the helix promoted by this Gly deregulates the interaction of Phe376 with Phe94 in the binding site (Figure 7). Phe376 and Phe94 are replaceable functionally but contribute to coordination of substrate in the binding site. Both are conserved as aromatic residues in NAT/NCS2



transporters of the NAT/COG2233 clade. Phe94 offers a stacking interaction with the ring of the nucleobase substrate and, also, a H bond from the amidic backbone nitrogen. Phe376 contributes to substrate coordination in the inward-facing conformation of the transporter. These conclusions derive from both the structural data on UapA<sup>2</sup> and UraA<sup>46,67</sup> and the MD studies on XanQ.<sup>37,8</sup> Mutations that retain the aromatic rings (F94Y, F376Y) in XanQ do not inhibit xanthine binding and transport, but replacement of one of the two phenyl rings leads to loss of 80–90% (F94C, F376C) or 40–60% of activity (F94I, F376L). F94Y, F94I<sup>36</sup> and F376L<sup>37</sup> show subtle changes in specificity, including a drop in the affinity for 3-MX with F94I, reminiscent of the loss of affinity for 3-MX seen with S377G (Table 2). In UraA and related uracil transporters, mutation of the equivalent Phe to Ala leads to enlargement of specificity towards binding and uptake of thymine.<sup>7</sup> Based on all the above, it is reasonable to assume that the Phe376-Phe94 deregulation effect (Figure 7) is linked with the enlarged specificity of the S377G mutant.

Although the specificity effect of S377G is rationalized here as an indirect effect on the binding site region, we cannot rule out the possibility that S377G might also change the dynamics of transport by uncoupling the sliding of the core (elevator) domain from substrate loading, in analogy with what was recently proposed for its fungal NAT/NCS2 homologs.<sup>19</sup> Such an effect is difficult to demonstrate with limited MD calculations but can clearly be a possibility. There are two lines of evidence that support this notion. First, S377 in XanQ is in the middle of TM12, a helix of the gate (scaffold) domain that forms part of the interface on which the core/elevator/binding-site slides, upon substrate loading in the binding site. It is reasonable to assume that, by tilting TM12 (Figure 8), S377G affects this elevator-like motion of the core domain and modulates the kinetics of transport. Several specific mutations at the elevator-scaffold interface in the fungal UapA (including mutations at TM12, at positions corresponding to M375, I379, R385 in XanQ) can broaden the specificity of this highly selective xanthine and uric acid transporter to allow recognition and/or transport of most purines or, even, uracil, and the effects were proposed to reflect changes in the dynamics of sliding of the core (elevator) domain, which includes the binding site, rather than changes in residues of the binding site *per se*.<sup>19</sup> S377G in XanQ might also affect specificity via a similar route. Second, S377G does not affect the affinity for xanthine or most xanthine analogs (based on the relevant  $K_M$  and  $K_i$  values, Tables 1 and 2), implying that it does not change the overall binding site topography, but displays high affinity for guanine and other non-canonical ligands (hypoxanthine and guanine/hypoxanthine analogs), implying that the change in the affinity for guanine and other analogs from non-

recognition at the mM level (wild-type) to low-micromolar or sub-micromolar affinity (S377G) cannot be due to a modulation of geometry/interactions in the binding site alone. On the other hand, the effects observed with ligands other than guanine were only on recognition/binding ( $K_i$  values) and not on transport (S377G does not transport the other non-canonical ligands), implying that modulation of the binding affinity and binding-site interactions is certainly a component of the S377G effect. In summary, it is probable that S377G can affect *both* the binding of guanine and other non-canonical substrates *and* the dynamics of transport. By analogy with the case proposed for UapA,<sup>19</sup> non-canonical substrates like guanine might also load to the binding site of XanQ, but due to distinct orientation and/or weaker binding might not unleash the sliding of the elevator/binding-site to the inward conformation. In the S377G mutant, uncoupling of the elevator from substrate loading may allow translocation of guanine because of the loosened or leaky sliding.

The specificity effect of S377G does not hold in all XanQ backgrounds. Coexistence of S377G with G27S or S312T or both (in mutant XanQ-5M) abolishes the specificity effect and reestablishes the xanthine-specific profile of XanQ. In contrast, coexistence with G191S does not affect the S377G phenotype. As shown in Supplementary Figure S9, the epistatic effect of G27S (TM1) might be due to reinforcement of an extensive H-bond network that stabilizes the core domain<sup>37</sup>; the additional interactions offered by G27S result in repositioning of the guanine/xanthine-binding residue Gln324 in the binding pocket, which may lead to restoration of the specificity constraints (i.e. abolishment of the binding of guanine and guanine-related analogs) in the context of the S377G mutant. Such an epistatic effect is not possible with G191S (TM6), a replacement at a gate-domain residue not participating in a similar H bond network.

The reverse mutation, G377S, was shown not to affect specificity in the broad-specificity guanine/xanthine permeases AncXanQ or NmXanQ. The same is true of the replacement S27G or the combined S27G/G377S mutant in AncXanQ. However, combination of four or five mutations including S27G or S27G/G377S in addition to A116L/S191G/T312S leads to restriction of specificity by abolishing guanine-related recognition and transport. This result is difficult to interpret but shows clearly that different epistatic interactions are possible in slightly different molecular backgrounds. In AncXanQ, for example, T312S might be more important for restricting specificity than S27G or G377S.

Thus, our mutagenesis data reveal that context-dependent epistatic interactions involving residues outside the binding site might have shaped the evolution of specificity in the bacterial XanQ clade.

Interestingly, cryptic and context-dependent epistatic interactions between spatially distinct mutations have also been suggested to be critical for the evolution of NAT/NCS2 specificity in the fungal homolog UapA and related eukaryotic transporters.<sup>40</sup> In the latter case, the potential context-dependent interactions involved specific mutations of UapA at the so-called NAT/NCS2 signature motif in TM10 and distantly located residues, especially F528 in TM14, combinations of which were found to change UapA from a xanthine and uric acid specific transporter to a transporter with a more promiscuous purine/pyrimidine profile. Thus, both bacterial and fungal permeases of the XanQ/UapA cluster appear to be genetically highly modifiable in terms of their substrate profiles and the combinations of replacements associated with such modifiability begin to provide clues on how their specificity might have evolved.

In summary, we showed that relatively few residues outside the binding site can control specificity of the purine permeases of the XanQ clade and could have been important for the evolution of xanthine-specific variants from broader-specificity transporters. Change of even one of these residues (S377G) would allow evolvability of a present-day XanQ from xanthine-specific to xanthine/guanine transporter. This amino acid difference is actually present in the xanthine/guanine-transporting XanQ from *Neisseria* genomes that lack a guanine-transporting GhxP/GhxQ homolog. The reverse change (G377S) is not sufficient to restrict the specificity of this present-day XanQ to xanthine and combination of several different changes would be needed in this case, as implied by the mutagenesis of AncXanQ. Thus, some of the other differences may represent “restrictive” mutations that minimize the odds for a stochastic loss of the guanine-transporting ability of XanQ in bacteria lacking a backup for guanine transport. Inversely, differences at the same sites, especially S27G (Figure S9) or T312S, may represent “permissive” mutations that would facilitate adoption of a new, guanine-transporting function upon occurrence of a critical additional change (S377G) in the xanthine-specific XanQ background.

## Materials and methods

### Phylogenetic analysis of the XanQ/UraA cluster of NAT/NCS2 family and elucidation of the XanQ clade

We first performed a detailed phylogenetic analysis of the homologs from Proteobacteria that belong to the XanQ/UapA cluster of the NAT/NCS2 family. As shown previously,<sup>37</sup> the NAT/NCS2 homologs falling in the XanQ/UapA cluster derive primarily from bacteria (phyla Proteobacteria, Firmicutes, Actinobacteria, Cyanobacteria, Bacteroidetes), but also archaea, protists, and fungi.

The homologs from Proteobacteria are distributed to at least 5 distinct monophyletic groups, one of which includes XanQ.<sup>37</sup> For a more detailed analysis, we selected all genomes in phylum Proteobacteria from the IGM/M database at JGI with a genome status Finished (2,795 in total, as of Jan. 2017). The genomes were classified according to class (Alpha-, Beta-, Gamma-, Delta-, Epsilon-proteobacteria). Using *E. coli* XanQ as a query we performed BLAST-p search in each one of the five classes and retrieved all homologous sequences (cutoff *E* value 1e-5). These sequences were subjected to Neighbor Joining phylogenetic analysis with MEGA7<sup>43</sup> to identify all homologs pertaining to the XanQ/UapA cluster. We then selected and pooled all homologs of the XanQ/UapA cluster from the five analyses, yielding a total of 1,652 sequences. We reduced the data size by taking homologs from only one strain per species (selecting a strain containing the maximum number of homologs for each species), which reduced the number of sequences to 442 and, at a second step, one species per genus (selecting a strain containing the maximum number of homologs for each genus), which reduced the number of sequences to 150. Both sets of sequences were subjected to Maximum Likelihood (ML) phylogenetic analysis. The results for the set of 150 sequences is shown in [Supplementary Figure S1](#).

The inferred phylogenetic tree shows that the XanQ/UapA cluster is split in two major groups, the one including the *E. coli* xanthine permease XanQ and the other the uric acid permeases UacT, SmUacT1, SmUacT2, and the xanthine/uric acid permease SmXUacT (Figure S1). The XanQ-containing group comprises four subgroups, one of which is a monophyletic group of 12 (when considering the set of homologs representing one species per genus) or 20 sequences (when considering the set representing one strain per species) including the sequence of XanQ from *E. coli* K-12. We refer to this small monophyletic group as the *XanQ clade*.

### Ancestral sequence reconstruction of the common ancestor of the XanQ clade

Our ancestral reconstruction protocol is based on the workflow implemented by Thornton.<sup>60</sup> We applied this protocol to the XanQ clade in an effort to understand the unique specificity of XanQ in the NAT/NCS2 family. The most likely amino acid sequence of the ancestral node of the XanQ clade was inferred using Maximum Likelihood statistics in MEGA7 (see Supplement, [Source File S1](#), for an account of all Posterior Probabilities for each amino acid at each position of the inferred sequence). A DNA sequence that codes for this protein sequence was designed, codon-optimized for expression in *E. coli* and synthesized commercially (Eurofins Genomics GmbH, Ebersberg, Germany). The synthesized nucleotide sequence was cloned

into plasmid vector pT7-5/-BAD (see below) using *Bam*HI and *Apal* sites that had also been engineered in the synthetic construct.

### Materials and general considerations for wet-lab experiments

[2,8-<sup>3</sup>H]-adenine (31.8 Ci mmol<sup>-1</sup>), [8-<sup>3</sup>H]-guanine (21.2 Ci mmol<sup>-1</sup>), [2,8-<sup>3</sup>H]-hypoxanthine (27.7 Ci mmol<sup>-1</sup>), [8-<sup>3</sup>H]-xanthine (22.8 Ci mmol<sup>-1</sup>), [8-<sup>14</sup>C]-uric acid (58 mCi mmol<sup>-1</sup>), [5,6-<sup>3</sup>H]-uracil (32.5 Ci mmol<sup>-1</sup>), [methyl-<sup>3</sup>H]-thymine (50.6 Ci mmol<sup>-1</sup>) and [<sup>3</sup>H]-cytosine (18.8 Ci mmol<sup>-1</sup>) were from Moravék Biochemicals (Brea, CA). Non-radioactive nucleobases and analogs were from Sigma-Aldrich (St. Louis, MO). Carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP) (Sigma-Aldrich) was prepared in dimethyl sulfoxide (DMSO). Cells were grown in Luria-Bertani broth (LB) in aerobic conditions. For all liquid media cultures, cells were grown with shaking at 220 r.p.m. at 37 °C. Oligodeoxynucleotides were synthesized from Eurofins Genomics GmbH. High-fidelity DNA polymerase was from Kapa Biosystems; restriction endonucleases, alkaline phosphatase and T4 DNA ligase from Takara Clontech; horseradish peroxidase (HRP)-conjugated streptavidin from Millipore. All reagents were of analytical grade and obtained from commercial sources.

### Bacterial strains, coding sequences and plasmids

Genomic DNA of *Neisseria meningitidis* (strain ID: 41,615 isolate W-5295) was obtained from the National Meningitis Reference Laboratory, Department of Public Health Policy, School of Public Health, University of West Attica, Athens, Greece (kindly prepared and provided by Dr. Konstantinos Kesanopoulos), and used as a template for amplification of gene *NmxanQ*. *N. meningitidis* NmXanQ is identical to NCBI Ref. Seq. **WP\_00221138**. *Aeromonas veronii* pamvotica was isolated from surface sediments of Lake Pamvotis (Ioannina, Greece). Genomic DNA of this strain (GenBank no. **NZ\_MRUJ01000001.1**) was used as a template for amplification of gene *AvpxanQ*. *AvpxanQ* is GenBank no. **OLF59922.1**. *XanQ* (UniProt no. **P67444**) was mobilized from *E. coli* T184 (*lacI<sup>+</sup>O<sup>+</sup>Z<sup>-</sup>Y<sup>-</sup>(A)*, *prsl*, *met<sup>-</sup>*, *thr<sup>-</sup>*, *recA*, *hsdR/F<sup>-</sup>*, *lacI<sup>+</sup>O<sup>+</sup>ZD<sup>118</sup>*).<sup>59</sup> The sequence of *anc-xanQ* was produced by gene synthesis (Eurofins Genomics GmbH) (see above).

The coding sequences of genes were transferred to a previously described version of plasmid vector pT7-5 including the DNA sequence of the biotin-acceptor domain (BAD) of the oxaloacetate decarboxylase from *Klebsiella pneumoniae* as an insert between the *Apal* and *Hind*III sites. This

vector is designated pT7-5/-BAD. After insertion of the NAT/NCS2 coding sequences in pT7-5/-BAD, at the appropriate orientation and frame, the resulting constructs contain the BAD sequence as a C-terminal tag of each NAT/NCS2. Following expression of these genes, the gene products are biotinylated *in vivo* during bacterial growth and allow monitoring of the protein levels in the *E. coli* membrane by western blotting.<sup>34</sup>

*E. coli* TOP10F<sup>+</sup> (Invitrogen) was used for initial propagation of recombinant plasmids to prepare plasmid DNA for sequencing. *E. coli* T184<sup>59</sup> was used for expression of the pT7-5/-BAD-borne genes from the *lacZ* promoter/operator by induction with isopropyl-β-D-1-thiogalactopyranoside (IPTG) and for analysis of the uptake of xanthine<sup>34</sup> and of uric acid.<sup>50</sup> Similarly, *E. coli* JW4025 (Keio collection; *ghxP* knock-out) was used for the analysis of the uptake of guanine<sup>51</sup> and JW3692 (Keio collection; *adeP* knock-out) was used for the analysis of the uptake of adenine and hypoxanthine, whereas JW2482 (Keio collection; *uraA* knock-out) was used for the uptake of uracil and thymine<sup>7</sup> and JW0327 (Keio collection; *codB* knock-out) for the uptake of cytosine (Danielsen *et al.*, 1992). The Keio collection single-gene knockout strains<sup>5</sup> were provided from the *E. coli* Genetic Stock Center. All *E. coli* strains used were transformed according to Inoue *et al.*<sup>33</sup>

### Molecular cloning

The coding sequences of NAT/NCS2 genes were amplified by PCR on the template of genomic DNA and transferred to pT7-5/-BAD by restriction fragment replacement between the *Bam*HI and *Apal* sites. All site-directed mutants of AncXanQ, XanQ or NmXanQ were constructed using two-stage (overlap extension) PCR.<sup>29</sup> The sequences of synthetic oligodeoxynucleotides used as PCR primers are given in Supplement, Table S2. The coding sequence of all constructs was verified by double-strand DNA sequencing (Eurofins Genomics GmbH).

### Growth of bacteria

*E. coli* Keio collection strains harboring given plasmids were grown aerobically at 37 °C in LB containing kanamycin (0.025 mg/ml) and ampicillin (0.1 mg/ml). *E. coli* T184 was grown at the same conditions except that streptomycin (0.01 mg/mL) was used instead of kanamycin. Fully grown cultures (1 mL) were diluted 10-fold, allowed to grow to mid-logarithmic phase to a cell density of OD<sub>600 nm</sub> 2.0, induced with IPTG (0.5 mM) for an additional 105 min at 37 °C, and harvested for use in transport assays or western blotting.



## Transport assays and kinetic analysis

*E. coli* T184, JW4025 or JW3692 were washed twice in KPi, 0.1 M, pH 7.5, normalized to an OD<sub>420 nm</sub> of 10.0 (corresponding to 35 µg of total protein per 50 µL) in the same buffer and assayed for transport of radiolabeled substrates. *E. coli* JW2482 were prepared in MK buffer (MES 5 mM, pH 6.5, containing KCl, 0.15 M), as described.<sup>7</sup> Before initiating the transport reaction, the cells were energized by addition of glycerol to a final concentration of 20 mM and equilibrated in the assay buffer for 3 min at 25 °C. All the transport reactions were performed at 25 °C. After termination of the reactions, samples were rapidly filtered through Whatman GF/C filters, washed twice immediately with 3 mL of ice-cold KL buffer (KPi, 0.1 M, pH 5.5, LiCl, 0.1 M) and taken for liquid scintillation counting. To determine  $K_M$  and  $V_{max}$  values for the uptake of purine nucleobases, *E. coli* T184 were assayed for active transport of [<sup>3</sup>H]-xanthine and *E. coli* JW4025 for active transport of [<sup>3</sup>H]-guanine, at 5–15 s, at a range of substrate concentrations, and data were fitted to the Michaelis-Menten equation using *Prism7*.

For competitive inhibition experiments, the uptake rate of [<sup>3</sup>H]-xanthine (1 µM) or [<sup>3</sup>H]-guanine (0.1 µM) was assayed at 5–15 s in the absence or presence of non-radiolabeled analogs. To obtain IC<sub>50</sub> values, the data were fitted to the equation  $y = B + (T - B) / (1 + 10^{((\log IC_{50} - \log x) \cdot h)})$  for sigmoidal dose–response (variable slope), using *Prism7*, where  $x$  is the concentration variable,  $y$  (the transport rate) ranges from  $T$  (top) to  $B$  (bottom) and  $h$  is the Hill coefficient. In all cases,  $h$  was close to  $-1$ , consistent with competition for a single binding site.  $K_i$  values were calculated from the IC<sub>50</sub> values, based on the equation:  $K_i = IC_{50} / [1 + (S/K_M)]$  (where  $S$  is the concentration of [<sup>3</sup>H]-xanthine or [<sup>3</sup>H]-guanine used; and  $K_M$  is the  $K_M$  value obtained for the uptake of xanthine or guanine by the relevant permease or mutant).<sup>12</sup> It should be noted that the  $K_i$  value is an affinity constant implying binding to the transporter but does not indicate whether the ligand is being transported across the membrane.

## Western blot analysis

*E. coli* cells were washed twice in Tris-HCl (0.05 M), pH 8.0, containing NaCl (0.1 M) and Na<sub>2</sub>EDTA (1 mM), supplemented with 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF) (0.2 mM), and used to prepare membrane fractions by osmotic shock, treatment with EDTA/lysozyme and sonication, as described.<sup>34</sup> Membrane fractions prepared from 10 mL cell cultures were harvested by ultracentrifugation in an Optima MAX-XP Ultracentrifuge (Beckman Coulter), normalized to a protein concentration of 100 µg per 50 µL in sample loading buffer, and subjected to SDS-PAGE (12%). After electrophore-

sis, proteins were electroblotted to a polyvinylidene difluoride membrane (Parablot PVDF; Macherey-Nagel GmbH) and the BAD-tagged proteins were probed with HRP-conjugated streptavidin which was used at a dilution of 1:50,000. Signals were developed with enhanced chemiluminescence (ECL).

## Molecular simulations

**Protein model construction.** A homology model of XanQ has already been constructed in an inward-open conformation<sup>8</sup> based on the UapA structure **5I6C**.<sup>2</sup> The XanQ model was used as a template to generate AncXanQ model (see Supplement, [Source File S2](#) and [Source File S3](#)) by Homology Modeling using Prime 2018–4 (Schrödinger, LLC, New York, NY, 2018) on Maestro platform (Maestro, version 2018–4, Schrödinger, LLC, New York, NY, 2018). The two transporters, AncXanQ and XanQ, share 76% sequence identity.

**Protein preparation.** The modeled transporter was prepared using the Protein Preparation Protocol implemented in Schrödinger suite (Schrödinger Suite 2018, Protein Preparation Wizard) which is accessible within the Maestro suite (Maestro, version 2018–4, Schrödinger, LLC, New York, NY, 2018). The preparation procedure starts by adding at first hydrogen atoms. Then, we performed optimization of the orientation of hydroxyl groups of Gln and Asn and of the protonation state of His guided by the maximization of hydrogen bonding. The final step was energy minimization of the transporter, using the OPLS3 force field.<sup>26</sup>

**Ligand preparation.** Ligand preparation was performed with LigPrep application (LigPrep, version 2018–4, Schrödinger, LLC, New York, NY, 2018), which consists of a series of steps that perform conversions, apply corrections to the structure, generate ionization states and tautomers, and optimize the geometries. The force field used was OPLS3.<sup>26</sup>

**Induced-fit docking (IFD).** For IFD, Schrödinger Suite protocol was used (Schrödinger Suite 2018–4 Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2018), taking into account both the side chain or backbone movements, upon ligand binding. In the first softened-potential docking step, of the protocol, 20 poses per ligand were retained. In the second step, for each docking pose, a full cycle of protein refinement was performed, with Prime 2018–4 (Prime, version 3.0, Schrödinger, LLC, New York, NY, 2018) on all residues within 5 Å of any out of the 20 ligand poses. The Prime refinement starts by performing conformational search and by

minimizing the side chains of the selected residues. After convergence to a low-energy result, an additional minimization of all selected residues (side chain and backbone) is performed with the Truncated-Newton algorithm using the OPLS3 parameter set<sup>26</sup> and a surface Generalized Born implicit solvent model. The ranking of the obtained complexes is implemented according to Prime calculated energy (molecular mechanics and solvation), and the complexes within 30 kcal/mol of the minimum energy structure are used in the last step of the process, redocking with Glide (Schrödinger, LLC, New York, NY, 2016) using standard precision, and scoring. Finally, the ligands used in the first docking step are redocked into each of the receptor structures retained from the Prime refinement step. The final ranking of the complexes is performed by a composite score which takes into account the receptor – ligand interaction energy (GlideScore) and receptor strain and solvation energies (Prime energy).

**Molecular dynamics (MD).** In order to construct the protein–ligand complex we used the CHARMM-GUI platform. Each model was inserted into a heterogeneous fully hydrated bilayer 120 Å × 120 Å × 120 Å, consisting of DPPC lipids. The membrane embedded system was solvated with TIP3P water molecules, neutralizing with counter ions, and adding 150 mM Na<sup>+</sup> and Cl<sup>-</sup>. CHARMM36m<sup>32</sup> force field was used for protein and lipids, while the ligand was prepared using Antechamber<sup>70</sup> and the general Amber force field.<sup>71</sup> The protein orientation in the membrane was calculated using the PPM server.<sup>44</sup> The assembled simulation system consisted of ~130,000 atoms. The systems were simulated using GROMACS software.<sup>1</sup> The models were energy-minimized and equilibrated to obtain stable structures. Minimization was carried out for 2000 steps with a step size of 0.001 kJ/mol applying a steepest descent followed by a conjugate gradient algorithm, and the system was equilibrated for 20 ns by gradually heating and releasing the restraints to expedite stabilization. Finally, the system was further simulated free of restraints at a constant temperature of 300 K for 100 ns, using Nose-Hoover thermostat<sup>20</sup> and Parrinello-Rahman semi-isotropic pressure coupling<sup>53</sup> and compressibility at 4.5e-5 bar-1. The van der Waals and electrostatic interactions were smoothly switched off at 1.2 nm by switching function, while long-range electrostatic interactions were calculated using the particle mesh Ewald method.<sup>15</sup> The calculations performed were of 500 ns in time length. Each calculation was produced three times using different seed numbers.

## Data Availability Statement

Data that support the findings of this study are available in Supplemental Information (Figures

S1-S12 and Tables S1-S2; and Source Files S1-S4).

## CRedit authorship contribution statement

**Ekaterini Tatsaki:** Investigation, Visualization. **Eleni Anagnostopoulou:** Investigation. **Iliana Zantza:** Investigation, Visualization. **Panayiota Lazou:** Investigation. **Emmanuel Mikros:** Supervision, Methodology, Validation. **Stathis Frillingos:** Conceptualization, Methodology, Funding acquisition, Supervision, Visualization.

## Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme “Human Resources, Development, Education and Lifelong Learning” in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432) implemented by the State Scholarships Foundation (IKY). We also acknowledge support of this work by the project “INSPIRED” (MIS-5002550), under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020). This work was also supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC Facility -ARIS- under project pr010020. We thank Dr. Grigoris Amoutzias for many helpful discussions on the bioinformatic part of the study. We thank Dr. Konstantinos Kesanopoulos (National Meningitis Reference Laboratory; Athens, Greece) for providing the genomic DNA from *Neisseria meningitidis*. E. T. was supported by the State Scholarship Foundation (IKY) scholarship. I. Z. was supported by a “Stavros S. Niarchos Foundation” scholarship.

## Author Contributions

S.F. conceived the study and designed the experiments. E.T. performed the phylogenetic analyses and the functional study of transporters and mutants thereof. E.A. performed experiments with the homolog from *Neisseria meningitidis* and mutagenesis studies. P.L. performed experiments with the homolog from *Aeromonas veronii*. E.A., P.L. and E.T. analyzed their results with E.T. and S.F.. I.Z. performed the molecular simulations and analyzed these results with E.M. S.F. wrote the manuscript with E.T., I.Z. and E.M.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2021.167329>.

Received 8 June 2021;

Accepted 19 October 2021;

Available online 25 October 2021

### Keywords:

xanthine/guanine;  
permease;  
evolution;  
*Neisseria*;  
*Escherichia coli*

### Abbreviations used:

NAT/NCS2, Nucleobase-Ascorbate Transporter/  
Nucleobase-Cation Symporter-2; CCCP, carbonyl  
cyanide m-chlorophenyl hydrazone; HX, hypoxanthine; 6-  
MP, 6-mercaptopurine; 6-SG, 6-thioguanine; 1-MG, 1-  
methylguanidine; 8-azaG, 8-azaguanine; A, adenine; 2,6-  
DAP, 2,6-diaminopurine; 7-MG, 7-methylguanidine; MD,  
molecular dynamics; HP, horseradish peroxidase

## References

- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- Alguel, Y., Amillis, S., Leung, J., Lambrinidis, G., Capaldi, S., Scull, N.J., Craven, G., Iwata, S., et al., (2016). Structure of eukaryotic purine/H<sup>+</sup> symporter UapA suggests a role for homodimerization in transport activity. *Nat. Comm.* **7**, 11336. <https://doi.org/10.1038/ncomms11336>.
- Amillis, S., Kostis, V., Pantazopoulou, A., Mikros, E., Dhallinas, G., (2011). Mutational analysis and modelling reveal functionally critical residues in transmembrane segments 1 and 3 of the UapA transporter. *J. Mol. Biol.* **411**, 567–580. <https://doi.org/10.1016/j.jmb.2011.06.024>.
- Arakawa, T., Kobayashi-Yurugi, T., Alguel, Y., Iwanari, H., Hatae, H., Iwata, M., Abe, Y., Abe, Y., et al., (2015). Crystal structure of the anion exchanger domain of human erythrocyte band 3. *Science* **350**, 680–684. <https://doi.org/10.1126/science.aaa4335>.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection 2006.0008 *Mol. Syst. Biol.* **2** <https://doi.org/10.1038/msb4100050>.
- Bolotin, E., Hershberg, R., (2016). Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. *Sci. Rep.* **6**, 35168. <https://doi.org/10.1038/srep35168>.
- Botou, M., Lazou, P., Papakostas, K., Lambrinidis, G., Evangelidis, T., Mikros, E., Frillingos, S., (2018). Insight on specificity of uracil permeases of the NAT/NCS2 family from analysis of the transporter encoded in the pyrimidine utilization operon of *Escherichia coli*. *Mol. Microbiol.* **108**, 204–219. <https://doi.org/10.1111/mmi.13931>.
- Botou, M., Yalelis, V., Lazou, P., Zantza, I., Papakostas, K., Charalambous, V., Mikros, E., Flemetakis, E., et al., (2020). Specificity profile of NAT/NCS2 purine transporters in *Sinorhizobium (Ensifer) meliloti*. *Mol. Microbiol.* **114**, 151–171. <https://doi.org/10.1111/mmi.14503>.
- Bürzle, M., Suzuki, Y., Ackermann, D., Miyazaki, H., Maeda, N., Cléménçon, B., Burrier, R., Hediger, M.A., (2013). The sodium-dependent ascorbic acid transporter family SLC23. *Mol. Aspects Med.* **34**, 436–454. <https://doi.org/10.1016/j.mam.2012.12.002>.
- Chalioitis, A., Vlastaridis, P., Ntountoumi, C., Botou, M., Yalelis, V., Lazou, P., Tatsaki, E., Mossialos, D., Frillingos, S., Amoutzias, G.D., (2018). NAT/NCS2-hound: A webserver for the detection and evolutionary classification of prokaryotic and eukaryotic nucleobase-cation symporters of the NAT/NCS2 family. *Gigascience* **7**, giy133. <https://doi.org/10.1093/gigascience/giy133>.
- Chang, B.S., Jonsson, K., Kazmi, M.A., Donoghue, M.J., Sakmar, T.P., (2002). Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489. <https://doi.org/10.1093/oxfordjournals.molbev.a004211>.
- Cheng, Y., Prusoff, W.H., (1973). Relationship between the inhibition constant ( $K_i$ ) and the concentration of inhibitor which causes 50 per cent inhibition ( $I_{50}$ ) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108. [https://doi.org/10.1016/0006-2952\(73\)90196-2](https://doi.org/10.1016/0006-2952(73)90196-2).
- Chi, X., Jin, X., Chen, Y., Lu, X., Tu, X., Li, X., Zhang, Y., Lei, J., et al., (2020). Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. *Cell Discov.* **6**, 55. <https://doi.org/10.1038/s41421-020-00193-7>.
- Dandanell, G., Szczepanowski, R.H., Kierdaszuk, B., Shugar, D., Bochtler, M., (2005). *Escherichia coli* purine nucleoside phosphorylase II, the product of the *xapA* gene. *J. Mol. Biol.* **348**, 113–125. <https://doi.org/10.1016/j.jmb.2005.02.019>.
- Darden, T., York, D., Pedersen, L., (1993). Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092. <https://doi.org/10.1063/1.464397>.
- Dean, P., Sendra, K.M., Williams, T.A., Watson, A.K., Major, P., Nakjang, S., Kozhevnikova, E., Goldberg, A.V., et al., (2018). Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. *Nat. Comm.* **9**, 1709. <https://doi.org/10.1038/s41467-018-03923-4>.
- Dhallinas, G., (2014). Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. *Front. Pharmacol.* **12**, 207. <https://doi.org/10.3389/fphar.2014.00207>.
- Dhallinas, G., (2016). Dissection of transporter function: From genetics to structure. *Trends Genet.* **32**, 576–590. <https://doi.org/10.1016/j.tig.2016.06.003>.
- Dhallinas, G., (2021). Transporter specificity: A tale of loosened elevator-sliding. *Trends Biochem. Sci.* **46**, 708–717. <https://doi.org/10.1016/j.tibs.2021.03.007>.



20. Evans, D.J., Holian, B.L., (1985). The Nose-Hoover thermostat. *J. Chem. Phys.* **83**, 4069–4074. <https://doi.org/10.1063/1.449071>.
21. Frillingos, S., (2012). Insights to the evolution of Nucleobase-Ascorbate transporters (NAT/NCS2) from the Cys-scanning analysis of xanthine permease XanQ. *Int. J. Biochem. Mol. Biol.* **3**, 250–272 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3476789>.
22. Garaeva, A.A., Slotboom, D.J., (2020). Elevator-type mechanisms of membrane transport. *Biochem. Soc. Trans.* **48**, 1227–1241. <https://doi.org/10.1042/bst20200290>.
23. Geertsma, E.R., Chang, Y.-N., Shaik, F.R., Neldner, Y., Pardon, E., Steyaert, J., Dutzler, R., (2015). Structure of a prokaryotic fumarate transporter reveals the architecture of the SLC6 family. *Nat. Struct. Mol. Biol.* **22**, 803–808. <https://doi.org/10.1038/nsmb.3091>.
24. Georgopoulou, E., Mermelekas, G., Karena, E., Frillingos, S., (2010). Purine substrate recognition by the nucleobase-ascorbate transporter motif in the YgfO xanthine permease: Asn-325 binds and Ala-323 senses substrate. *J Biol Chem* **285**, 19422–19433. <https://doi.org/10.1074/jbc.M110.120543>.
25. Goudela, S., Karatza, P., Koukaki, M., Frillingos, S., Diallinas, G., (2005). Comparative substrate recognition by bacterial and fungal purine transporters of the NAT/NCS2 family. *Mol. Membr. Biol.* **22**, 263–275. <https://doi.org/10.1080/09687860500093016>.
26. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., Wang, L., Lupyan, D., et al., (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **12**, 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>.
27. Harms, M.J., Thornton, J.W., (2010). Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366. <https://doi.org/10.1016/j.sbi.2010.03.005>.
28. Hartwich, K., Poehlein, A., Daniel, R., (2012). The purine-utilizing bacterium *Clostridium acidurici* 9a: A Genome-Guided metabolic reconsideration. *PLoS One* **7**, <https://doi.org/10.1371/journal.pone.0051662> e51662.
29. Heckman, K.L., Pease, L.R., (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* **2**, 924–932. <https://doi.org/10.1038/nprot.2007.132>.
30. Hershberg, R., Tang, H., Petrov, D.A., (2007). Reduced selection leads to accelerated gene loss in *Shigella*. *Genome Biol.* **8**, R164. <https://doi.org/10.1186/gb-2007-8-8-r164>.
31. Hochberg, G.K.A., Thornton, J.W., (2017). Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269. <https://doi.org/10.1146/annurev-biophys-070816-033631>.
32. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., MacKerell, A.D., (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73. <https://doi.org/10.1038/nmeth.4067>.
33. Inoue, H., Nojima, H., Okayama, H., (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28. [https://doi.org/10.1016/0378-1119\(90\)90336-p](https://doi.org/10.1016/0378-1119(90)90336-p).
34. Karatza, P., Frillingos, S., (2005). Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Mol. Membr. Biol.* **22**, 251–261. <https://doi.org/10.1080/09687860500092927>.
35. Karena, E., Frillingos, S., (2009). Role of intramembrane polar residues in the YgfO xanthine permease: His-31 and Asn-93 are crucial for affinity and specificity, and Asp-304 and Glu-272 are irreplaceable. *J. Biol. Chem.* **284**, 24257–24268. <https://doi.org/10.1074/jbc.M109.030734>.
36. Karena, E., Frillingos, S., (2011). The role of transmembrane segment TM3 in the xanthine permease XanQ of *Escherichia coli*. *J. Biol. Chem.* **286**, 39595–39605. <https://doi.org/10.1074/jbc.M111.299164>.
37. Karena, E., Tatsaki, E., Lambrinidis, G., Mikros, E., Frillingos, S., (2015). Analysis of conserved NCS2 motifs in the *Escherichia coli* xanthine permease XanQ. *Mol. Microbiol.* **98**, 502–517. <https://doi.org/10.1111/mmi.13138>.
38. Kosti, V., Lambrinidis, G., Myriantopoulos, V., Diallinas, G., Mikros, E., (2012). Identification of the substrate recognition and transport pathway in a eukaryotic member of the nucleobase-ascorbate transporter (NAT) family. *PLoS One* **7**, <https://doi.org/10.1371/journal.pone.0041939> e41939.
39. Kourkoulou, A., Pittis, A., Diallinas, G., (2018). Evolution of substrate specificity in the Nucleobase-Ascorbate Transporter (NAT) protein family. *Microb. Cell* **5**, 280–292. <https://doi.org/10.15698/mic2018.06.636>.
40. Kourkoulou, A., Zantza, I., Foti, K., Mikros, E., Diallinas, G., (2021). Context-dependent cryptic roles of specific residues in substrate selectivity of the UapA purine transporter. *J. Mol. Biol.* **433**, 166814. <https://doi.org/10.1016/j.jmb.2021.166814>.
41. Kryptou, A., Lambrinidis, G., Evangelidis, T., Mikros, E., Diallinas, G., (2014). Modelling, substrate docking and mutational analysis identify residues essential for function and specificity of the major fungal purine transporter AzgA. *Mol. Microbiol.* **93**, 129–145. <https://doi.org/10.1111/mmi.12646>.
42. Kryptou, A., Scazzocchio, C., Diallinas, G., (2015). Functional characterization of NAT/NCS2 proteins of *Aspergillus brasiliensis* reveals a genuine xanthine-uric acid transporter and an intrinsically misfolded polypeptide. *Fungal Genet. Biol.* **75**, 56–63. <https://doi.org/10.1016/j.fgb.2015.01.009>.
43. Kumar, S., Stecher, G., Tamura, K., (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
44. Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376. <https://doi.org/10.1093/nar/gkr703>.
45. Lee, C., Kang, H.J., von Ballmoos, C., Newstead, S., Uzdavinyis, P., Dotson, D.L., Iwata, S., Beckstein, O., et al., (2013). A two-domain elevator mechanism for sodium/proton antiport. *Nature* **501**, 573–577. <https://doi.org/10.1038/nature12484>.
46. Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., et al., (2011). Structure and mechanism of the

- uracil transporter UraA. *Nature* **472**, 243–246. <https://doi.org/10.1038/nature09885>.
47. Mansfield, T.A., Schultes, N.P., Mourad, G.S., (2009). AtAzg1 and AtAzg2 comprise a novel family of purine transporters in *Arabidopsis*. *FEBS Lett.* **583**, 481–486. <https://doi.org/10.1016/j.febslet.2008.12.048>.
48. Niopek-Witz, S., Deppe, J., Lemieux, M.J., Möhlmann, T., (2014). Biochemical characterization and structure-function relationship of two plant NCS2 proteins, the nucleobase transporters NAT3 and NAT12 from *Arabidopsis thaliana*. *BBA-Biomembr.* **1838**, 3025–3035. <https://doi.org/10.1016/j.bbamem.2014.08.013>.
49. Mira, A., Ochman, H., Moran, N.A., (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596. [https://doi.org/10.1016/s0168-9525\(01\)02447-7](https://doi.org/10.1016/s0168-9525(01)02447-7).
50. Papakostas, K., Frillingos, S., (2012). Substrate selectivity of YgfU, a uric acid transporter from *Escherichia coli*. *J. Biol. Chem.* **287**, 15684–15695. <https://doi.org/10.1074/jbc.M112.355818>.
51. Papakostas, K., Botou, M., Frillingos, S., (2013). Functional identification of the hypoxanthine/guanine transporters YjcD and YgfQ and the adenine transporters PurP and YicO of *Escherichia coli* K-12. *J. Biol. Chem.* **288**, 36827–36840. <https://doi.org/10.1074/jbc.M113.523340>.
52. Papakostas, K., Georgopoulou, E., Frillingos, S., (2008). Cysteine-scanning analysis of putative helix XII in the YgfO xanthine permease: Ile-432 and Asn-430 are important. *J. Biol. Chem.* **283**, 13666–13678. <https://doi.org/10.1074/jbc.M800261200>.
53. Parrinello, M., Rahman, A., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190. <https://doi.org/10.1063/1.328693>.
54. Patching, S.G., (2018). Recent developments in nucleobase cation symporter-1 (NCS1) family transport proteins from bacteria, archaea, fungi and plants. *J. Biosci.* **43**, 797–815. <https://doi.org/10.1007/s12038-018-9780-3>.
55. Sauer, D.B., Trebesch, N., Marden, J.J., Cocco, N., Song, J., Koide, A., Koide, S., Tajkhorshid, E., et al., (2020). Structural basis for the reaction cycle of DASS dicarboxylate transporters. *eLife* **9**, <https://doi.org/10.7554/eLife.61350> e61350.
56. Savory, F.R., Milner, D.S., Miles, D.C., Richards, T.A., (2018). Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol. Biol. Evol.* **35**, 1887–1900. <https://doi.org/10.1093/molbev/msy082>.
57. Seeger, M.A., (2018). Membrane transporter research in times of countless structures. *BBA-Biomembr.* **1860**, 804–808. <https://doi.org/10.1016/j.bbamem.2017.08.009>.
58. Shek, R., Hilaire, T., Sim, J., French, J.B., (2019). Structural determinants for substrate selectivity in guanine deaminase enzymes of the amidohydrolase superfamily. *Biochemistry* **58**, 3280–3292. <https://doi.org/10.1021/acs.biochem.9b00341>.
59. Teather, R.M., Bramhill, J., Riede, I., Wright, J.K., Furst, M., Aichele, G., Wilhelm, V., Overath, P., (1980). Lactose carrier protein of *Escherichia coli*: Construction and expression of plasmids carrying the Y gene of the lac operon. *Eur. J. Biochem.* **108**, 223–231. <https://doi.org/10.1111/j.1432-1033.1980.tb04715.x>.
60. Thornton, J.W., (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**, 366–375. <https://doi.org/10.1038/nrg1324>.
61. Thornton, J.W., Need, E., Crews, D., (2003). Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717. <https://doi.org/10.1126/science.1086185>.
62. Thurtle-Schmidt, B.H., Stroud, R.M., (2016). Structure of Bor1 supports an elevator transport mechanisms for SLC4 anion exchangers. *Proc. Natl. Acad. Sci. USA* **113**, 10542–10546. <https://doi.org/10.1073/pnas.1612603113>.
63. Yamamoto, S., Inoue, K., Murata, T., Kamigaso, S., Yasujima, T., Maeda, J., Yoshida, Y., Ohta, K., Yuasa, H., (2010). Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *J. Biol. Chem.* **285**, 6522–6531. <https://doi.org/10.1074/jbc.M109.032961>.
64. Yan, N., (2017). A glimpse of membrane transport through structures—advances in the structural biology of the GLUT glucose transporters. *J Mol Biol.* **429**, 2710–2725. <https://doi.org/10.1016/j.jmb.2017.07.009>.
65. Yang, G., Miton, C.M., Tokuriki, N., (2020). A mechanistic view of enzyme evolution. *Protein Sci.* **29**, 1724–1747. <https://doi.org/10.1002/pro.3901>.
66. Yao, X., Fan, X., Yan, N., (2020). Cryo-EM analysis of a membrane protein embedded in the liposome. *Proc. Natl. Acad. Sci. USA* **117**, 18497–18503. <https://doi.org/10.1073/pnas.2009385117>.
67. Yu, X., Yang, G., Yan, C., Baylon, J.L., Jiang, J., Fan, H., Lu, G., Hasegawa, K., et al., (2017). Dimeric structure of the uracil:proton symporter UraA provides mechanistic insights into the SLC4/23/26 transporters. *Cell Res.* **27**, 1020–1033. <https://doi.org/10.1038/cr.2017.83>.
68. Yuasa, H., Yasujima, T., Inoue, K., (2020). Current understanding of the intestinal absorption of nucleobases and analogs. *Biol. Pharm. Bull.* **43**, 1293–1300. <https://doi.org/10.1248/bpb.b20-00342>.
69. Vos, S., Parry, R.J., Burns, M.R., de Jersey, J., Martin, J.L., (1998). Structures of free and complexed forms of *Escherichia coli* xanthine-guanine phosphoribosyltransferase. *J. Mol. Biol.* **282**, 875–889. <https://doi.org/10.1006/jmbi.1998.2051>.
70. Wang, J., Wang, W., Kollman, P.A., Case, D.A., (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260. <https://doi.org/10.1016/j.jmglm.2005.12.005>.
71. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A., (2004). Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. <https://doi.org/10.1002/jcc.20035>.
72. Xi, H., Schneider, B.L., Reitzer, L., (2000). Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J. Bacteriol.* **182**, 5332–5341. <https://doi.org/10.1128/jb.182.19.5332-5341.2000>.
73. Zhu, D., Wei, Y., Yin, J., Liu, D., Ang, E.L., Zhao, H., Zhang, Y., (2020). A pathway for degradation of uracil to acetyl coenzyme A in *Bacillus megaterium*. *Appl. Environ. Microbiol.* **86**, e02837–e2919. <https://doi.org/10.1128/aem.02837-19>.

# 1 **Uracil/H<sup>+</sup> symport by the FurE transporter challenges the rocking-** 2 **bundle mechanism of transport in APC transporters**

3  
4 Iliana Zantza,<sup>1</sup> Georgia F. Papadaki,<sup>2†</sup> Stefano Raniolo,<sup>3†</sup> Yiannis Pyrris,<sup>2</sup> George  
5 Lambrinidis,<sup>1</sup> Vittorio Limongelli,<sup>3,4\*</sup> George Diallinas,<sup>2,5\*</sup> Emmanuel Mikros<sup>1,6\*</sup>

6  
7 <sup>1</sup>Department of Pharmacy, National and Kapodistrian University of Athens, Panepistimiopolis,  
8 Athens, 15771, Greece.

9 <sup>2</sup>Department of Biology, National and Kapodistrian University of Athens, Panepistimiopolis,  
10 Athens, 15781, Greece.

11 <sup>3</sup>Faculty of Biomedical Sciences, Euler Institute, Università della Svizzera italiana (USI), Lugano,  
12 6900, Switzerland

13 <sup>4</sup>Department of Pharmacy, University of Naples “Federico II”, Naples, 80131, Italy

14 <sup>5</sup>Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology,  
15 Heraklion, 70013, Greece.

16 <sup>6</sup>Athena Research and Innovation Center in Information Communication & Knowledge  
17 Technologies, Marousi, 15125, Greece.

## 18 19 **Corresponding Authors**

20 \*Vittorio Limongelli, [vittoriolimongelli@gmail.com](mailto:vittoriolimongelli@gmail.com)

21 \*George Diallinas, [diallina@biol.uoa.gr](mailto:diallina@biol.uoa.gr)

22 \*Emmanuel Mikros [mikros@pharm.uoa.gr](mailto:mikros@pharm.uoa.gr)

23  
24 † These authors contributed equally to this work.

25  
26 Key words: *Aspergillus nidulans*/NCS1/fungal/specificity/structure-function



## 32 **Abstract**

33

34 Transporters mediate the uptake of solutes, metabolites and drugs across the cell membrane. The  
35 eukaryotic FurE nucleobase/H<sup>+</sup> symporter of *Aspergillus nidulans* has been used as a model protein  
36 to address structure-function relationships in the APC transporter superfamily, members of which  
37 are characterized by the LeuT-fold and seem to operate by the so-called ‘rocking-bundle’  
38 mechanism. In this study, we reveal the binding mode, translocation and release pathway of  
39 uracil/H<sup>+</sup> by FurE, using path collective variable, funnel metadynamics and rationally designed  
40 mutational analysis. Our study reveals a step-wise, induced-fit, mechanism of ordered sequential  
41 transport of proton and uracil, which in turn suggests that the FurE symporter, and probably  
42 structurally similar transporters, functions as a multi-step gated pore, rather than employing  
43 ‘rocking’ of compact domains, as generally proposed for APC transporters. In addition, our work  
44 further supports the emerging concept that specific elements of cytosolic terminal regions of  
45 transporters might be functionally important.

46

## 47 **Introduction**

48

49 Secondary active transporters are transmembrane proteins that mediate the transport of nutrients,  
50 metabolites and drugs in or out of cells. They select and translocate their substrates using the energy  
51 provided by the electrochemical gradient of the membrane, via a mechanism that involves the  
52 symport or antiport of mostly Na<sup>+</sup>/H<sup>+</sup> cations with other solutes. Structural studies revealed that  
53 although secondary active transporters may be structurally, functionally or evolutionary distinct,  
54 they share common folds, which are related to specific protein conformational changes associated  
55 with the transport cycle. The general model for the transport mechanism is known as the  
56 ‘alternating-access model’, where the transporter accepts or releases the substrate at one side of the  
57 cell membrane by changing conformations from an *outward-open* (OO) state facing the  
58 extracellular environment to an *inward-open* (IO) state facing the cytosol.<sup>1-5</sup> Depending on the  
59 folding and specific conformational rearrangements of the transporter, three major mechanisms  
60 have been proposed, namely the rocker-switch, the rocking-bundle and the sliding-elevator.<sup>5-9</sup>

61 Important structural and functional information about the rocking-bundle mechanism, which  
62 characterizes one of two largest transporter families, the so-called Amino Acid-Polyamine-  
63 Organocation (APC) superfamily, rise from seminal studies on the bacterial transporter LeuT,  
64 specific for leucine and alanine.<sup>3,5,10</sup> LeuT adopts the 5+5 helical inverted repeat (5HIRT), formed

65 by the first 10 transmembrane helices whose structural elements and conformational changes  
66 determine substrate recognition and transport. In total, LeuT and most APC transporters possess  
67 twelve transmembrane  $\alpha$ -helical segments (TMSs), however the role of TMS11 and TMS12 is not  
68 yet clarified. The rocking-bundle model assumes that translocation of the substrate following the  
69 OO-to-IO conformational change is facilitated by the relative motion between two motifs, the so-  
70 called ‘hash’/scaffold domain (TMS3, TMS4, TMS8, TMS9) and the ‘bundle’/core domain (TMS1,  
71 TMS2, TMS6, TMS7), with TMS5 and TMS10 functioning as gates. It has been suggested that  
72 substrate binding in the OO conformation is assisted by the simultaneous binding of a positive  
73 charge ion ( $\text{Na}^+$  or  $\text{H}^+$ ), which elicits the conformational change of the protein towards the IO  
74 conformation. This mechanism of substrate translocation has been supported by studies on the  
75 eukaryotic dopamine (DAT)<sup>11</sup> and serotonin (SERT)<sup>12</sup> transporters (neurotransmitter/sodium  
76 symporter family-NSS), and a number of mostly prokaryotic transporters.<sup>13-22</sup>

77 Although all transporters conforming to the 5+5 APC structure share the same ‘bundle-hash’ fold,  
78 topological differences have been found during the transition from the OO to the IO state. LeuT  
79 and MhsT crystal structures suggest that the ‘bundle’ domain (TMS1, TMS2, TMS6, TMS7)  
80 undergoes significant conformational changes during the OO/IO transition, pivoting around the  
81 ‘hash’ domain (TMS3, TMS4, TMS8, TMS9), while there are two additional rearrangements  
82 functioning as opening-closing gates. In LeuT, specifically, the displacement of TMS1b, TMS6a  
83 acts as an extracellular gate, along with a 45-degree kink of the TMS1a followed by a local  
84 unwinding of TMS5, which functions as the intracellular gate. In contrast, the Mhp1 transporter  
85 transits from the outward- to the inward-state by rocking a mobile ‘hash’ motif around the ‘bundle’  
86 domain, which also promotes TMS10 to move towards TMS1b and TMS6a to pack the substrate in  
87 the occluded conformation. Additionally, a flexible TMS5 bending, rather than movements in  
88 TMS1a of LeuT, opens the inward facing cavity and facilitates substrate release, thus functioning  
89 as the inner gate.

90 Several fungal members of the nucleobase cation symporter 1 (NCS1) family, which are  
91 structurally related to the APC superfamily, have been extensively studied by Diallinas and co-  
92 workers, unveiling important information about regulation of expression, subcellular trafficking  
93 and turnover, transport kinetics, and substrate specificity.<sup>23-29</sup> Transporters of this family function  
94 as  $\text{H}^+$  symporters selective for uracil, cytosine, allantoin, uridine, thiamine or nicotinamide riboside  
95 and secondarily for uric acid and xanthine.<sup>23,24,30</sup> In previous studies, we have modeled several  
96 NCS1 transporters of *Aspergillus nidulans* using the prokaryotic Mhp1 benzyl-hydantoin/ $\text{Na}^+$   
97 transporter as a structural template, and assessed structure-function relationships via extensive

98 mutational analyses. From these studies, we defined the substrate binding site and revealed the  
99 important role of the cytosolic N-and C-terminal segments in regulating endocytic turnover,  
100 transport kinetics and surprisingly substrate specificity,<sup>25-29</sup> the importance of N-terminus in  
101 transporter function has been also proved in the case of hSERT.<sup>31</sup>

102 Here, we sought to describe the functional conformational changes associated with the transport  
103 activity of the most extensively studied fungal NCS1 member, namely the FurE uracil/uric  
104 acid/allantoin transporter, from the OO to the IO state. To this end, we employed metadynamics  
105 calculations<sup>32</sup> and additional mutational analyses, rationally designed to assess our *in silico*  
106 findings. Overall, we were able to characterize the large-scale conformational changes of FurE from  
107 the OO to the IO state, including several intermediate states, elucidating the role and the  
108 internalization order of both substrate (uracil) and H<sup>+</sup> (in the form of H<sub>3</sub>O<sup>+</sup>) and their binding modes,  
109 thus providing a comprehensive novel picture that challenges aspects of the rigid-domain rocking  
110 mechanism of APC transporters.

111

## 112 **Results**

113

### 114 **FurE 3D structure**

115 The FurE structure in three different conformational states, Outward Open (OO), Occluded (Occ)  
116 and Inward Open (IO), was built through homology modeling using the corresponding Mhp1 crystal  
117 structures (**Figure 1 and Figure S1**).<sup>20-22</sup> Upon visual inspection of the structures, it emerges that  
118 interactions between residues are expected to be crucial for the structure and function of the  
119 transporter. For example, R123 (TMS3) can form a salt bridge with D261 (TMS6), mimicking the  
120 interaction observed in Mhp1 between K110 (TMS3) and D229 (TMS6) (**Figure S2A**). Another  
121 important interaction is between E51 at the edge of TMS1a and K199 of TMS5 (**Figure S2B**).  
122 Interestingly, in both OO and Occ cases, the K199 side-chain amino group is situated in the position  
123 of the co-crystallized Na<sup>+</sup> cation in the Occ conformation of Mhp1 and very close to that of the  
124 second Na<sup>+</sup> (Na2) present in the equivalent structure of LeuT (**Figure S2C**). Additionally K252  
125 (TMS6), which has been shown to affect substrate specificity,<sup>28</sup> might also form a second salt bridge  
126 with E51 (**Figure S2B**). Finally, the cytoplasmically located N-terminal D28 appears to interact  
127 with K188 (TMS5) in OO, as also reported by Papadaki *et al.*,<sup>29</sup> and with R264 in Occ (**Figure**  
128 **S2A**). Apart from the aforementioned ‘static’ salt bridges, additional interactions, possibly involved  
129 in the function of the outer gate could be between the Q59, T63, S64 side chains (TMS1b) and

130 F385, S386 (TMS10) (**Figure S2D**), while hydrophobic interactions involving W39 might control  
131 the inner gate.

132

### 133 **Mutational analysis confirms the crucial role of specific residues in FurE transport function**

134 The FurE structural models highlight two salt bridges, R123-D261 and E51-K199, and a polar  
135 interaction between Q59 and S384 or S386 as crucial in the conformational transitions of FurE.  
136 Additional residues predicted to be related to conformational changes were W39, T63, S64, R193,  
137 F196, R264, N347 and F385. In order to support these predictions, we performed respective Ala  
138 substitutions. Other residues predicted to be important for transport activity, such as the interaction  
139 of D28 with K188, and the critical role of K252 in substrate binding and specificity, have been  
140 previously studied by analogous Ala substitutions.<sup>29</sup> Mutated versions of FurE, C-terminally tagged  
141 with GFP, were analyzed in an *A. nidulans* ( $\Delta 7$ ) strain that genetically lacks all major nucleobase-  
142 related transporters, as previously described<sup>28,29</sup>. **Figure 2A** (upper left panel) summarizes growth  
143 phenotypes of mutants and control strains. As expected, the positive control strain expressing wild-  
144 type FurE grows on allantoin and uric acid and is sensitive to 5-fluorouracil (5-FU), whereas the  
145 negative control strain not expressing FurE shows a N starvation growth phenotype and is resistant  
146 to 5-FU. Ala substitutions in residues predicted to form the two major salt bridges (R123-D261 and  
147 E51-K199) scored as loss-of-function mutations, reflected in abolishment or dramatic reduction of  
148 growth on allantoin or uric acid and relatively increased resistance to 5-FU, mostly evident in  
149 R123A and D261A. Similar dramatic loss of FurE transport activity was obtained in R264A and  
150 F385A mutants, while Q59A and S386A FurE versions seem to have lost their transport activity for  
151 uric acid or 5-FU, but conserved some capacity for allantoin transport. Thus, the mutational analysis  
152 confirms the essential functional role of the interactions between R123-D261, E51-K199 and Q59-  
153 S385-S386, as well as the importance of R264, which is predicted to interact with N-terminal D28.  
154 The mutational analysis also revealed an important role of W39, as its substitution led to loss of  
155 FurE-mediated uric acid and allantoin transporter, although 5-FU transport to this drug is retained.  
156 Our findings further showed that Ala substitution of T63, S64, R193 or F196 have moderate  
157 negative effects on FurE apparent activity, reflected in reduction of growth on uric acid and some  
158 increase in 5-FU resistance (e.g., F196A), whereas residues N347 and S384 proved non-essential  
159 for FurE activity.

160 Epifluorescence microscopic analysis, shown in the right panel of **Figure 2A**, confirmed that  
161 mutational disruption of the major interactions tested (R123-D261, E51-K199 and Q59-F385-S386)  
162 did not affect the normal PM localization and stability of FurE, which confirms that the associated

163 growth defects in specific mutants reflect defects in FurE transport activity *per se*, rather than an  
164 effect on protein folding or subcellular trafficking. Direct transport assays—showed that FurE-  
165 mediated radiolabeled uracil transport was abolished in the respective mutants (**Figure 2B**).  
166 Noticeably, only in the case of R264A mutant the apparent loss-of-function proved to be the result  
167 of abolishment of trafficking to the PM, due to ER-retention of FurE. In conclusion, mutations of  
168 residues proposed, via homology modeling and initial MDs, to be functionally important validated  
169 the structural models constructed.

### 171 **The binding mode of hydronium**

172 Contrastingly to Mhp1, which is a Na<sup>+</sup>-driven NCS1 symporter, all characterized fungal NCS1  
173 transporters function via proton (H<sup>+</sup>) symport. Nevertheless, proton interactions are not elucidated  
174 for none of them, including FurE. In a first step we aspire to determine possible interactions  
175 implying the proton and residues located towards the outer gate of the transporter. In this aspect we  
176 investigated the binding of a hydronium molecule (H<sub>3</sub>O<sup>+</sup>) to FurE by employing Funnel-  
177 Metadynamics (FM), developed by our group and widely used to study ligand-protein systems.<sup>33</sup>  
178 During the FM calculations the whole binding pathway was simulated and all possible binding sites  
179 were energetically evaluated (**Figure S3A**). The preferential binding site of hydronium was  
180 identified as the lowest energy state in the Binding Free Energy Surface (BFES) (**Figure 3A**) and  
181 proved to be the same site identified for Na<sup>+</sup> in both Mhp1 and LeuT sodium co-crystallized  
182 structures. This site, located at the interface of TMSs 1 and 8, involves E51 (TMS1b) and T336  
183 (TMS8), the later residue conserved also in LeuT and Mhp1 (**Figure 3B**). The structural stability  
184 of the binding complex FurE/H<sub>3</sub>O<sup>+</sup> was further assessed by a 150 ns MD simulation.

### 186 **The binding mode of uracil**

187 Next, we simulated the binding process of uracil (K<sub>m</sub> = 1 mM) to its putative binding site in FurE  
188 using FM. The putative binding site was confined between TMS1, TMS3, TMS6 and TMS8 as  
189 suggested by previous mutagenesis data, as well as structural studies on Mhp1 and other NCS1  
190 transporters.<sup>26</sup> As performed in the case of H<sub>3</sub>O<sup>+</sup>, we simulated the binding process of uracil from  
191 its fully solvated state to the binding site in the Occ state, using the uracil distance to the binding  
192 site as CV. Uracil starting structure was generated by docking calculations, which does not affect  
193 the final result since FM calculations explore all the possible binding poses. To ensure a wide area  
194 sampling around the binding site we have set a large cone section in the FM simulation (**Figure**

195 **S3B**).  $\text{H}_3\text{O}^+$  was also included at the binding mode previously identified according to the existing  
196 literature.<sup>19,22</sup>

197 Derived from the global minimum of the BFES (**Figure 3C**), the selected model of uracil binding  
198 mode (**Figure 3D**) was found to be remarkably similar to that of (5S)-5-benzylimidazolidine-2,4-  
199 dione (hydantoin analogue) in the Mhp1 crystal structure (**Figure S4**).<sup>22</sup> In more detail, we observed  
200 H-bond interactions between T254 (TMS6) and uracil C2=O, N341 (TMS8) and uracil C4=O and  
201 N3, and  $\pi$ - $\pi$  stacking interactions between W130 (TMS3) and uracil. Two additional minima were  
202 found at higher energy values that represent probable intermediate binding poses of the ligand along  
203 its path to the final binding site. Uracil appears first to interact with Q59 (TMS1b), via a bidentate  
204 interaction with N3 and C2=O, and a  $\pi$ - $\pi$  stacking with W307 (L7 loop) (**Figure 3C**). Subsequently,  
205 it moves lower in the FurE binding cavity, where it interacts with Q137 (TMS3) via a bidentate  
206 bond involving C4=O and N3 (**Figure 3D**). Finally, uracil and W130 both interacted with F385  
207 (TMS10) through  $\pi$ - $\pi$  and T-shaped stacking interactions.

### 209 **The conformational transition of FurE from OO to IO**

210 To thoroughly describe the large-scale conformational transition of FurE, and the relative order with  
211 which hydronium and uracil are transported, we employed a dimensionality reduction approach,  
212 called path collective variables (PCVs).<sup>34</sup> In this case, the aforementioned transitions can be  
213 discretized by providing a set of frames describing the required movement (see Methods). These  
214 frames include the positions of key atoms from the beginning to the end of the conformational  
215 change, allowing us to track the transition stage during the simulation and also accelerate its  
216 sampling through Metadynamics. The whole transition of FurE from OO to IO was investigated  
217 through two set of simulations, the first describing the OO-to-Occ transition and the second the  
218 Occ-to-IO. For each of them, four systems were investigated considering all possible combinations  
219 of ligand stoichiometry: i) FurE -  $\text{H}_3\text{O}^+$  - uracil (*apo*); ii) FurE +  $\text{H}_3\text{O}^+$  - uracil; iii) FurE -  $\text{H}_3\text{O}^+$  +  
220 uracil; iv) FurE +  $\text{H}_3\text{O}^+$  + uracil (the total simulation time for each metadynamics is shown in **Table**  
221 **S1**). In the simulations where hydronium and uracil are present, they occupy the binding mode  
222 previously identified. FurE structures representing the global minimum at the calculated FES were  
223 extracted and clustered. The centroid structure of the most populated cluster was selected and  
224 subjected to a 100ns standard MD simulation in order to assess its stability. The interactions  
225 between the most important residues were monitored within the extracted structures and statistics  
226 are shown in **Figure 4A**.



## 228 The OO-to-Occ transition

229 *i) Apo state (FurE - H<sub>3</sub>O<sup>+</sup> - uracil)* - The FES shows one single, wide energy minimum between  
230 OO and Occ (**Figure 4B**) indicating a relative conformational flexibility of FurE when no ligand is  
231 present, confirmed also by standard MD simulations (**Figure S5**). Notably, in the energy minimum  
232 structures the initial part of TMS10 is positioned much closer to TMS1a with respect to the starting  
233 OO state. The relative orientation between ‘hash’ and ‘bundle’ motives remained very similar to  
234 Mhp1. The most stable interactions in the *apo* state are engaged by K199-T336, E51-T336 and  
235 R123-D261, while E51-K199 and R264-D28 interact at a minor extent (**Figure 4A**). Additionally,  
236 S386 and Q59 can form a water bridge (**Figure S6**), while Q134 interacts with T336 through a  
237 water molecule (**Figure S7A**).

238 *ii) Hydronium bound (FurE + H<sub>3</sub>O<sup>+</sup> - uracil)* - When H<sub>3</sub>O<sup>+</sup> cation is bound to FurE, the FES is  
239 rather similar to that of the *apo* form, albeit the minimum is narrower indicating a reduced flexibility  
240 of the transporter and in particular of TMS10 (**Figure 4B**). This finding suggests that the presence  
241 of H<sub>3</sub>O<sup>+</sup> influences the free energy landscape, leading TMS10 in a position competent to bind the  
242 substrate. More specifically, H<sub>3</sub>O<sup>+</sup> engages in a salt bridge with E51 and H-bond with T336,  
243 disrupting the bond between K199 and E51, between E51 and T336, and the water bridge between  
244 Q134 and T336. Consequently, T336 interacts only with K199 so that the H-bond between T336  
245 and Q134 is lost, making Q134 available to interact with uracil (**Figure S7B**).

246 *iii) Uracil bound (FurE - H<sub>3</sub>O<sup>+</sup> + uracil)* - When only uracil is bound to FurE, the FES minimum  
247 was located close to the Occ state (**Figure 4B**). However, in this pose the first part of TMS10 is still  
248 relatively distant to TMS1a. Unbiased MD simulations performed on this system show that uracil  
249 is not stable in the binding pocket, leaving the binding mode after 20ns, while TMS10 fluctuated  
250 between an OO and Occ state (**Figure S5**).

251 *iv) Hydronium and uracil bound (FurE + H<sub>3</sub>O<sup>+</sup> + uracil)* - In the case both hydronium and uracil  
252 are bound, the lowest energy minimum represents the Occ state (**Figure 4B**). Comparing this FurE  
253 state to the crystallized Occ state of Mhp1, minor differences are observed in TMS5, where a tilt  
254 was noted towards the IO conformation, and in TMS3 and TMS9. A network of interactions  
255 between F385, S386, F388, L389 with Q59, V60 and W130 contributes in stabilizing TMS10 in the  
256 occluded position, with a consequent motion of TMS9, which however is not observed in the Mhp1  
257 crystal structure. Furthermore, the slight tilt of TMS5 suggests that FurE in Occ has already moved  
258 in a conformation closer to the IO state, foreshadowing a low energy barrier between the occluded  
259 and an Inward Occlude state (IOcc). Additionally, binding of uracil stabilizes Q134 through an H-

bond, in a position capable of making an H-bond network with Q59 and water molecules. S386 (TMS10) can interact with Q59 either via a water molecule or directly (**Figure S6**).

Taken together, our results provide unprecedented structural insight into the OO-to-Occ transition of FurE. In detail, it is clearly shown that the presence of hydronium stabilizes the FurE conformation competent for binding the uracil and that the binding of both hydronium and uracil is necessary to lock FurE in Occ state. Our observations agrees with experimental data concerning Mhp1, where the symported  $\text{Na}^+$  increase substrate affinity without inducing major conformation changes,<sup>19,35</sup> while in LeuT  $\text{Na}^+$  binding shifts the conformational equilibrium to the Occ state.<sup>36,37</sup> Occ state in FurE is a very stable state as demonstrated by the low RMSD values ( $\sim 1$  Å) computed for the backbone  $\text{C}\alpha$  atoms of the transporter in unbiased MD calculations. Additionally, the break of initial bonds that stabilized TMS5 in a closed position and retained a stable ‘hash-bundle’ domain orientation (K199-E51, K199-T336, E51-T336), allow the FurE structure to move towards IO state.

### The Occ-to-IO transition

*i) Hydronium and uracil bound (FurE +  $\text{H}_3\text{O}^+$  + uracil)*- When uracil and  $\text{H}_3\text{O}^+$  are both bound to FurE, the transporter assumes a low energy structure that approaches the IO state, albeit not reaching it. This can be defined as the Inward Occluded (IOcc) state. Here, TMS3 is tilted, inducing TMS4 and TMS5 to assume a semi-open state. At the same time the central part of TMS8 is tilted away from TMS1a. The TMS3 motion is characterized by the break of the electrostatic interaction between D261 (TMS6) and R123 (TMS3) (**Figure 4A**), which instead interacts with T254 of the uracil binding site and uracil itself.  $\text{H}_3\text{O}^+$  in a more interior position, approaches D28 in the N-terminal loop. In addition, the E51 side chain rotates following the cation motion and this results in a more stable interaction with K252 (**Figure 4A**).

*ii) Uracil bound (FurE -  $\text{H}_3\text{O}^+$  + uracil)* – In this state, the FES shows a minimum close to the IO conformation. Such minimum is narrow, confined by a high-energy barrier (**Figure 4C**). This finding suggests that first  $\text{H}_3\text{O}^+$  unbinds FurE, then the transporter is stabilized in a close-to-IO conformation useful for uracil release. Compared to the Mhp1 inward structure, the tilt of TMS5 is more pronounced, while TMS8 is not tilted anymore leading to a maximum distance from TMS1a. Both conformational changes elicit a rearrangement of the other helices of the “hash” motif TMS3, TMS4 and TMS9 (**Figure S8**).

*iii) Apo state (FurE -  $\text{H}_3\text{O}^+$  - uracil)* - In the *apo* system, the FES shows the lowest energy minimum close to IO, in a position similar to the uracil bound state. However, here the minimum is wider, indicating a larger conformational freedom of the transporter in the *apo* state. TMS5 is rather

flexible, while TMS3 is slightly bent if compared with the Occ state. As FurE assumes the *apo* state after the release of both the ligands, such conformational freedom might be instrumental to favor the reverse transition of the transporter to the outward state. The FurE flexibility was confirmed by standard MD simulations carried out on the structure of the minimum.

*iv) Hydronium bound (FurE + H<sub>3</sub>O<sup>+</sup> - uracil)* - When only H<sub>3</sub>O<sup>+</sup> is bound to FurE, the energy minimum structure is between Occ and IO (**Figure 4C**), with TMS5 very close to the position assumed in Occ. This hints that in the absence of uracil the protein is not able to reach the IO state.

Our calculations show that when H<sub>3</sub>O<sup>+</sup> is still bound to the protein the conformation is stabilized in an intermediate state between Occ and IO. This suggests that the sequence of events includes first dislocation and dissociation of the H<sub>3</sub>O<sup>+</sup>, while uracil is needed in order to shift the Occ to the final IO state. Furthermore, this transition from Occ to IO is related to TMS3 and TMS8 tilting, a shift associated also with both H<sub>3</sub>O<sup>+</sup> and substrate interactions.

### The internalization pathway of H<sub>3</sub>O<sup>+</sup> cation

Based on our PCV calculations on the FurE Occ-to-IO transition, hydronium is the first to be released in the intracellular environment. Therefore, we investigated the unbinding of H<sub>3</sub>O<sup>+</sup> from the transporter by means of FM simulations (**Figure S3C**). Our results show that hydronium is able to move towards the intracellular region of FurE passing through different binding modes (**Figure 5A, 5B**). First, H<sub>3</sub>O<sup>+</sup> breaks the interactions with T336 to H-bond with S339, while maintaining the salt bridge with E51. This corresponds to minimum D in the FES reported in **Figure 5A**. Then, hydronium binds to a cleft created by F47, F262 and E51, corresponding to minimum C (**Figure 5A**). Afterwards, the interaction with E51 is lost and H<sub>3</sub>O<sup>+</sup> binds to D28, D26 of the cytosolic N-terminal terminus<sup>29</sup> and N347, corresponding to minimum B of the FES (**Figure 5A**). Finally, H<sub>3</sub>O<sup>+</sup> reaches the lowest energy pose A, binding to D28 and D26 (**Figure 5B**), before being fully released in the cytoplasm. The motion of E51 along with the H<sub>3</sub>O<sup>+</sup> unbinding elicit a break of E51-K199 and D28-R264 interactions (**Figure 4A**). Notably, our simulations indicate that the flexibility of the cytosolic N-terminal segment of FurE plays a major role in the release of hydronium in the cytoplasm.

### The internalization pathway of uracil

Once hydronium unbinds, uracil can be released in the intracellular environment. We investigated the unbinding of uracil from the FurE IO state investigating all the possible exiting pathways from

325 the binding pocket to the TMS5 inner gate (**Figure S3D**). The FES and the ligand energetically  
326 relevant poses are represented in **Figure 5C, 5D**, respectively. During uracil unbinding three  
327 residues, W130, E51 and W39, play a major role (**Figure 5D**). W130 keeps a vertical conformation  
328 to the z axis of the membrane, thus closing *de facto* the access to the extracellular part, while E51  
329 forms a H-bond with uracil favoring the translocation of the ligand towards the TMS5 inner gate.  
330 Finally, W39 forms  $\pi$ - $\pi$  and T-shaped stacking interactions with uracil justifying the relevance of  
331 W39 as highlighted by mutagenesis data. It should be noticed that W39 (TMS1) is part of a  
332 hydrophobic cleft consisting of F262 (TMS6), Y265 (TMS6), V343 (TMS8) and V189 (TMS5)  
333 contributing to the stability of the OO and Occ states where TMS1a, TMS6b, TMS8 and TMS5 are  
334 close, while in IO TMS8 and TMS5 move away as  $\text{H}_3\text{O}^+$  and uracil are transported intracellularly.

## 336 Discussion

337 Here we used the extensively studied at the genetic and functional level FurE protein, a eukaryotic  
338 transporter that is structurally similar to APC superfamily members, to address the mechanism of  
339 substrate/ $\text{H}^+$  symport using state-of-art free-energy calculations, named funnel-metadynamics  
340 (FM), focusing on the conformational rearrangements of the transporter structure that accompany  
341 transport catalysis. At variance with other binding molecular simulation methods, FM allows the  
342 sampling of the binding process without knowing *a priori* the binding mode of the ligand(s), and  
343 thus provides a unique and thorough classification of all possible binding modes. Importantly,  
344 rational mutational analysis validates the outcome and conclusions obtained via the theoretical FM  
345 calculations. Overall, this work reveals the operation mode and identifies the step-wise  
346 conformational changes that underlie the symport of uracil/ $\text{H}^+$  by FurE. Our principal findings are  
347 highlighted schematically in **Figure 6** and discussed in more detail below.

348 A principal novelty of this work is that it addresses proton symport, by introducing  $\text{H}_3\text{O}^+$  as a  
349 second distinct substrate. Thus, we obtained compelling evidence that during the whole process  
350  $\text{H}_3\text{O}^+$  interacts with three negatively charged residues, namely E51, D28 and D26. The initial  
351 binding location of  $\text{H}_3\text{O}^+$  (E51) was found to be exactly at the same place where  $\text{Na}^+$  is co-  
352 crystallized in the homologous prokaryotic transporter Mhp1.  $\text{H}_3\text{O}^+$  binding stabilized the rather  
353 flexible *apo* structure in an intermediate conformation between the initially constructed OO and  
354 Occ models (i.e. outward-occluded or OOcc). FurE- $\text{H}_3\text{O}^+$  interaction was found to trigger local  
355 amino acid rearrangements that permit Q134 to bind uracil, without promoting other major protein  
356 conformational changes, rather similar to what has been found in Mhp1.<sup>21,35</sup> This local dynamic

357 change elicited by cation-binding alone differs in other APC transporters, such as LeuT, dDAT,  
358 hDAT and SERT, where Na<sup>+</sup> binding favors a fully Occ conformation.<sup>11,12,37,38</sup> In FurE, only when  
359 both substrate (uracil) and H<sub>3</sub>O<sup>+</sup> are bound, the lowest energy conformation shifted towards the Occ  
360 structure, a state where both TMS10 (outer gate) and TMS5 (inner gate) are closed. Noticeably, in  
361 the FurE Occ state we detected relatively small changes in the ‘hash’ helices. More specifically, the  
362 relative motion of W130 (TMS3) interacting with uracil elicited a small bend in the last part of  
363 TMS3 assisted by a G132, and this was followed by a similar bend of the first part of TMS8.  
364 Furthermore, TMS9 followed the movement of TMS10 and induced a small shift to TMS4 (see  
365 **Figure 4**). A network of interactions between TMS10 and TMS1b residues, namely F385, L389  
366 and Q59, also contributed to the stabilization of the Occ state and play a critical role in the substrate  
367 specificity, as supported by the mutational analysis.

368 After acquiring the Occ structure, with uracil and H<sub>3</sub>O<sup>+</sup> bound, FurE assumes an intermediate  
369 structure between Occ and IO (i.e., inward-occluded or IOcc), where H<sub>3</sub>O<sup>+</sup> cation moved towards  
370 the intracellular domain. In this state, the N-terminal D28 loses the interaction with R264 in order  
371 to be engaged in the translocation of H<sub>3</sub>O<sup>+</sup>, while other critical rearrangements involved K199-E51-  
372 K252 and R123-D261 interactions. These events also trigger a relative motion of TMS3, TMS4 and  
373 the first part of TMS5, followed by a major tilt of TMS8 (see **Figure S8**). In this IOcc state, in  
374 which both uracil and H<sub>3</sub>O<sup>+</sup> are still bound, we observed an initial bending of the first part of the  
375 unleashed TMS5 (at P204), which reflects the opening an inner gate. In both LeuT<sup>39</sup> and DAT<sup>40</sup>,  
376 two sodium binding sites have been identified and related to both Occ state stabilization and  
377 substrate internalization.<sup>10</sup> In FurE, K199 side chain group is located in the same position of Na ion  
378 in Mhp1 and Na2 in LeuT, and corresponds to K158 in ApcT, which is the only proton symporter  
379 crystalized today.<sup>18,41</sup> In addition, the flexible side chain of K252 was very often located close to  
380 the LeuT Na1 site, in our simulations. Both K199 and K252 residues have been identified  
381 experimentally as crucial for substrate specific recognition and transport via their interaction with  
382 E51, triggering the necessary protein conformational alterations for transport activity (**Figure 4**). It  
383 thus seems that specific lys residues in H<sup>+</sup> symporters might alleviate the need for Na binding  
384 needed in other APC carriers.

385 From the IOcc, in order to reach the IO state from, our simulations showed that H<sub>3</sub>O<sup>+</sup> must be  
386 released first, as only in this case the FES is shifted to IO. A similar finding has been found in DAT  
387 and LeuT.<sup>39,42-46</sup> Notably, however, internalization of H<sub>3</sub>O<sup>+</sup> was accompanied by neutralization of  
388 D28 and D26 and subsequent relocation of the cytoplasmic N-terminal segment known as LID.<sup>29</sup>  
389 At this state, when uracil is ready to leave the transporter, TMS5 (the inner gate) opens, the upper

390 part of TMS3 bends, TMS8 is not tilted anymore, while TMS4 and TMS9 are relocated following  
391 the movements of TMS5, TMS8 and TMS3 (**Figure S8**). H<sub>3</sub>O<sup>+</sup> release interrupted the E51-H<sub>3</sub>O<sup>+</sup>-  
392 T336 interaction bridging the ‘hash’ and ‘bundle’ motives. This allowed the middle part of TMS8  
393 to shift away from TMS1, which in turn resulted in maximum TMS5 bending, creating sufficient  
394 space for uracil to be in contact with the intracellular medium and leave the binding site. This is  
395 possible only by breaking a network of interactions between residues TMS1a, TMS6b, TMS8 and  
396 the cytosolic N-terminal LID, a change also contributing to substrate specificity. Overall, proton  
397 release triggers concerted conformational bending in TMS3, TMS5 and TMS8, possible due to the  
398 presence of G132, P204 and G335, respectively. Importantly, these findings suggest a deviation  
399 from the rigid-body motion of the ‘hash’ motif, shown in Mhp1 by spin label experiments, or the  
400 rocking of the ‘bundle’ domain in other APCs.<sup>35</sup>

401 After substrate dissociation FurE is in the *apo* form, and the energy minimum structure is close  
402 to IO (**Figure 4**). This finding agrees with the evidence that Mhp1 has been crystalized also in the  
403 *apo* form conformation.<sup>20</sup> However, the FES minimum in the *apo* form is wide, suggesting that the  
404 transporter might assume several alternative conformations between Occ and IO. The fact that the  
405 presence of H<sub>3</sub>O<sup>+</sup> stabilizes a FurE state close to Occ (**Figure 4C**) prompts to suggest that H<sub>3</sub>O<sup>+</sup>  
406 binding might be essential also for the backward transition of FurE to OO.

407 In conclusion, we showed that H<sup>+</sup>/uracil binding and transport shape the energy landscape by  
408 eliciting induced-fit conformational changes that lead to sequential movements of specific TMS  
409 principally in the ‘hash’ domain, and less so in the ‘bundle’ domain, associated also with opening  
410 and closing of outer (TMS10) and inner (TMS5) gates. Our results infer that the ‘hash’ motif helices  
411 exhibit flexibility and tilt upon substrate binding in the OO-to-Occ conformational rearrangement,  
412 while in the Occ-to-IO TMS3, TMS5 and TMS8 exhibit local substrate binding-dependent  
413 flexibility, questioning the rigid rocking-movement of either the ‘bundle’ or the ‘hash’ motif, as  
414 proposed for LeuT or Mhp1, respectively. Thus, the unified picture emerging from this work is that  
415 the FurE symporter, and probably other homologous carriers, might function as a multi-step gated  
416 pore, rather than employing dramatic changes in rigid body compact domains. Finally, this work  
417 strongly supports the importance of the cytosolic N-terminal LID sequence for completion of  
418 substrate release in the cytoplasm, as also suggested by mutational analysis.<sup>50</sup>

419

## 420 **Materials and methods**

### 421 **Protein Model Construction**



422 Model of FurE was constructed based on homology modeling using Prime 2018-4 (Schrödinger,  
423 LLC, New York, NY, 2018) on Maestro platform (Maestro, version 2018-4, Schrödinger, LLC, New  
424 York, NY, 2018). Mhp1 was used as query in the three conformations: OO(2JLN), Occ(4D1B),  
425 IO(2X79), sharing with FurE a 35% similarity, while the sequence alignment was formulated  
426 according to previous work.<sup>29</sup> In order to correctly represent TMS9 in the case of IO as in 2X79 IO  
427 Mhp1 crystal structure a part of it was coil, we started with the OO FurE structure and using  
428 Targeted Molecular Dynamics in plumed-v2 software,<sup>51</sup> a constant force of 500000 kj/(mol\*nm<sup>2</sup>)  
429 was applied on the Ca atoms of the helices to create FurE in occluded and inward state. The constant  
430 force was gradually turned to zero and the system was further subjected to stabilization.

### 431 **System Setup**

432 In order to construct the protein-ligand complex CHARMM-GUI<sup>52</sup> platform was used. Each  
433 model was inserted into a heterogeneous fully hydrated bilayer 120 Å × 120 Å × 120 Å, consisting  
434 of YOPC, POPI lipids and ergosterol at a ratio of 40:40:20. The membrane embedded system was  
435 solvated with TIP3P water molecules. The solution contained neutralizing counter ions and 150  
436 mM Na<sup>+</sup> and 150 mM Cl<sup>-</sup>. In the case that H<sub>3</sub>O<sup>+</sup> was present, a water molecule was replaced, and  
437 the system was neutralized with Cl<sup>-</sup> counter ions. The assembled simulation system consisted of  
438 ~160,000 atoms.

### 439 **Molecular Dynamics (MD) / Metadynamics**

440 All simulations were conducted using GROMACS software, version 2019.2.<sup>53</sup> CHARMM36m<sup>54</sup>  
441 force field was chosen for protein and lipids, H<sub>3</sub>O<sup>+</sup> was provided from Bryce group<sup>55</sup> while the  
442 ligand and H<sub>3</sub>O<sup>+</sup> were prepared using Antechamber<sup>56</sup> and the general Amber force field.<sup>57</sup> The  
443 protein orientation into the membrane was calculated using the OPM database of the PPM server.<sup>58</sup>  
444 All model systems were minimized and equilibrated to obtain stable structures. Minimization was  
445 carried out for 5,000 steps with a step size of 0.001 kJ/mol applying a steepest descent followed by  
446 a conjugate gradient algorithm, and the system was equilibrated for 20ns by gradually heating and  
447 releasing the restraints to expedite stabilization. Finally, the system proceeded to further simulations  
448 free of restraints at a constant temperature of 300K using Nose-Hoover thermostat,<sup>59</sup> the pressure  
449 was kept constant at 1 bar using Parrinello-Rahman semi-isotropic pressure coupling<sup>60</sup> and  
450 compressibility at 4.5e-5 bar<sup>-1</sup>. The van der Waals and electrostatic interactions were smoothly  
451 switched off at 1.2 nm, while long-range electrostatic interactions were calculated using the particle  
452 mesh Ewald method.<sup>61</sup> All bonds were constrained using the LINCS algorithm,<sup>62</sup> allowing a time-  
453 step of 2.0 fs. The trajectories were further examined for structural stability by RMSD calculation

454 of protein Ca (up to 1.2 Å) and by visual inspection with VMD platform<sup>63</sup> thus ensuring that the  
455 thermalization did not cause any structural distortion.

456 For metadynamics<sup>32,33,34,64</sup> simulations the plumed-v2 software was used.<sup>51</sup>

### 457 **Funnel-Metadynamics for H<sub>3</sub>O<sup>+</sup> cation.**

458 a) The FurE transporter used was in outward-open (OO) conformation. Since no data are available  
459 concerning the binding site of the H<sub>3</sub>O<sup>+</sup> cation, a wide area around the equivalent one of the Na<sup>+</sup>  
460 cation in Mhp1 was circumvented by the funnel cone. The cone's starting point was T332 Ca, while  
461 the cylinder had a direction towards the extracellular waters. The funnel had a switching point  
462 between the cone and cylinder region at 4.0 nm, the amplitude of the cone was 0.27 rad, the radius  
463 of the cylinder section for the unbound region at 0.1 nm, the minimum and maximum value sampled  
464 as projection of the ligand's center of mass (COM) along the funnel axis was at 0.25 and 4.6 nm  
465 respectively, as long as, the lowest and highest value for fps.lp used to construct the funnel-shape  
466 restraint potential was at 0.00 and 4.8 nm respectively. The value for the spring constant of the  
467 funnel-shape restraint potential was set to 5000 kj/(mol\*nm<sup>2</sup>). As collective variable (CV) was  
468 selected the distance between the Ca atom of T332 and the center of mass of H<sub>3</sub>O<sup>+</sup> cation. The  
469 width of the Gaussian functions was set to 0.05 nm, the height to 2 kj/mol and the deposition stride  
470 to 500 simulation steps. The rescaling factor of the Gaussian function's height was 20 as we  
471 performed well-tempered metadynamics.

472 b) The study of binding/unbinding process of the H<sub>3</sub>O<sup>+</sup> cation in the cytoplasmic solvent was  
473 initiated by using well-tempered metadynamics with the FM method on FurE in the Occ state of  
474 FurE. Uracil was included in the system, placed at the binding site, whereas an upper wall of 20000  
475 kj/(mol\*nm<sup>2</sup>) enforced the COM in distances lower than 0.9 nm from the Cg of N341. The  
476 constructed funnel included all the possible routes that could lead the H<sub>3</sub>O<sup>+</sup> to the exit to the  
477 cytoplasm. The cone region started at Ca of S56. The direction of the funnel axis was cytoplasm-  
478 oriented passing through Asp348 Cb atom. The switching point between the cone and the cylinder  
479 region was at 3.6 nm, the amplitude of the cone section was set to 0.5 rad and the radius of the  
480 cylinder for the unbound region to 0.1 nm. The minimum and maximum value sampled as  
481 projection of the ligand's COM along the funnel axis was set to 0.29 and 4.2 nm respectively, the  
482 lowest and highest value for fps.lp used to construct the funnel-shape restraint potential was set to  
483 0.2 and 4.3 nm respectively, while, the value for the spring constant of the funnel-shape restraint  
484 potential to 7000 kj/(mol\*nm<sup>2</sup>). As CV was selected the distance between the Ca of E51 and the  
485 center of mass of H<sub>3</sub>O<sup>+</sup> cation. The width of the Gaussian functions was calculated at 0.01 nm, the

486 height was arranged at 2 kJ/mol with a rescaling factor of the Gaussian function at 20 and the  
487 deposition stride was set to 500 simulation steps.

### 488 **Funnel-Metadynamics for Uracil**

489 a) FM were performed aiming to highlight the binding mode of uracil in the binding site and the  
490 binding mechanism as it approaches the binding pocket from the extracellular. FurE was in the  
491 occluded state and the  $\text{H}_3\text{O}^+$  cation was included in the system, as in crystallographic results of  
492 other transporters, in particular Mhp1, ion and substrate co-exist in the Occ state. In detail, the  
493 funnel used, oriented from Ca atom of V323 deep in the binding area, with direction of the funnel  
494 axis to the extracellular solute. The switching point between the cone and cylinder region was set  
495 to 2.7 nm, the amplitude of the cone section to 0.37 rad, the radius of the cylinder section for the  
496 unbound region to 0.1 nm, the minimum and maximum value sampled as projection of the ligand's  
497 COM along the funnel axis to 0.2 and 3.3 nm respectively, the lowest and highest value for fps.lp  
498 used to construct the funnel-shape restraint potential to 0.05 and 3.6 nm respectively. The value for  
499 the spring constant of the funnel-shape restraint potential was 30000 kJ/(mol\*nm<sup>2</sup>). As CV was  
500 selected the distance between the Ca of N341 and the center of mass of uracil. The width of the  
501 Gaussian functions was 0.01 nm, the height was arranged at 2 kJ/mol and the deposition stride at  
502 500 simulation steps. The rescaling factor of the Gaussian function's height was 20.

503 b) The uracil internalization process was implemented using again, well-tempered metadynamics  
504 with the FM method, on FurE transporter in IO conformation containing uracil and not  $\text{H}_3\text{O}^+$ , as  
505 the latter was already proved from PCV simulations that leaves first the transporter in order to allow  
506 uracil to exit too (see Main Text). The funnel was constructed as to include all the possible exiting  
507 pathways from the binding pocket to the TMS5 outer gate. The cone restraint started at backbone  
508 C atom of PF53, while the direction of the funnel axis was cytoplasm-oriented passing through  
509 S342 O atom. The switching point between the cone and the cylinder region was set to 3.4 nm, the  
510 amplitude of the cone section to 0.49 rad and the radius of the cylinder for the unbound region to  
511 0.1 nm. The minimum and maximum value sampled as projection of the ligand's COM along the  
512 funnel axis was set to 0.21 and 4.1 nm respectively, the lowest and highest value for fps.lp used to  
513 construct the funnel-shape restraint potential was set to 0.1 and 4.2 nm respectively, while, the value  
514 for the spring constant of the funnel-shape restraint potential was set to 30000 kJ/(mol\*nm<sup>2</sup>). As  
515 CV was chosen the distance between the backbone of A50 and the center of mass of uracil. The  
516 width of the Gaussian functions was calculated at 0.01 nm, the height was arranged at 2 kJ/mol with  
517 a rescaling factor of the Gaussian function at 20 and the deposition stride was set to 500 simulation  
518 steps.

## 519 **Metadynamics Simulations with Path Collective Variable (PCV):**

520 a) OO-to-Occ path: In this case we used the C $\alpha$  atoms of the residues belonging to FurE helices  
521 involved in hash and bundle motif. This choice was found to be appropriate because the calculated  
522 FESs were well reproducible. The initial path was obtained through a carefully chosen set of frames  
523 with equally distant RMSDs, derived from a steered MD simulation where the OO FurE was biased  
524 to Occ conformation using a stable force on Ca atoms of helices. 6 frames were used to construct  
525 the path in total, while the average distance between adjacent frames was 0.13 nm. The RMSD  
526 matrix was constructed and plotted, confirming that the frames were appropriate for the  
527 calculation. The  $\lambda$  value calculated for s was equal to 200 nm<sup>2</sup>. The width of the Gaussian functions  
528 for hills deposition was 0.035 nm<sup>2</sup> based on the structure fluctuation in unbiased MD, the height  
529 was arranged at 0.5 kJ/mol and the deposition stride at 500 simulation steps. An upper wall of  
530 500000 kJ/(mol\*nm<sup>2</sup>) was set to constrain the distance from the path at a value lower than 0.06,  
531 based in unbiased MD simulations of more than 200 ns where the cv's fluctuation did not reach  
532 values higher than 0.03. If uracil is part of the system, it is constrained in the previously calculated  
533 position in the binding site with a distance restraint of 20000 kJ/(mol\*nm<sup>2</sup>) at 0.7 nm between the  
534 center of mass of the substrate and Cd atom of Q134. The same constraint was applied on the  
535 distance of H<sub>3</sub>O<sup>+</sup> cation from Cd atom of E51 at 0.45 nm.

536 b) The same rationale and method were used in the Occ-to-IO case. Here, the  $\lambda$  value for s was  
537 equal to 110 nm<sup>2</sup>, the width of the Gaussian functions for hills deposition was 0.037 nm<sup>2</sup>, the upper  
538 wall of 500000 kJ/(mol\*nm<sup>2</sup>) was set to constrain the z at a value lower than 0.1.

## 539 **Media, strains and growth conditions**

540 Standard complete (CM) and minimal media (MM) for *A. nidulans* growth were used,  
541 supplemented with necessary auxotrophies at concentrations given in <http://www.fgsc.net>. Glucose  
542 1% (w/v) was used as carbon source. 10 mM of sodium nitrate (NO<sub>3</sub><sup>-</sup>) or 0.5 mM of uric acid (UA),  
543 xanthine (XAN) or allantoin (ALL) were used as nitrogen sources. The uracil toxic analog 5-FU  
544 was used at 100  $\mu$ M in the presence of 10 mM NO<sub>3</sub><sup>-</sup> as N source. All media and chemical reagents  
545 were obtained from Sigma-Aldrich (Life Science Chemilab SA, Hellas) or AppliChem (Bioline  
546 Scientific SA, Hellas).

547 A *AfurD::riboBAfurA::riboBAfcyB::argBAazgAAuapAAuapC::AfpYrGAcntA::riboB pabaA1*  
548 *pantoB100* mutant strain named  $\Delta 7$ , was the recipient strain in transformations with plasmids  
549 carrying FurE mutant versions, based on complementation of the pantothenic acid auxotrophy  
550 *pantoB100*.<sup>65</sup> The  $\Delta 7$  strain has an intact endogenous FurE gene transporter, but this is very little  
551 expressed under standard conditions and thus does not contribute to detectable transport of its

552 physiological substrates (UA, ALL) or to sensitivity in 5-FU<sup>26</sup>. *A. nidulans* protoplast isolation and  
553 transformation was performed as previously described.<sup>66</sup> Growth tests were performed at 37 °C for  
554 48 h, at pH 6.8.

### 555 **Standard molecular biology manipulations and plasmid construction**

556 Genomic DNA extraction from *A. nidulans* was performed as described in FGSC  
557 (<http://www.fgsc.net>). Plasmids, prepared in *Escherichia coli*, and DNA restriction or PCR  
558 fragments were purified from agarose 1% gels with the Nucleospin Plasmid Kit or Nucleospin  
559 Extract II kit, according to the manufacturer's instructions (Macherey–Nagel, Lab Supplies  
560 Scientific SA, Hellas). Standard PCR reactions were performed using KAPATaq DNA polymerase  
561 (Kapa Biosystems). PCR products used for cloning, sequencing and re-introduction by  
562 transformation in *A. nidulans* were amplified by a high-fidelity KAPA HiFi HotStart Ready Mix  
563 (Kapa Biosystems) polymerase. DNA sequences were determined by VBC-Genomics (Vienna,  
564 Austria). Site-directed mutagenesis was carried out according to the instructions accompanying the  
565 Quik-Change® Site-Directed Mutagenesis Kit (Agilent Technologies, Stratagene). The principal  
566 vector used for most *A. nidulans* mutants is a modified pGEM-T-easy vector carrying a version of  
567 the *gpdA* promoter, the *trpC* 3' termination region and the *panB* selection marker.<sup>26</sup> Mutations in  
568 FurE were constructed by oligonucleotide-directed mutagenesis or appropriate forward and reverse  
569 primers. Transformants with intact FurE alleles were identified by PCR analysis.

### 570 **Epifluorescence microscopy**

571 Samples for standard epifluorescence microscopy were prepared as previously described.<sup>67</sup>

572 In brief, sterile 35-mm l-dishes with a glass bottom (Ibidi, Germany) containing liquid MM  
573 supplemented with NaNO<sub>3</sub> and 0.1% glucose were inoculated from a spore solution and incubated  
574 for 18 h at 25 °C. The images were obtained using an inverted Zeiss Axio Observer Z1 equipped  
575 with an Axio Cam HR R3 camera. Image processing and contrast adjustment were made using the  
576 ZEN 2012 software while further processing of the TIFF files was made using Adobe Photoshop  
577 CS3 software for brightness adjustment, rotation, alignment and annotation.

### 578 **Uptake assays**

579 FurE transport activity was measured by estimating uptake rates of [<sup>3</sup>H]-uracil (40 Ci mmol<sup>-1</sup>,  
580 Moravek Biochemicals, CA, USA), as previously described.<sup>65</sup>

581 In brief, [<sup>3</sup>H]-uracil uptake was assayed in *A. nidulans* conidiospores germinating for 4 h at 37 °C,  
582 at 140 rpm, in liquid MM (pH 6.8). Initial velocities were measured on 10<sup>7</sup> conidiospores/100 µL  
583 incubated with a concentration of 0.2µM of [<sup>3</sup>H]-uracil at 37 °C. All transport assays were carried

584 out in triplicates in at least two independent experiments. Results were analyzed using the GraphPad  
585 Prism software. Standard deviation was less than 20% in all calculations.

586

## 587 **References**

- 588 1. O. Jardetzky, Simple allosteric model for membrane pumps. *Nature*. **211**, 969–970 (1966).
- 589 2. D. Drew, O. Boudker, Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev.*  
590 *Biochem.* **85**, 543–572 (2016).
- 591 3. A. Penmatsa, E. Gouaux, How LeuT shapes our understanding of the mechanisms of sodium-  
592 coupled neurotransmitter transporters. *J Physiol.* **592**, 863–869 (2014).
- 593 4. L. R. Forrest, Y.-W. Zhang, M. T. Jacobs, J. Gesmonde, L. Xie, B. H. Honig, G. Rudnick,  
594 Mechanism for alternating access in neurotransmitter transporters. *Proc Natl Acad Sci U S A.*  
595 **105**, 10338–10343 (2008).
- 596 5. K. Kazmier, D. P. Claxton, H. S. Mchaourab, Alternating access mechanisms of LeuT-fold  
597 transporters: trailblazing towards the promised energy landscapes. *Current Opinion in*  
598 *Structural Biology.* **45**, 100–108 (2017).
- 599 6. H. Krishnamurthy, C. L. Piscitelli, E. Gouaux, Unlocking the molecular secrets of sodium-  
600 coupled transporters. *Nature*. **459**, 347–355 (2009).
- 601 7. L. R. Forrest, G. Rudnick, The Rocking Bundle: A Mechanism for Ion-Coupled Solute Flux  
602 by Symmetrical Transporters. *Physiology (Bethesda)*. **24**, 377–386 (2009).
- 603 8. N. Yan, Structural Biology of the Major Facilitator Superfamily Transporters. *Annual Review*  
604 *of Biophysics.* **44**, 257–283 (2015).
- 605 9. A. A. Garaeva, D. J. Slotboom, Elevator-type mechanisms of membrane transport. *Biochem*  
606 *Soc Trans.* **48**, 1227–1241 (2020).
- 607 10. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial  
608 homologue of Na<sup>+</sup>/Cl<sup>-</sup>-dependent neurotransmitter transporters. *Nature*. **437**, 215–223  
609 (2005).



- 610 11. K. H. Wang, A. Penmatsa, E. Gouaux, Neurotransmitter and psychostimulant recognition by  
611 the dopamine transporter. *Nature*. **521**, 322–327 (2015).
- 612 12. J. A. Coleman, V. Navratna, D. Antermite, D. Yang, J. A. Bull, E. Gouaux, Chemical and  
613 structural investigation of the paroxetine-human serotonin transporter complex. *eLife*. **9**,  
614 e56427 (2020).
- 615 13. M. Quick, A. M. Abramyan, P. Wiriyasermkul, H. Weinstein, L. Shi, J. A. Javitch, The LeuT-  
616 fold neurotransmitter:sodium symporter MhsT has two substrate sites. *PNAS*. **115**, E7924–  
617 E7931 (2018).
- 618 14. S. Faham, A. Watanabe, G. M. Besserer, D. Cascio, A. Specht, B. A. Hirayama, E. M. Wright,  
619 J. Abramson, The crystal structure of a sodium galactose transporter reveals mechanistic  
620 insights into Na<sup>+</sup>/sugar symport. *Science*. **321**, 810–814 (2008).
- 621 15. S. Ressler, A. C. Terwisscha van Scheltinga, C. Vonnrhein, V. Ott, C. Ziegler, Molecular basis  
622 of transport and regulation in the Na<sup>(+)</sup>/betaine symporter BetP. *Nature*. **458**, 47–52 (2009).
- 623 16. L. Tang, L. Bai, W. Wang, T. Jiang, Crystal structure of the carnitine transporter and insights  
624 into the antiport mechanism. *Nature Structural & Molecular Biology*. **17**, 492–496 (2010).
- 625 17. Y. Fang, H. Jayaram, T. Shane, L. Kolmakova-Partensky, F. Wu, C. Williams, Y. Xiong, C.  
626 Miller, Structure of a Prokaryotic Virtual Proton Pump at 3.2 Å Resolution. *Nature*. **460**,  
627 1040–1043 (2009).
- 628 18. P. L. Shaffer, A. Goehring, A. Shankaranarayanan, E. Gouaux, Structure and mechanism of a  
629 Na<sup>+</sup> independent amino acid transporter. *Science*. **325**, 1010–1014 (2009).
- 630 19. S. Weyand, A. D. Cameron, T. Shimamura, O. Beckstein, M. S. P. Sansom, S. Iwata, P. J. F.  
631 Henderson, The alternating access mechanism of transport as observed in the sodium-  
632 hydantoin transporter Mhp1. *J Synchrotron Radiat*. **18**, 20-23, (2011).
- 633 20. T. Shimamura, S. Weyand, O. Beckstein, N. G. Rutherford, J. M. Hadden, D. Sharples, M. S.  
634 P. Sansom, S. Iwata, P. J. F. Henderson, A. D. Cameron, Molecular Basis of Alternating  
635 Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1. *Science*. **328**, 470–  
636 473 (2010).

- 637 21. S. Weyand, T. Shimamura, S. Yajima, S. Suzuki, O. Mirza, K. Krusong, E. P. Carpenter, N.  
638 G. Rutherford, J. M. Hadden, J. O'Reilly, P. Ma, M. Saidijam, S. G. Patching, R. J. Hope, H.  
639 T. Norbertczak, P. C. J. Roach, S. Iwata, P. J. F. Henderson, A. D. Cameron, Structure and  
640 Molecular Mechanism of a Nucleobase-Cation-Symport-1 Family Transporter. *Science*. **322**,  
641 709–713 (2008).
- 642 22. K. J. Simmons, S. M. Jackson, F. Brueckner, S. G. Patching, O. Beckstein, E. Ivanova, T.  
643 Geng, S. Weyand, D. Drew, J. Lanigan, D. J. Sharples, M. S. Sansom, S. Iwata, C. W.  
644 Fishwick, A. P. Johnson, A. D. Cameron, P. J. Henderson, Molecular mechanism of ligand  
645 recognition by membrane transport protein, Mhp1. *EMBO J*. **33**, 1831–1844 (2014).
- 646 23. A. Pantazopoulou, G. Diallinas, Fungal nucleobase transporters. *FEMS Microbiol Rev*. **31**,  
647 657–675 (2007).
- 648 24. Z. Hamari, S. Amillis, C. Drevet, A. Apostolaki, C. Vágvölgyi, G. Diallinas, C. Scazzocchio,  
649 Convergent evolution and orphan genes in the Fur4p-like family and characterization of a  
650 general nucleoside transporter in *Aspergillus nidulans*. *Molecular Microbiology*. **73**, 43–57  
651 (2009).
- 652 25. E. Kryptou, V. Kosti, S. Amillis, V. Myrianthopoulos, E. Mikros, G. Diallinas, Modeling,  
653 Substrate Docking, and Mutational Analysis Identify Residues Essential for the Function and  
654 Specificity of a Eukaryotic Purine-Cytosine NCS1 Transporter \*. *Journal of Biological*  
655 *Chemistry*. **287**, 36792–36803 (2012).
- 656 26. E. Kryptou, T. Evangelidis, J. Bobonis, A. A. Pittis, T. Gabaldón, C. Scazzocchio, E. Mikros,  
657 G. Diallinas, Origin, diversification and substrate specificity in the family of NCS1/FUR  
658 transporters: Origin and specificity of Fur transporters. *Molecular Microbiology*. **96**, 927–950  
659 (2015).
- 660 27. G. Sioupouli, G. Lambrinidis, E. Mikros, S. Amillis, G. Diallinas, Cryptic purine transporters  
661 in *Aspergillus nidulans* reveal the role of specific residues in the evolution of specificity in the  
662 NCS1 family. *Mol Microbiol*. **103**, 319–332 (2017).
- 663 28. G. F. Papadaki, S. Amillis, G. Diallinas, Substrate Specificity of the FurE Transporter Is  
664 Determined by Cytoplasmic Terminal Domain Interactions. *Genetics*. **207**, 1387–1400 (2017).

- 665 29. G. F. Papadaki, G. Lambrinidis, A. Zamanos, E. Mikros, G. Diallinas, Cytosolic N- and C-  
666 Termini of the *Aspergillus nidulans* FurE Transporter Contain Distinct Elements that Regulate  
667 by Long-Range Effects Function and Specificity. *Journal of Molecular Biology*. **431**, 3827–  
668 3844 (2019).
- 669 30. C. Girke, M. Daumann, S. Niopek-Witz, T. Mählmann, Nucleobase and nucleoside transport  
670 and integration into plant metabolism. *Front. Plant Sci.* **5** (2014),  
671 doi:10.3389/fpls.2014.00443.
- 672 31. S. Sucic, S. Dallinger, B. Zdrzil, R. Weissensteiner, T. N. Jørgensen, M. Holy, O. Kudlacek,  
673 S. Seidel, J. H. Cha, U. Gether, A. H. Newman, G. F. Ecker, M. Freissmuth, H. H. Sitte, The  
674 N Terminus of Monoamine Transporters Is a Lever Required for the Action of Amphetamines.  
675 *Journal of Biological Chemistry*. **285**, 10924–10938 (2010).
- 676 32. A. Laio, M. Parrinello, Escaping free-energy minima. *PNAS*. **99**, 12562–12566 (2002).
- 677 33. V. Limongelli, M. Bonomi, M. Parrinello, Funnel metadynamics as accurate binding free-  
678 energy method. *Proceedings of the National Academy of Sciences*. **110**, 6358–6363 (2013).
- 679 34. D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space. *J. Chem.*  
680 *Phys.* **126**, 054103 (2007).
- 681 35. K. Kazmier, S. Sharma, S. M. Islam, B. Roux, H. S. Mchaourab, Conformational cycle and  
682 ion-coupling mechanism of the Na<sup>+</sup>/hydantoin transporter Mhp1. *Proceedings of the National*  
683 *Academy of Sciences*. **111**, 14752–14757 (2014).
- 684 36. P. C. Gedeon, M. Indarte, C. K. Surratt, J. D. Madura, Molecular dynamics of leucine and  
685 dopamine transporter proteins in a model cell membrane lipid bilayer. *Proteins*. **78**, 797–811  
686 (2010).
- 687 37. K. Kazmier, S. Sharma, M. Quick, S. M. Islam, B. Roux, H. Weinstein, J. A. Javitch, H. S.  
688 McHaourab, Conformational dynamics of ligand-dependent alternating access in LeuT. *Nat*  
689 *Struct Mol Biol.* **21**, 472–479 (2014).
- 690 38. D. P. Claxton, M. Quick, L. Shi, F. D. de Carvalho, H. Weinstein, J. A. Javitch, H. S.  
691 McHaourab, Ion/substrate-dependent conformational dynamics of a bacterial homolog of  
692 neurotransmitter:sodium symporters. *Nat Struct Mol Biol.* **17**, 822–829 (2010).

- 693 39. L. Shi, M. Quick, Y. Zhao, H. Weinstein, J. A. Javitch, The mechanism of a  
694 neurotransmitter:sodium symporter – inward release of Na<sup>+</sup> and substrate is triggered by  
695 substrate in a second binding site. *Mol Cell*. **30**, 667–677 (2008).
- 696 40. L. Borre, T. F. Andreassen, L. Shi, H. Weinstein, U. Gether, The Second Sodium Site in the  
697 Dopamine Transporter Controls Cation Permeation and Is Regulated by Chloride. *J Biol*  
698 *Chem*. **289**, 25764–25773 (2014).
- 699 41. L. Shi, H. Weinstein Conformational rearrangements to the intracellular open states of the  
700 LeuT and ApcT transporters are modulated by common mechanisms. *Biophys J*. **99**, 103-105,  
701 (2010)
- 702 42. A. M. Razavi, G. Khelashvili, H. Weinstein, A Markov State-based Quantitative Kinetic  
703 Model of Sodium Release from the Dopamine Transporter. *Sci Rep*. **7** (2017),  
704 doi:10.1038/srep40076.
- 705 43. L. Malinauskaite, M. Quick, L. Reinhard, J. A. Lyons, H. Yano, J. A. Javitch, P. Nissen, A  
706 Mechanism for Intracellular Release of Na<sup>+</sup> by Neurotransmitter: Sodium Symporters. *Nat*  
707 *Struct Mol Biol*. **21**, 1006–1012 (2014).
- 708 44. C. Zhao, S. Y. Noskov, The role of local hydration and hydrogen-bonding dynamics in ion  
709 and solute release from ion-coupled secondary transporters. *Biochemistry*. **50**, 1848–1856  
710 (2011).
- 711 45. G. Khelashvili, N. Stanley, M. A. Sahai, J. Medina, M. V. LeVine, L. Shi, G. De Fabritiis, H.  
712 Weinstein, Spontaneous inward opening of the dopamine transporter is triggered by PIP2-  
713 regulated dynamics of the N-terminus. *ACS Chem Neurosci*. **6**, 1825–1837 (2015).
- 714 46. J. Kniazeff, L. Shi, C. J. Loland, J. A. Javitch, H. Weinstein, U. Gether, An Intracellular  
715 Interaction Network Regulates Conformational Transitions in the Dopamine Transporter. *J*  
716 *Biol Chem*. **283**, 17691–17701 (2008).
- 717 47. G. Khelashvili, S. G. Schmidt, L. Shi, J. A. Javitch, U. Gether, C. J. Loland, H. Weinstein,  
718 Conformational Dynamics on the Extracellular Side of LeuT Controlled by Na<sup>+</sup> and K<sup>+</sup> Ions  
719 and the Protonation State of Glu290. *J Biol Chem*. **291**, 19786–19799 (2016).

- 720 48. G. Khelashvili, N. Stanley, M. A. Sahai, J. Medina, M. V. LeVine, L. Shi, G. De Fabritiis, H.  
721 Weinstein, Spontaneous Inward Opening of the Dopamine Transporter Is Triggered by PIP2-  
722 Regulated Dynamics of the N-Terminus. *ACS Chem Neurosci.* **6**, 1825–1837 (2015).
- 723 49. G. Torres, R. Gainetdinov, M. Caron, Torres, G.E., Gainetdinov, R.R. & Caron, M.G. Plasma  
724 membrane monoamine transporters: structure, regulation and function. *Nat. Rev. Neurosci.* **4**,  
725 13-25. *Nature reviews. Neuroscience.* **4**, 13–25 (2003).
- 726 50. E. Mikros, G. Diallinas, Tales of tails in transporters. *Open Biol.* **9**:190083 (2019)
- 727 51. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers  
728 for an old bird. *Computer Physics Communications.* **185**, 604–613 (2014).
- 729 52. CHARMM-GUI: A web-based graphical user interface for CHARMM - Jo - 2008 - Journal  
730 of Computational Chemistry - Wiley Online Library, (available at  
731 <https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20945>).
- 732 53. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS:  
733 High performance molecular simulations through multi-level parallelism from laptops to  
734 supercomputers. *SoftwareX.* **1–2**, 19–25 (2015).
- 735 54. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D.  
736 MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered  
737 proteins. *Nat Methods.* **14**, 71–73 (2017).
- 738 55. AMBER parameter database (Bryce Group: Computational Biophysics and Drug Design -  
739 University of Manchester), (available at <http://amber.manchester.ac.uk/>).
- 740 56. L. Wang, B. J. Berne, R. A. Friesner, On achieving high accuracy and reliability in the  
741 calculation of relative protein-ligand binding affinities. *Proceedings of the National Academy*  
742 *of Sciences.* **109**, 1937–1942 (2012).
- 743 57. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of  
744 a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
- 745 58. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM  
746 web server: resources for positioning of proteins in membranes. *Nucleic Acids Research.* **40**,  
747 D370–D376 (2012).

- 748 59. D. J. Evans, B. L. Holian, The Nose–Hoover thermostat. *The Journal of Chemical Physics*.  
749 **83**, 4069–4074 (1985).
- 750 60. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular  
751 dynamics method. *Journal of Applied Physics*. **52**, 7182–7190 (1981).
- 752 61. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums  
753 in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- 754 62. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver  
755 for molecular simulations. *Journal of Computational Chemistry*. **18**, 1463–1472 (1997).
- 756 63. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of*  
757 *Molecular Graphics*. **14**, 33–38 (1996).
- 758 64. S. Raniolo, V. Limongelli, Ligand binding free-energy calculations with funnel  
759 metadynamics. *Nature Protocols*. **15**, 2837–2866 (2020).
- 760 65. E. Kryptou, G. Diallinas, Transport assays in filamentous fungi: kinetic characterization of  
761 the UapC purine transporter of *Aspergillus nidulans*. *Fungal Genet Biol.* **63**, 1–8 (2014).
- 762 66. M. Koukaki, E. Giannoutsou, A. Karagouni, G. Diallinas, A novel improved method for  
763 *Aspergillus nidulans* transformation. *J Microbiol Methods*. **55**, 687–695 (2003).
- 764 67. C. Gournas, S. Amillis, A. Vlanti, G. Diallinas, Transport-dependent endocytosis and turnover  
765 of a uric acid-xanthine permease. *Molecular Microbiology*. **75**, 246–260 (2010).

766

## 767 **Acknowledgments**

768 We would like to thank Dr. Simone Aureli for his assistance on the computational part, Dr. Paolo  
769 Conflitti for many helpful discussions.

770

## 771 **Funding**

772 This research was funded by a Stavros Niarchos Foundation research grant (KE14319).



773 IZ was supported by a Stavros Niarchos Foundations scholarship in National and Kapodistrian  
774 University of Athens and a COST action CA15135 funded a scholarship for the mobility of IZ to  
775 Università della Svizzera Italiana.

776 Work in the laboratory of GD was supported by a "Stavros Niarchos Foundation" Research Grant  
777 (KE14315) and by a Research Grant (KE18458) from the "Hellenic Foundation for Research and  
778 Innovation (HFRI)".

779 This work was also supported by computational time granted from the Greek Research &  
780 Technology Network (GRNET) in the National HPC Facility "ARIS" under project pr010020.

781

#### 782 **Author contributions:**

783 Conceptualization: GD and EM.

784 Methodology and supervision:

785 IZ carried out the calculations with help by SR and supervision by EM and VL.

786 GL has constructed the initial models together with IZ.

787 GFP and YP carried out the experiments supervised by GD.

788 Writing—original draft: IZ, EM and GD

789 Writing—review & editing: SR, VL, GFP

790

#### 791 **Competing interests:**

792 The authors declare no competing financial interests.

793

794 **Data and materials availability:** All data and materials used in the analyses are available to any  
795 researcher for purposes of reproducing or extending the analyses.

796

#### 797 **Figures and Tables**

798 **Figure 1 Model structures of FurE.** The three homology models, Outward Open (OO),  
799 Occluded (Occ), Inward Open (IO), based on corresponding Mhp1 template crystal structures are  
800 shown in side view (orientation parallel to the membrane lipids). The 'bundle' helices are colored  
801 blue, the 'hash' helices are colored cyan, the outer and inner gates are colored red, and the TMS11,  
802 and TMS12 are grey. The yellow dashed lines represent the membrane plane.

803 **Figure 2: Functional analysis of FurE mutants. (A)** Growth tests of isogenic strains expressing  
804 distinct FurE mutant versions in a  $\Delta 7$  genetic background (i.e., genetically lacking other  
805 nucleobase-related transporters), compared to a positive (FurE) and a negative ( $\Delta 7$ ) control strain

806 (for strain details see materials and methods). NO<sub>3</sub>, UA, ALL, Xan denote MM supplemented with  
807 nitrate, uric acid, allantoin or xanthine as sole N source. 5FU is MM+NO<sub>3</sub> supplemented with 5-  
808 FU. WT denotes a standard *A. nidulans* wild-type strain expressing all major nucleobase  
809 transporters. *In vivo* epifluorescence microscopy of the same strains is shown in the right panel. All  
810 FurE mutants are functionally tagged with GFP. Notice that all FurE mutant versions, except  
811 R264A, exhibit normal (i.e., wt FurE-like) plasma membrane localization and vacuolar turnover.  
812 R264A is trapped in the perinuclear ER rings, typical of misfolded versions of FurE or other  
813 transporters (for details see Materials and methods) (B) Direct uptake assays of selected FurE  
814 mutants, using 0.2 μM [3H]-radiolabeled uracil. The figure shows relative % initial uptake rates (1  
815 min) of mutants, when wild-type FurE transport is taken as 100%, performed with 107 germinated  
816 conidiospores, as described by Kryptou and Diallinas, 2014 (for details see Materials and  
817 methods)

818 **Figure 3: Uracil and H<sub>3</sub>O<sup>+</sup> internalization as simulated by Funnel Metadynamics.** (A) The  
819 BFES of H<sub>3</sub>O<sup>+</sup> binding process. Contour lines are shown every 2 kcal/mol. (B) The binding mode  
820 of H<sub>3</sub>O<sup>+</sup> cation in FurE transporter as derived from the global energy minimum in the FES. (C) The  
821 BFES of uracil binding process in FurE transporter. (D) The intermediate states (local minima in  
822 the BFES) of uracil entering FurE transporter and the binding mode in the binding site (inset) as  
823 derived from the BFES in C.

824 **Figure 4. FurE structural alterations, residue interactions and Free Energy Surface plots**  
825 **during transport conformational changes.** (A) Side chain interactions of important residues have  
826 been monitored in all structures collected in each FES global minimum and are represented as  
827 percentage over the ensemble of the structures. (B) The FESs of the OO-to-Occ transition using  
828 different stoichiometry of ligands bound to the transporter (colour code: protein in the *apo* form in  
829 blue, complexed only with H<sub>3</sub>O<sup>+</sup> in cyan, complexed with both uracil and H<sub>3</sub>O<sup>+</sup> in green, complex  
830 only with uracil in grey). Each tick in the x axis represents one unit. (C) The FESs of the Occ-to-  
831 IO transition are shown using different stoichiometry of ligands bound to the transporter. The  
832 system containing both uracil and H<sub>3</sub>O<sup>+</sup> is represented in red, the system containing only uracil is  
833 represented in brown, the *apo* form is represented in magenta, the system containing only H<sub>3</sub>O<sup>+</sup> is  
834 represented in yellow.

835 **Figure 5. The unbinding process of H<sub>3</sub>O<sup>+</sup> and uracil to the cytoplasm as simulated by Funnel**  
836 **Metadynamics.** (A) The BFES of H<sub>3</sub>O<sup>+</sup> internalization process. The separation between contours  
837 is 2 kcal/mol. (B) The binding sites of H<sub>3</sub>O<sup>+</sup> cation in FurE transporter along the internalization  
838 pathway, as derived from the low energy states in the BFES in A. In orange is represented the

839 cytosolic N-terminal LID when  $H_3O^+$  is bound in D26, D28 and N347, while in yellow when  $H_3O^+$   
840 is released in the cytoplasm. (C) The BFES of uracil internalization process. The separation between  
841 contours is 2 kcal/mol. (D) The intermediate states of uracil internalization pathway while exiting  
842 the FurE transporter, as derived from the BFES in C.

843 **Figure 6: Schematic representation of the transport mechanism.** The mobile FurE ‘hash’  
844 motif helices (TMS3, TMS4, TMS8, TMS9) and outer and inner gates (TMS10 and TMS5) are  
845 shown relative to the ‘bundle’ motif, shown as cyan background, which is considered relatively  
846 immobile during uracil and  $H_3O^+$  internalization. In the Outward Open (OO) state, FurE is in *apo*  
847 form.  $H_3O^+$  binding results in local residue rearrangement but does not cause rearrangement of the  
848 gross tertiary structure. Uracil binding induces the closing of the TMS10 outer gate (salmon) and  
849 the kink and tilt of TMS8 (green) and TMS3 (cyan), respectively, reaching the Occluded (Occ)  
850 state.  $H_3O^+$  moves toward the TMS5 inner gate (red), which slightly bends, while TMS3 and TMS8  
851 also display structural rearrangements, initiating the Inward Occluded (IOcc) state. After  $H_3O^+$  is  
852 released in the intracellular space, TMS5 bends more, while TMS8 is not tilted anymore moving  
853 away from the ‘bundle’. TMS4 and TMS9 are shifted by TMS5, TMS8 and TMS3 bending  
854 introducing the Inward Open (IO) conformation. After the release of both  $H_3O^+$  and uracil, TMS5  
855 slightly returns to the previous bend position. An inward-facing unknown return state, probably  
856 introduced by a  $H_3O^+$ , is represented in grey.

857

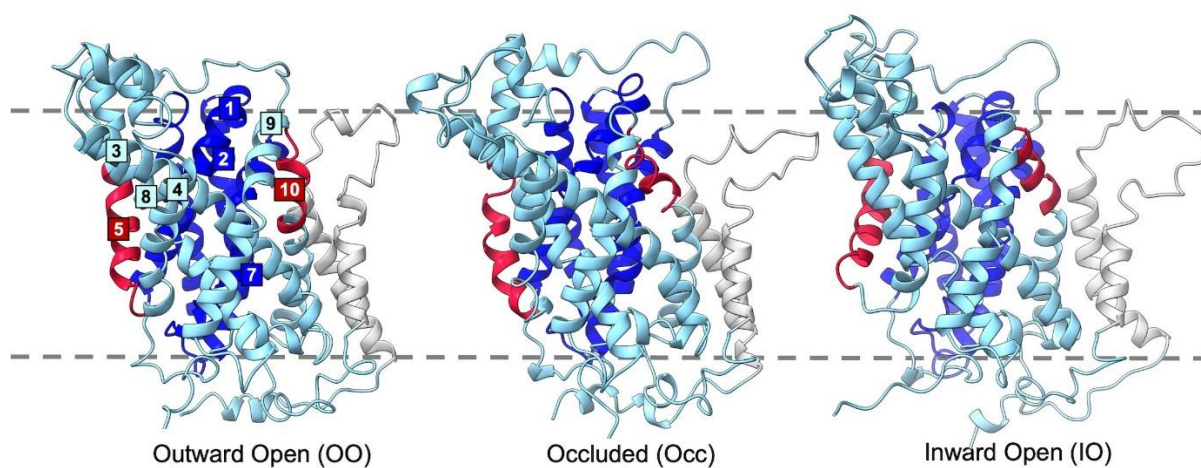
### 858 **Supplementary Materials**

859 Supporting material file contains Table S1: The simulation time of each case of the PCV  
860 Metadynamics simulations. Figures S1-S8: Alignment of FurE and Mhp1, Details of interactions  
861 between residues in FurE models, Relative orientation of transmembrane helices of the ‘hash’ motif  
862 compared to the ‘bundle’, Funnel dimensions used for the four cases of FM simulations Comparison  
863 of the binding mode of substrates in FurE and Mhp1, RMSD diagrams, Substrate-residue  
864 interactions in different intermediate conformations and conformational changes of ‘hash’ motif  
865 helices during Occ to IO transition.

866

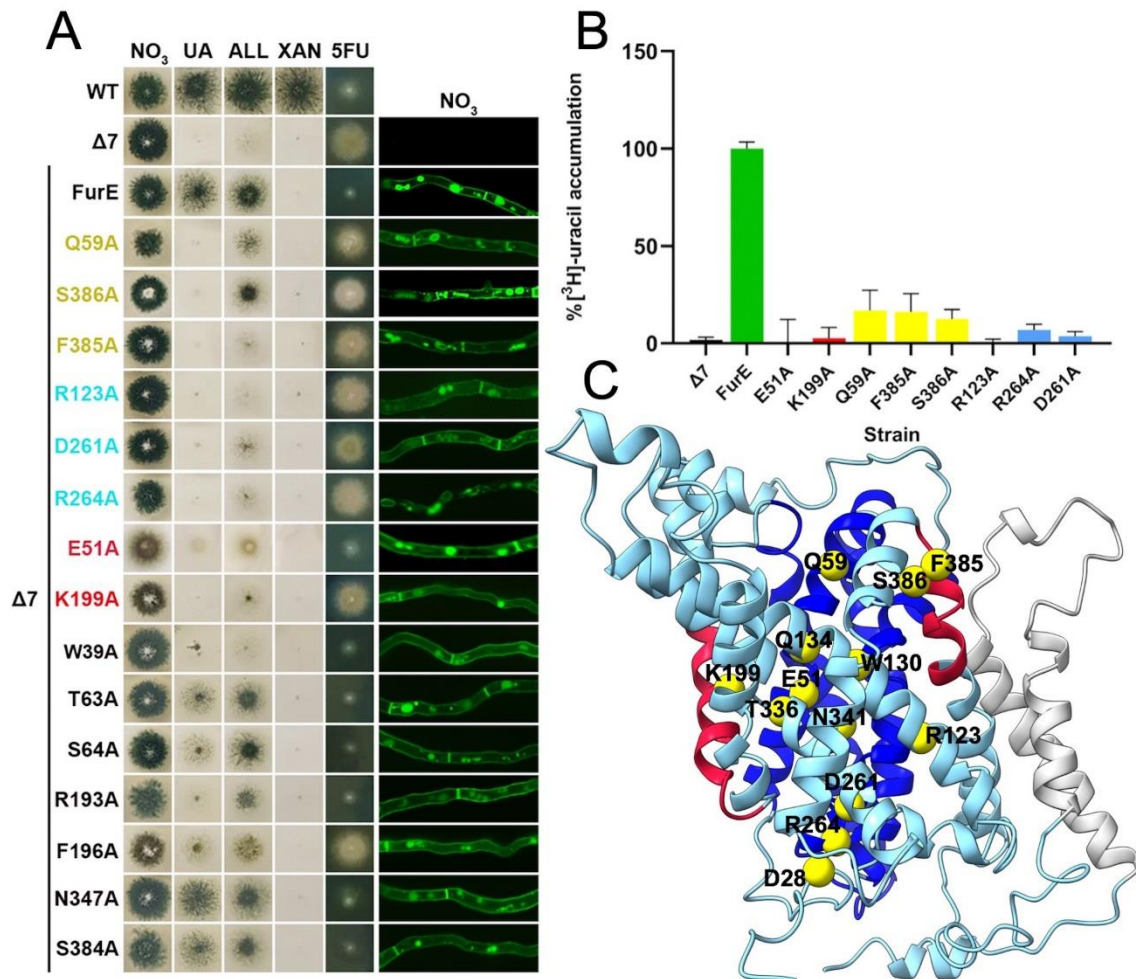
867

868



**Figure 1 Model structures of FurE.** The three homology models, Outward Open (OO), Occluded (Occ), Inward Open (IO), based on corresponding Mhp1 template crystal structures are shown in side view (orientation parallel to the membrane lipids). The ‘bundle’ helices are colored blue, the ‘hash’ helices are colored cyan, the outer and inner gates are colored red, and the TMS11, and TMS12 are grey. The yellow dashed lines represent the membrane plane.

876



877

878

879

880

881

882

883

884

885

886

887

888

889

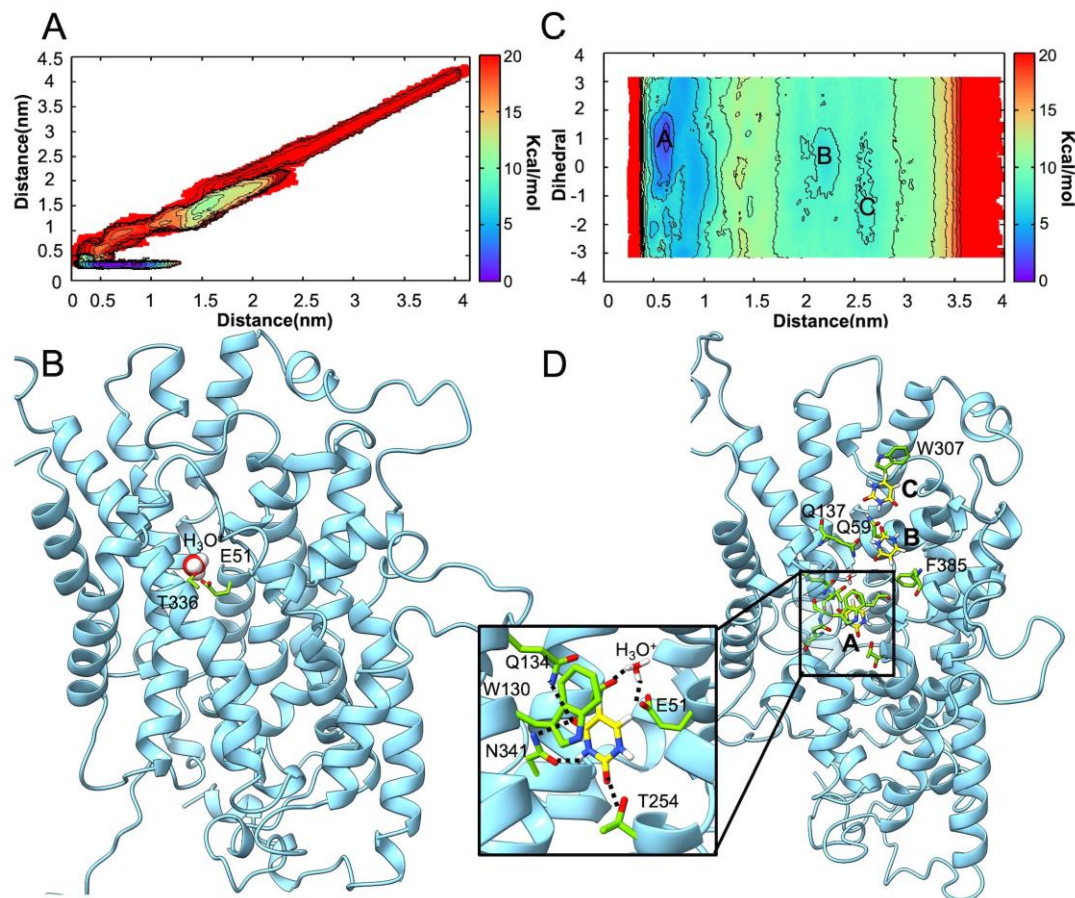
890

891

**Figure 2: Functional analysis of FurE mutants.** (A) Growth tests of isogenic strains expressing distinct FurE mutant versions in a Δ7 genetic background (i.e. genetically lacking other nucleobase-related transporters), compared to a positive (FurE) and a negative (Δ7) control strain (for strain details see materials and methods). NO<sub>3</sub>, UA, ALL, Xan denote MM supplemented with nitrate, uric acid, allantoin or xanthine as sole N source. 5FU is MM+NO<sub>3</sub> supplemented with 5-FU. WT denotes a standard *A. nidulans* wild-type strain expressing all major nucleobase transporters. *In vivo* epifluorescence microscopy of the same strains is shown in the right panel. All FurE mutants are functionally tagged with GFP. Notice that all FurE mutant versions, except R264A, exhibit normal (i.e. wt FurE-like) plasma membrane localization and vacuolar turnover. R264A is trapped in the perinuclear ER rings, typical of misfolded versions of FurE or other transporters (for details see Materials and methods) (B) Direct uptake assays of selected FurE mutants, using 0.2 μM [<sup>3</sup>H]-radiolabeled uracil. The figure shows relative % initial uptake rates (1 min) of mutants, when wild-type FurE transport is taken as 100%, performed with 107 germinated conidiospores, as described by Kryptou and Diallinas, 2014 (for details see Materials and methods)



892



893

894

895

896

897

898

899

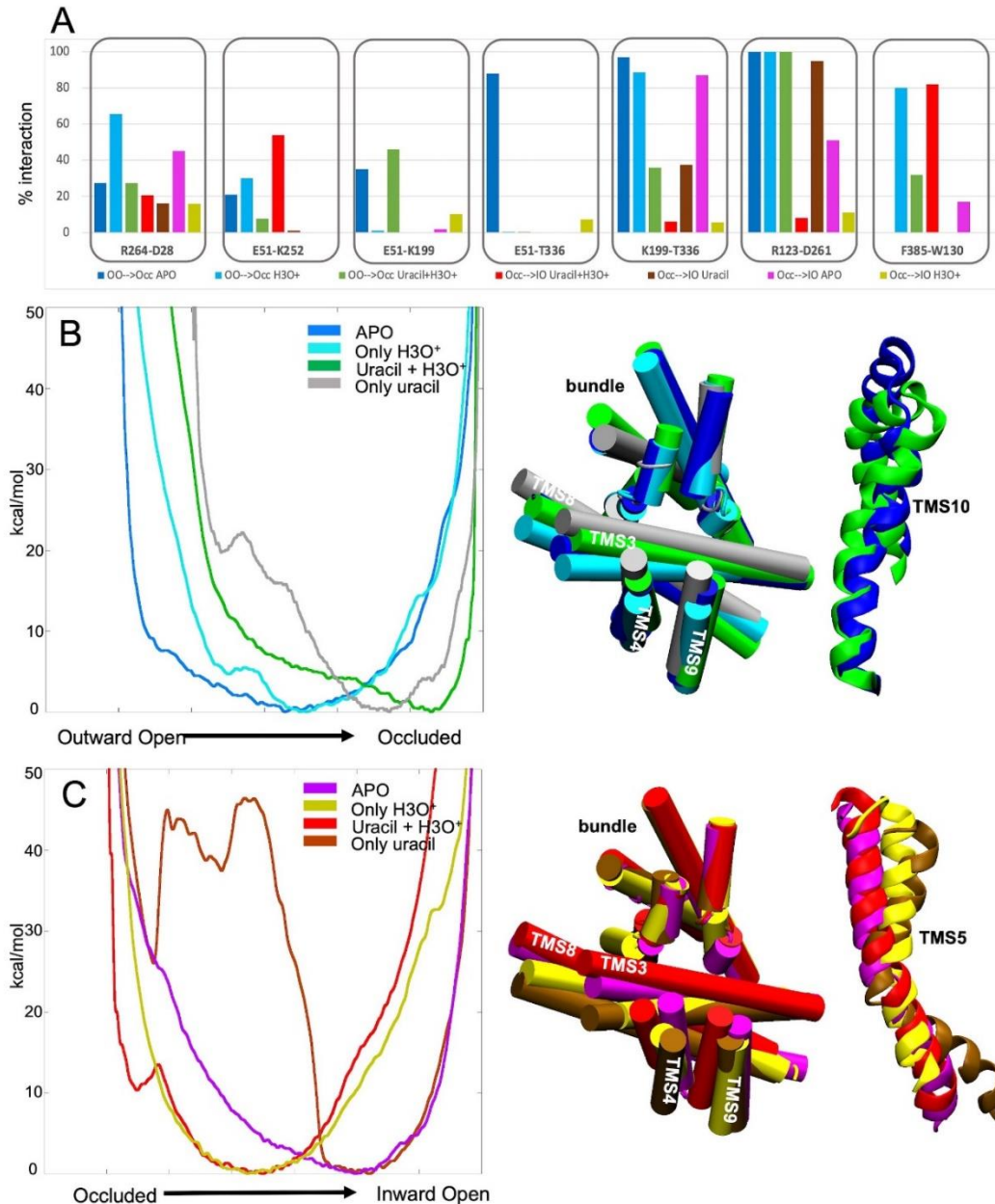
900

901

**Figure 3: Binding of H<sub>3</sub>O<sup>+</sup> and uracil as simulated by Funnel Metadynamics.** (A) The BFES of H<sub>3</sub>O<sup>+</sup> binding process. Contour lines are shown every 2 kcal/mol. (B) The binding mode of H<sub>3</sub>O<sup>+</sup> cation in FurE transporter as derived from the global energy minimum in the FES. (C) The BFES of uracil binding process in FurE transporter. (D) The intermediate states (local minima in the BFES) of uracil entering FurE transporter and the binding mode in the binding site (inset) as derived from the BFES in C.



902



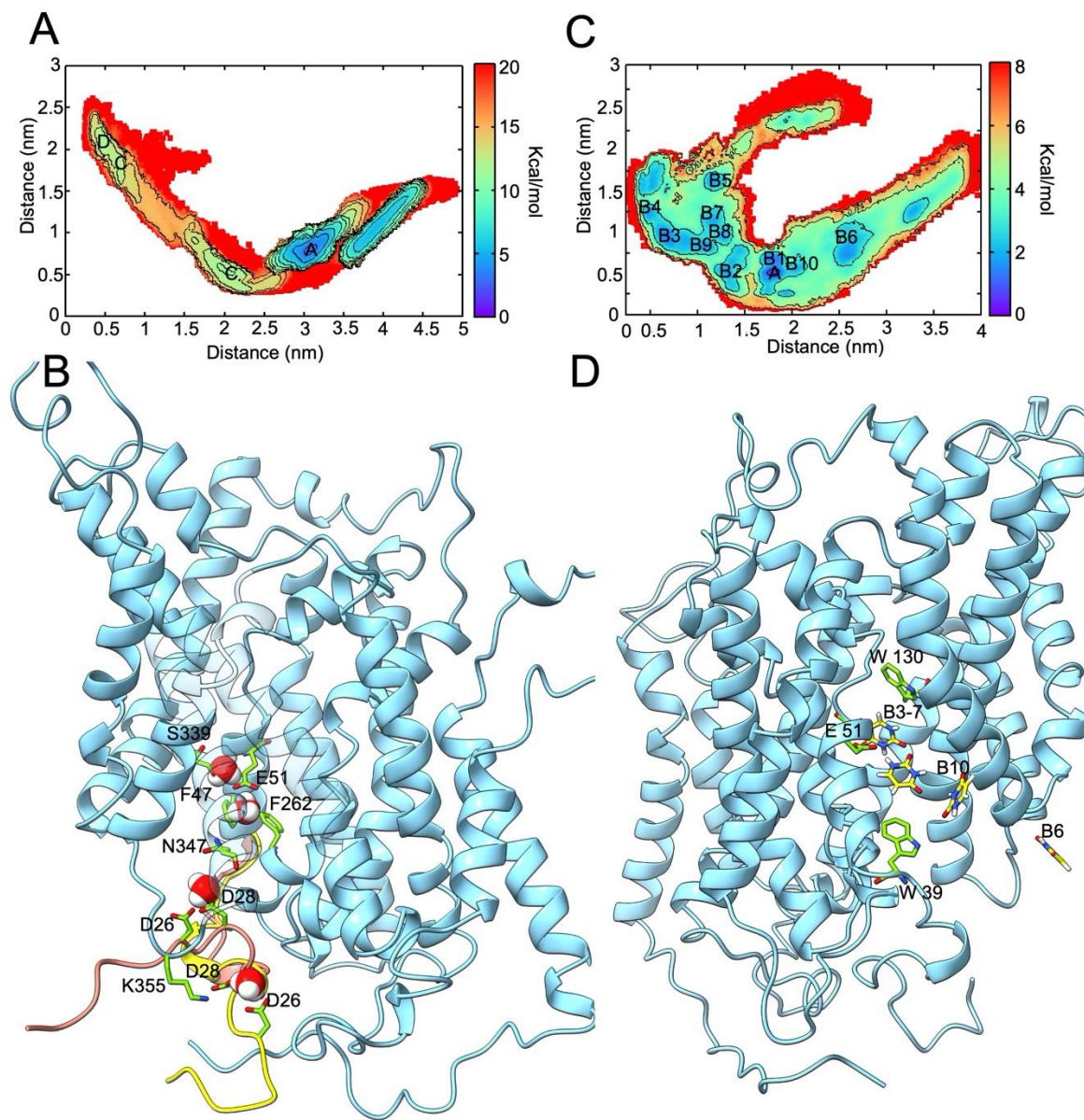
903

904 **Figure 4. FurE structural alterations, residue interactions and Free Energy Surface plots**  
 905 **during transport conformational changes.** (A) Side chain interactions of important residues have  
 906 been monitored in all structures collected in each FES global minimum and are represented as  
 907 percentage over the ensemble of the structures. (B) The FESs of the OO-to-Occ transition using  
 908 different stoichiometry of ligands bound to the transporter (colour code: protein in the *apo* form in  
 909 blue, complexed only with  $H_3O^+$  in cyan, complexed with both uracil and  $H_3O^+$  in green, complex  
 910 only with uracil in grey). Each tick in the x axis represents one unit. (C) The FESs of the Occ-to-  
 911 IO transition are shown using different stoichiometry of ligands bound to the transporter. The  
 912 system containing both uracil and  $H_3O^+$  is represented in red, the system containing only uracil is

913 represented in brown, the *apo* form is represented in magenta, the system containing only  $\text{H}_3\text{O}^+$  is  
914 represented in yellow.

915

916



917

918 **Figure 5. The unbinding process of H<sub>3</sub>O<sup>+</sup> and uracil to the cytoplasm as simulated by Funnel**

919 **Metadynamics.** (A) The BFES of H<sub>3</sub>O<sup>+</sup> internalization process. The separation between contours

920 is 2 kcal/mol. (B) The binding sites of H<sub>3</sub>O<sup>+</sup> cation in FurE transporter along the internalization

921 pathway, as derived from the low energy states in the BFES in A. In orange is represented the

922 cytosolic N-terminal LID when H<sub>3</sub>O<sup>+</sup> is bound in D26, D28 and N347, while in yellow when H<sub>3</sub>O<sup>+</sup>

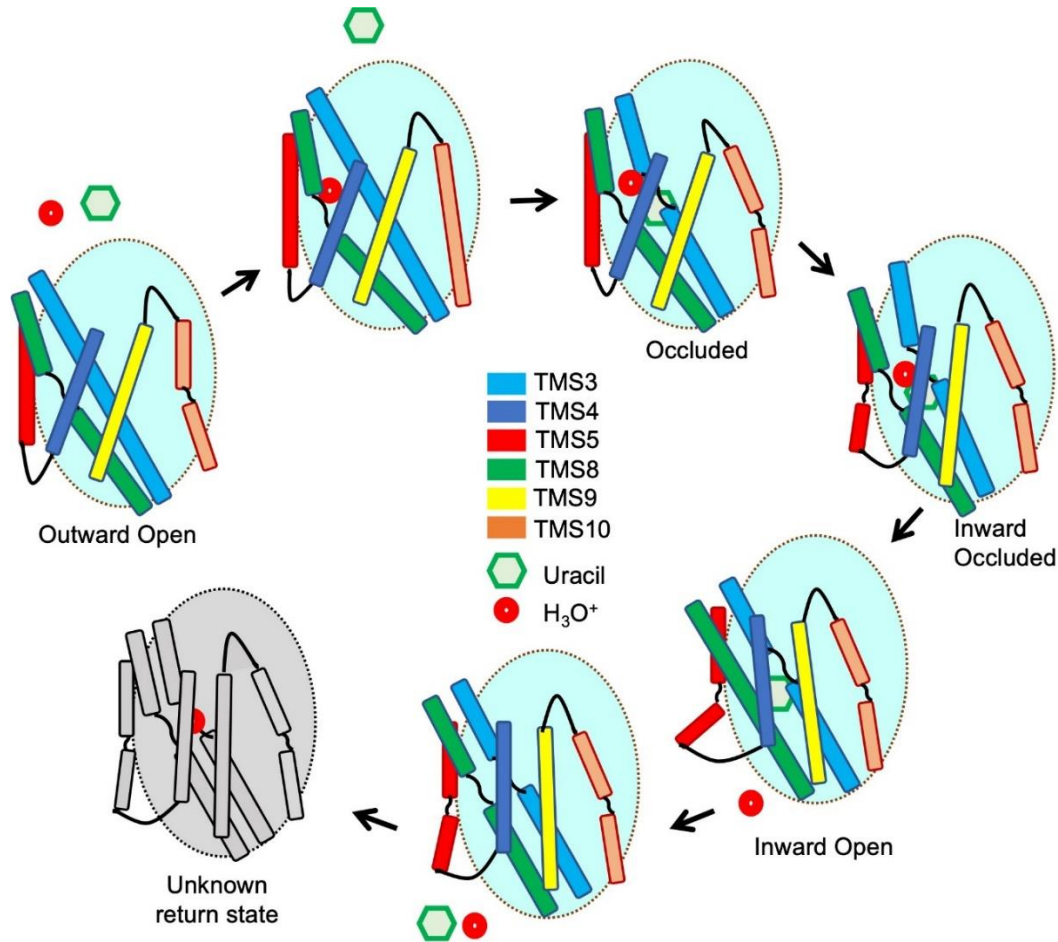
923 is released in the cytoplasm. (C) The BFES of uracil internalization process. The separation between

924 contours is 2 kcal/mol. (D) The intermediate states of uracil internalization pathway while exiting

925 the FurE transporter, as derived from the BFES in C.

926

927



**Figure 6: Schematic representation of the transport mechanism.** The mobile FurE ‘hash’ motif helices (TMS3, TMS4, TMS8, TMS9) and outer and inner gates (TMS10 and TMS5) are shown relative to the ‘bundle’ motif, shown as cyan background, which is considered relatively immobile during uracil and  $\text{H}_3\text{O}^+$  internalization. In the Outward Open (OO) state, FurE is in *apo* form.  $\text{H}_3\text{O}^+$  binding results in local residue rearrangement but does not cause rearrangement of the gross tertiary structure. Uracil binding induces the closing of the TMS10 outer gate and the kink and tilt of TMS8 and TMS3, respectively, reaching the Occluded (Occ) state.  $\text{H}_3\text{O}^+$  moves toward the TMS5 inner gate, which slightly bends, while TMS3 and TMS8 also display structural rearrangements, initiating the Inward Occluded (IOcc) state. After  $\text{H}_3\text{O}^+$  is released in the intracellular space, TMS5 bends more, while TMS8 is not tilted anymore moving away from the ‘bundle’. TMS4 and TMS9 are shifted by TMS5, TMS8 and TMS3 bending introducing the Inward Open (IO) conformation. After the release of both  $\text{H}_3\text{O}^+$  and uracil, TMS5 slightly returns to the previous bend position. An inward-facing unknown return state, probably introduced by a  $\text{H}_3\text{O}^+$ , is represented in grey.