NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF PHYSICS
SECTION OF CONDENSED MATTER PHYSICS

Electronic structure of aperiodic and natural nucleic acid segments, and influence of mutations in charge transfer and transport properties

Marilena Mantela
Ph.D. Thesis

ATHENS 2023

PREFACE

The present dissertation was conducted at the National and Kapodistrian University of Athens, between November 2017 and February 2023, under the supervision of Associate Professor Constantinos Simserides. He is the first person I would like to thank, not only for our collaboration during this dissertation, but during the years of my undergraduate and master studies as well. He trusted in me even when I was still an inexperienced undergraduate physics student back in 2014, and soon I started working on fascinating research projects. His excellent guidance, support and patience - even in times when I was frustrated and impatient, in a combination with his love for physics, is what whetted my appetite for the scientific research. I heartily thank him for all these - and many more.

I would also like to thank the other two members of my Advisory Committee, Associate Professor Spyros Gardelis and Associate Professor Giannis Lelidis for their support and the smooth collaboration we had during all these years. I am also greatly honoured that Associate Professor Vlasis Lykodimos, Professor Nikolaos Stefanou, Professor Dimitrios Frantzeskakis, and Professor Rosa di Felice accepted to be members of my Examination Commitee, and kindly offered their time to read and comment on the present Thesis.

I could not skip on mentioning the past and present members of our research group, "Physics of Nanosctructures and Biomaterials", whom I had the pleasure to collaborate and share an office with; Maria Tassi, Konstantinos Kaklamanis, Marina Theodorakou, Christina Zacharaki, Christina Vantaraki, and Panagiota Bilia. I really enjoyed working with you guys! Special thanks to the post-doctoral researcher Konstantinos Lambropoulos and the Ph.D. student Andreas Morphis, members of the aforementioned group, for the fruitful discussions we had all these years, and for being there to provide consistent help in understanding and solving demanding physical and mathematical problems.

I would like to thank all of my friends for their endless support, and for making my life meaningful, even in the most difficult times.

Last but not least, I would like to deeply thank my father Thanasis for his multilevel support during my whole life, and especially this last year that we went through difficult times as a family. Also, I would like to express my heartfelt

gratitude to my beloved brother Nikos; he constantly proves that he is the best brother a person could have. Finally, not only these lines, but the whole Ph.D. Thesis was written having on my mind the memory of the person that made me who I am today, raised me, loved me, believed in me and supported me more than anyone, in any possible way, during my whole life, as well as during the years of my Ph.D. research. Even if she is not with us anymore, my deepest gratitude goes to her: to my late mom, Chrysoula.

I consider physics as the finest of sciences. It helps us feel and describe the world around us, sparking important unanswered questions, constantly pushing us for a quest of patterns, connections, and relations between objects and phenomena, giving the space to think in different perspectives. Nevertheless, this worthy to explore world, becomes meaningful only through people. As poet Giorgos Seferis stated in his Nobel prize acceptance speech in 1963: "In our gradually shrinking world, everyone is in need of all the others. We must look for man wherever we can find him.".

Abstract

The present Ph.D. Thesis aims to the study of the electronic structure of nucleic acid bases and similar almost planar organic molecules, as well as the investigation of charge transfer and charge transport properties along B-DNA sequences. It also discusses the possibility of using charge transport as a diagnostic tool in discrimination between pathogenic and nonpathogenic mutations. The semi-empirical Linear Combination of Atomic Orbitals (LCAO) method is introduced to calculate the ionization and excitation energies of nucleic acid bases and similar biologically important molecules as well as assemblies of DNA bases, along with a novel parameterization employing all valence orbitals. Then, we outline the Tight Binding method for charge transfer of an extra carrier along DNA, and we also introduce the physical quantities studied. More specifically, we employ a TB wire model, where the base pairs are the sites of the chain, to study the spectral and charge transfer properties of periodic sequences with increasing repetition unit, as well as deterministic aperiodic DNA segments. In addition, we address the impact of structural flexibility on the electronic structure and charge transfer ability of B-DNA. To this end, we apply our LCAO method to 20 AA and GG dimers, extracted from representative structures in a classical MD trajectory of a 20mer, and study some useful physical quantities. Finally, we move on to investigate charge transport along DNA molecules, using the time-independent Schrödinger equation together with the transfer-matrix method in order to finally obtain current-voltage I-V curves. We examine ideal and natural geometries concerning two categories of mutations: (i) DNA sequences that contain point substitution mutations, and (ii) sequences extracted from segments of human chromosomes, modified by expansion of the CAG triplet to mimic diseases.

ΠΕΡΙΛΗΨΗ

Τα νουκλεϊκά οξέα είναι σύνθετα βιολογικά μακρομόρια, αποτελούμενα από αλληλουχίες νουκλεοτιδίων και περιέχουν γενετική πληροφορία. Τα νουκλεϊκά οξέα βρίσκονται στα κύτταρα όλων των έμβιων οργανισμών και είναι το δεοξυριβο-νουκλεϊκό οξύ (DNA) και το ριβονουκλεϊκό οξύ (RNA). Το DNA περιέχει τις γενετικές πληροφορίες που καθορίζουν τη βιολογική ανάπτυξη όλων των κυτταρι-κών μορφών ζωής και των περισσοτέρων ιών. Τα τελευταία χρόνια, το DNA και οι ιδιότητές του που σχετίζονται με τη μετακίνηση φορτίου κατά μήκος του, έχουν προσελκύσει το ενδιαφέρον διαφόρων επιστημονικών κλάδων, από τις βιοεπιστήμες και τη φυσικοχημεία μέχρι τη νανοτεχνολογία. Κάτι τέτοιο είναι λογικό αν σκεφτεί κανείς ότι το DNA, ως φορέας της γενετικής πληροφορίας, διατηρεί τον κεντρικό ρόλο στην ανάπτυξη, τη λειτουργία και την αναπαραγωγή των οργανισμών, οπότε η μελέτη του συνήθως συνδέεται με επιστημονικούς κλάδους όπως η γενετική, η βιολογία και η ιατρική. Επιπλέον όμως, τα ιδιαίτερα χαρακτηριστικά του μορίου, έχουν προσανατολίσει την έρευνα στην κατεύθυνση της μεταβίβασης και μεταφοράς φορτίου σε πολυμερικά συστήματα. Τα ζεύγη βάσεων του DNA, καθώς και η αλληλουχία τους, δημιουργούν έναν π-δρόμο, λόγω της επικάλυψης των π μοριακών τροχιακών τους. Αυτή η επικάλυψη επιτρέπει τη μετακίνηση φορέων ηλεκτρικού φορτίου, οπών ή ηλεκτρονίων. Η οξείδωση του μορίου δημιουργεί οπές και η αναγωγή επιπλέον ηλεκτρόνια. Ο όρος μετακίνηση περιλαμβάνει τη *μεταβίβαση* και τη *μεταφορά*. Με τον όρο *μεταβίβαση* (*transfer*) περιγράφεται η μετακίνηση ενός φορέα από μια τοποθεσία όπου δημιουργείται ή εγχέεται αρχικά σε άλλες ευνοϊκότερες τοποθεσίες χωρίς τη μεσολάβηση κάποιου εξωτερικού παράγοντα όπως διαφορά δυναμικού ή βαθμίδα θερμοκρασίας. Ο όρος *μεταφορά* (*trans-port*) αναφέρεται στη μετακίνηση φορτίου λόγω εφαρμογής εξωτερικής διαφοράς δυναμικού μέσω ηλεκτροδίων ή κάποιας άλλης βαθμίδας (π.χ. θερμοκρασίας).

Ένα χαρακτηριστικό γνώρισμα του DNA είναι το σχετικά μεγάλο μήκος ακαμ-ψίας του (persistence length), το οποίο εκτείνεται σε περίπου 150 ζεύγη βάσεων (50 nm). Το χαρακτηριστικό αυτό του επιτρέπει να παραμένει σχετικά άκαμπτο σε μεγάλα μήκη, έτσι ώστε να μπορεί να χρησιμοποιηθεί ως κανάλι (δίαυλος, σχεδόν μονοδιάστατο μοριακό σύρμα) σε ηλεκτρονικές νανοδιατάξεις και στην κατασκευή νανοκυκλωμάτων. Πέραν αυτού, το DNA παρέχει τη δυνατότητα κατασκευής μιας μεγάλης ποικιλίας διαφορετικών πολυμερών, συνδυάζοντας τα διαφορετικά ζεύγη βάσεων (Αδενίνη-Θυμίνη, Γουανίνη-Κυτοσίνη) με διαφορετικούς τρόπους. Κατ' αυτό τον τρόπο είναι δυνατό να δημιουργηθούν κατ' επιθυμίαν ακολουθίες, κατάλληλες να επιτελέσουν μια συγκεκριμένη λειτουργία, και κατά συνέπεια προ-σφέρεται η δυνατότητα ρύθμισης των ιδιοτήτων μιας διάταξης. Επιπλέον, η μετα-βίβαση φορτίου κατά μήκος του μορίου του DNA επιτελεί σημαντικό βιολογικό ρόλο, καθώς φαίνεται να σχετίζεται με την καρκινογένεση και τη μεταλλαξιγένεση, ενώ η μακράς εμβέλειας μεταβίβαση φαίνεται να συνδέεται με τις διαδικασίες κατα-

v

στροφής και επιδιόρθωσης του DNA.

Η παρούσα Διδακτορική Διατριβή επιχειρεί να αποτελέσει μια συστηματική θεωρητική μελέτη της ηλεκτρονικής δομής των νουκλεϊκών οξέων, των αζωτούχων βάσεων και παρομοίων οργανικών μορίων, καθώς και των ιδιοτήτων μεταβίβασης και μεταφοράς φορτίου στο DNA, όπως και να συμβάλει στην ανάπτυξη θεωρητικών εργαλείων για την ανίχνευση μεταλλάξεων. Για τη μελέτη χρησιμοποιήθηκε η μέθοδος του Γραμμικού Συνδυασμού των Ατομικών Τροχιακών (Linear Combination of Atomic Orbitals, LCAO) και το Πρότυπο της Ισχυρής Δέσμευσης (Tight Binding, TB). Η μέθοδος αυτή έχει το πλεονέκτημα της εύκολης και γρήγορης εφαρμογής σε μεγάλης έκτασης συστήματα, παρέχοντας αποτελέσματα προσαρμόσιμης ακρίβειας. Αυτές οι ιδιότητες της την καθιστούν κατάλληλη κυρίως για τη μελέτη μεγάλων συστημάτων, τα οποία επί του παρόντος δεν είναι διαχειρίσιμα με τις διαθέσιμες λεπτομερέστερες μεν, αλλά πολύ υψηλότερου υπολογιστικού κόστους, μεθόδους από πρώτες αρχές (ab-initio).

Η Διδακτορική Διατριβή αποτελείται από έξι Κεφάλαια.

Στο Κεφάλαιο 1 παρουσιάζονται τα βασικά δομικά χαρακτηριστικά των νουκλεϊκών οξέων, τα ερωτήματα στα οποία καλείται να απαντήσει η παρούσα Διατριβή, καθώς και η διάρθρωσή της.

Στο Κεφάλαιο 2 μελετάται η ηλεκτρονική δομή των βάσεων των νουκλεϊκών οξέων και παρόμοιων μορίων με τη μέθοδο του γραμμικού συνδυασμού των ατομικών τροχιακών (LCAO). Στους υπολογισμούς συμπεριλαμβάνονται όλα τα ατομικά τροχιακά σθένους και συγκρίνονται τα αποτελέσματα με αντίστοιχα υπολογισμών που λαμβάνουν υπόψη μόνο τα $2p_z$ ατομικά τροχιακά, υπολογισμών από πρώτες αρχές, καθώς και με πειραματικές τιμές των ενεργειών ιονισμού και διέγερσης. Η μέση απόκλιση μεταξύ θεωρητικών και πειραματικών τιμών είναι για τις ενέργειες ιονισμού 3.65% και για τις ενέργειες διέγερσης 6.49%. Επιπλέον, υπολογίζουμε τις αντίστοιχες τιμές για τα ζεύγη βάσεων (Αδενίνη-Θυμίνη, Γουανίνη-Κυτοσίνη).

Στο Κεφάλαιο 3, συζητείται η μεταβίβαση φορτίου (charge transfer) σε διάφορα πολυμερή τμήματα DNA. Αρχικά, θεμελιώνεται θεωρητικά το Πρότυπο Σύρματος (Wire Model) της Μεθόδου Ισχυρής Δέσμευσης (Tight-Binding Method) και εφαρμόζεται στην περιγραφή της μεταβίβασης φορτίου στο B-DNA σε επίπεδο ζευγών βάσεων. Εξετάζεται η συμπεριφορά ενός επιπλέον φορέα (ηλεκτρονίου ή οπής) αφού τοποθετηθεί σε ένα μονομερές, δηλαδή σε ένα ζεύγος βάσεων, και μελετάται η μετακίνησή του σε όλο το μήκος του πολυμερούς. Οι χρησιμοποιούμενες παράμετροι είναι οι επιτόπιες ενέργειες (on-site energies) του φορέα σε κάθε ζεύγος βάσεων και οι παράμετροι ή ολοκληρώματα αλληλεπίδρασης (interaction parameters or integrals) μεταξύ αμέσως γειτονικών ζευγών βάσεων. Αρχικά, υπολογίζονται οι παράμετροι μεταβίβασης χρήσει της παραμετροποίησης η οποία μεταχειρίζεται όλα τα ατομικά τροχιακά σθένους. Τα αποτελέσματα συγκρίνονται με πειραματικές τιμές, με τις οποίες υπάρχει καλή συμφωνία. Επίσης, συγκρίνουμε με αποτελέσματα παραμετροποιήσεων, οι οποίες χρησιμοποιούν μόνο τα $2p_z$ ατομικά τροχιακά.

Έχοντας το πλήρες σύνολο των απαιτούμενων για τη χρήση της μεθόδου Ισχυρής Δέσμευση παραμέτρων, μελετώνται ορισμένες περιοδικές ακολουθίες DNA με αυξανόμενη μονάδα επανάληψης, οι οποίες αποτελούνται είτε από το ίδιο (κατηγορία I από το «Identical», π.χ. GC..., GGCC... κ.ο.κ.) είτε από διαφορετικά (κατηγορία D από το «Different», π.χ. GA..., GGAA..., κ.ο.κ.) ζεύγη βάσεων.

Υπολογίζονται φυσικά μεγέθη όπως: το ενεργειακό φάσμα, η πυκνότητα καταστάσεων, η μέση πιθανότητα εύρεσης του φορέα σε κάθε μονομερές, το συχνοτικό περιεχόμενο (ο φορέας εκτελεί ταλάντωση σε σύρμα πεπερασμένου μήκους) και ο καθαρός μέσος ρυθμός μεταβίβασης. Βρίσκουμε ότι για τα I πολυμερή, με την αύξηση της μονάδας επανάληψης, όλες οι μελετώμενες ποσότητες προσεγγίζουν τις αντίστοιχες των ομοπολυμερών (π.χ. G...), ενώ για τα D πολυμερή οι αντίστοιχες ποσότητες προσεγγίζουν την ένωση των δύο πιθανών ομοπολυμερών (π.χ. G... και A ...). Στη συνέχεια του Κεφαλαίου αυτού, αναλύονται και σχολιάζονται τα αντίστοιχα χαρακτηριστικά μεγέθη που έχουν προκύψει από την εφαρμογή της μεθόδου Ισχυρής Δέσμευσης σε οιονεί περιοδικές (Fibonacci, Thue-Morse, Double-Period, Rudin-Shapiro) και μορφοκλασματικές (Cantor Dust, Asymmetric Cantor Set) ακολουθίες. Συγκρίνοντας τα αποτελέσματα για τα περιοδικά και τα απεριοδικά πολυμερή, διαπιστώνουμε ότι παρά το γεγονός ότι η μεταβίβαση φορτίου ευνοείται σε γενικές γραμμές στα περιοδικά συστήματα, εντούτοις υπάρχουν ορισμένα αρκετά αποδοτικά ως προς τη μεταβίβαση φορτίου απεριοδικά πολυμερή. Παρά ταύτα, τα ομοπολυμερή, τα οποία είναι και τα απλούστερα δομικά πολυμερή, αποτελούν άνω όριο για όλες τις κατηγορίες πολυμερών, παρουσιάζουν δηλαδή ευκολότερη μεταβίβαση. Επιπλέον, το τυχαίο ανακάτεμα ζευγών βάσεων των ντετερμινιστικών απεριοδικών αλληλουχιών οδηγεί γενικά σε αμελητέους ρυθμούς μεταβίβασης. Τέλος, η παραπάνω ανάλυση εφαρμόζεται για τον υπολογισμό των μέσων ρυθμών μεταβίβασης ορισμένων ακολουθιών οι οποίες έχουν χρησιμοποιηθεί σε πειράματα, και τα αποτελέσματα συγκρίνονται με τις αντίστοιχες τιμές της βιβλιογραφίας.

Ακολούθως, στο Κεφάλαιο 4, διερευνάται η επίδραση της δομικής μεταβλητότητας στην ηλεκτρονική δομή και στην ικανότητα μεταβίβασης φορτίου κατά μήκος του DNA. Για τη διερεύνηση αυτή επιστρατεύτηκαν 10 διμερή Αδενίνη-Θυμίνη/Αδενίνη-Θυμίνη και άλλα 10 Γουανίνη-Κυτοσύνη/Γουανίνη-Κυτοσύνη, τα οποία εξήχθησαν αφού εφαρμόστηκε η μέθοδος της Μοριακής Δυναμικής σε ένα 20μερές. Για τα συστήματα αυτά υπολογίστηκαν διάφορες παράμετροι (όπως το μέγιστο ποσοστό μεταβίβασης και τα ολοκληρώματα μεταβίβασης), σχετιζόμενες με δομικές παραμέτρους. Σε γενικές γραμμές, η μεταβίβαση φορτίου παραμένει ισχυρή, παρότι επηρεάζεται από τη δομική μεταβλητότητα. Επιπλέον, περιλαμβάνεται σύγκριση ορισμένων αποτελεσμάτων που προκύπτουν με χρήση του Προτύπου Ισχυρής Δέσμευσης, με λεπτομερέστερες μεθόδους από πρώτες αρχές. Το Πρότυπο Ισχυρής Δέσμευσης μπορεί να δώσει μια γρήγορη εκτίμηση των προς μελέτη ποσοτήτων.

Στο Κεφάλαιο 5, παρουσιάζεται συνοπτικά η μέθοδος των Πινάκων Μεταβίβασης (Transfer Matrix Method) και εφαρμόζεται στη μελέτη της ηλεκτρονικής δομής και της μεταφοράς φορτίου κατα μήκος ακολουθιών DNA που αφορούν περιπτώσεις: α) σημειακών μεταλλάξεων και β) μεταλλάξεων που προέρχονται από επανάληψη της τριπλέτας (CAG). Οι μεταλλάξεις της δεύτερης κατηγορίας ονομάζονται μεταλλάξεις Short Tandem Repeat (STR) expansions και στην παρούσα εργασία μελετώνται εξ αυτών οι: νόσος Huntington, νόσος Kennedy, νωτιαία παρεγκεφαλιδική αταξία 6 και νωτιαία παρεγκεφαλιδική αταξία 7. Μελετώνται τόσο περιπτώσεις ιδανικών πολυμερών, όσο και φυσικών. Εδώ ο όρος «ιδανικό» αναφέρεται σε ιδανική γεωμετρία, ενώ ο όρος «φυσικό» αναφέρεται σε γεωμετρία συλλεγμένη από βάσεις δεδομένων αλληλουχιών οι οποίες προέρχονται από οργανι-

σμούς. Το πιο χρήσιμο χαρακτηριστικό στη μελέτη της μεταφοράς φορτίου είναι οι χαρακτηριστικές καμπύλες I-V. Κατά την εισαγωγή μεταλλάξεων, οι καμπύλες αυτές αλλάζουν σημαντικά. Για το λόγο αυτό, ορίζεται ένα νέο μέγεθος που περιγράφει την κανονικοποιημένη απόκλιση της I-V από την αρχή των αξόνων (normalized deviation of the I-V curve from the origin, NDIV). Ένα βασικό συμπέρασμα που εξάγεται είναι ότι η NDIV γενικά δεν επηρεάζεται από τη θέση μιας σημειακής μετάλλαξης, αλλά επηρεάζεται από τον αριθμό των σημειακών μεταλλάξεων καθώς και από τον αριθμό των STR επαναλήψεων. Σε γενικές γραμμές θα μπορούσαμε να πούμε ότι στο μέλλον, η μελέτη της μεταφοράς φορτίου δύναται να αποτελέσει ένα νέο εργαλείο για την ανίχνευση μεταλλάξεων.

Το Κεφάλαιο 6, περιέχει έναν ευρύτερο συγκριτικό σχολιασμό των αποτελεσμάτων της Διδακτορικής Διατριβής, καθώς και των προοπτικών για μελλοντική έρευνα.

LIST OF PUBLICATIONS

A list of the published research activity of the author as a Ph.D. candidate.

Articles in Peer-Reviewed Journals

1. Periodic polymers with increasing repetition unit: Energy structure and carrier transfer.
   K. Lambropoulos, C. Vantaraki, P. Bilia, **M. Mantela**, and C. Simserides,
   Physical Review E **98** (2018) 032412 (14 pages and 23 pages Supplemental Material)
   https://doi.org/10.1103/PhysRevE.98.032412

2. Quasi-periodic and fractal polymers: Energy structure and carrier transfer.
   M. Mantela, K. Lambropoulos, M. Theodorakou, and C. Simserides,
   Materials **12** (2019) 2177 (30 pages)
   https://doi.org/10.3390/ma12132177

3. Effects of structural dynamics on charge carrier transfer in B-DNA: a combined MD and RT-TDDFT study.
   M. Mantela, A. Morphis, K. Lambropoulos, C. Simserides, and R. Di Felice,
   Journal of Physical Chemistry B **125 (16)** (2021) 3986 - 4003
   https://doi.org/10.1021/acs.jpcb.0c11489

4. LCAO Electronic Structure of Nucleic Acid Bases and Other Heterocycles and Transfer Integrals in B-DNA, Including Structural Variability.
   M. Mantela, C. Simserides, and R. Di Felice,
   Materials **14 (17)** (2021) 4930 (20 pages)
   https://doi.org/10.3390/ma14174930

5. Charge transport properties of ideal and natural DNA segments, as mutation detectors.
   M. Mantela, K. Lambropoulos, and C. Simserides,
   Physical Chemistry Chemical Physics **25** (2023) 7750 - 7762
   https://doi.org/10.1039/D3CP00268C

Articles in International Conference Proceedings

1. Frequency Content of Carrier Oscillations along B-DNA Polymers.
   K. Lambropoulos, **M. Mantela**, and <u>C. Simserides</u>
   PIERS (Progress In Electromagnetics Research Symposium) Proceedings,
   19-22 November 2017, Singapore,
   pp. 186-193, doi:10.1109/PIERS-FALL.2017.8293134

2. Frequency content of carrier oscillations along B-DNA aperiodic and periodic polymers.
   <u>**M. Mantela**</u>, K. Lambropoulos, C. Vantaraki, P. Bilia, A. Morphis, and C. Simserides
   PIERS (Progress In Electromagnetics Research Symposium) Proceedings,
   17-20 June 2019, Rome, Italy,
   pp. 831-838, doi:10.1109/PIERS-Spring46901.2019.9017899

Conference Presentations

1. 39th Progress In Electromagnetics Research Symposium in Singapore (PIERS 2017 in Singapore), 19-22 November 2017, Singapore
   Talk: "Frequency Content of Carrier Oscillations along B-DNA Polymers."
   K. Lambropoulos, **M. Mantela**, and <u>C. Simserides</u>

2. 33th Panhellenic Conference on Solid State Physics and Materials Science, University of Cyprus, 17-19 September 2018, Nicosia, Cyprus
   Talk: "Quasi-periodic and fractal polymers: energy structure and carrier transfer."
   <u>**M. Mantela**</u>, K. Lambropoulos, M. Theodorakou, and C. Simserides

3. 41st Progress In Electromagnetics Research Symposium in Rome (PIERS 2019 in Rome), 17-20 June 2019, Rome, Italy
   Talk: "Frequency content of carrier oscillations along B-DNA aperiodic and periodic polymers."
   <u>**M. Mantela**</u>, K. Lambropoulos, C. Vantaraki, P. Bilia, A. Morphis, and C. Simserides *Honorable Mention (Antennas & Microwave Technologies) for the presenter M. Mantela*

4. 18th International Conference on Nanosciences & Nanotechnologies (NN21), 6-9 July 2021, Thessaloniki, Greece
   Talk: "How structural dynamics affect hole transfer in B-DNA: A Combination of MD, RT-TDDFT and TB."
   <u>**M. Mantela**</u>, A. Morphis, K. Lambropoulos, C. Simserides, and R. Di Felice

5. The 2021 World Congress on Advances in Structural Engineering and Mechanics (ASEM21) and The 2021 World Congress on Advances in Nano, Bio, Robotics, and Energy (ANBRE21), 23-26 August 2021, Seoul, South Korea
   Talk: "Structural dynamics and hole transfer in B-DNA: combining MD, RT-TDDFT and TB."
   M. Mantela, A. Morphis, K. Lambropoulos, C. Simserides and R. Di Felice

6. 35th Panhellenic Conference on Solid State Physics and Materials Science, 26-29 September 2021, Athens, Greece
   Talk: "How structural variability affect hole transfer in B-DNA: A Combination of MD, RT-TDDFT and TB."
   M. Mantela, A. Morphis, K. Lambropoulos, C. Simserides and R. Di Felice

CONTENTS

Dictionary

The present list includes terms of the international scientific literature that are abbreviated in the present Ph.D. Thesis, in alphabetical order.

Term	Abbreviation	Description
Adenine	A	purine $C_5H_5N_5$
Asymmetric Cantor Set	ACS	fractal sequence
Cantor Set	CS	fractal sequence
Cytosine	C	pyrimidine $C_4H_5N_3O$
Density Functional Theory	DFT	computational quantum mechanical modelling method to investigate the electronic structure of many-body systems
Density Of States	DOS	number of states that can be occupied by electrons per energy interval
Deoxyribonucleic Acid	DNA	nucleic acid containing the genetic instructions for all known organisms and many viruses
Double-Period	DP	quasi-periodic sequence
Far infrared	FIR	radiation with wavelength of 8 μm - 1 mm, or 300 GHz - 37 THz
Guanine	G	purine $C_5H_5N_5O$
Highest Occupied Molecular Orbital	HOMO	
Integrated Density Of States	IDOS	number of states that can be occupied by electrons with energy smaller than E

Term	Abbreviation	Description
Linear Combination of Atomic Orbitals	LCAO	approximate method for calculating molecular orbitals, based upon quantum superposition of atomic orbitals
Linear Combination of Molecular Orbitals	LCMO	approximate method for calculating orbitals of bigger molecules, based upon quantum superposition of smaller component molecular orbitals
Lowest Unoccupied Molecular Orbital	LUMO	
Mid infrared	MIR	radiation with wavelength of 2.5 µm - 8 µm, or 37 THz - 120 THz
Near infrared	NIR	radiation with wavelength of 780 nm - 2.5 µm, or 120 THz - 384 THz
Real-Time Time-Dependent Density Functional Theory	RT-TDDFT	DFT-based approach to electronic excited states based on integrating the time-dependent Kohn-Sham equations in time
Ribonucleic Acid	RNA	nucleic acid containing essential information for many biological functions of the living organisms, while encoding the genetic information of many viruses
Rudin-Shapiro	RS	quasi-periodic sequence
Thue-Morse	TM	quasi-periodic sequence
Thymine	T	pyrimidine $C_5H_6N_2O_2$
Tight-Binding	TB	semi-empirical approach to the calculation of electronic band structure using a set of wave-functions based upon superposition of wavefunctions for isolated atoms located at each atomic site
Total Weighted Mean Frequency	TWMF	the sum of all WMFs multiplied by the corresponding probabilities to find the extra carrier at each monomer

Term	Abbreviation	Description
Transmission Coefficient	TC	the probability that a carrier transmits through a quantum system's eigenstates
Uracil	U	pyrimidine $C_4H_4N_2O_2$
Wire Model	WM	the simplest variation of the TB model, where a site is considered a monomer
Weighted Mean Frequency	WMF	mean frequency content of an extra carrier oscillation at a specific monomer

1 | INTRODUCTION

Today, the investigation of biomolecules' structure and properties - especially these of the nucleic acids - is emerged as a high significance issue. Nucleic acids are biopolymers, macromolecules, essential to all known forms of life. The *deoxyribonucleic acid* (*DNA*) carries genetic instructions for the development, functioning, growth and reproduction of all known organisms, as well as many viruses. The *ribonucleic acid* (*RNA*) is essential in various biological roles in coding, decoding, regulation and expression of genes, acts like a catalyst of biochemical reactions, and besides these, it encodes the genetic information of many viruses.

Both DNA and RNA are consisted of nucleotides. However, DNA has a double-stranded structure [1, 2], while RNA is single-stranded [3--5]. The two DNA strands are composed of simpler monomeric units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases, i.e., adenine (A), thymine (T) (substituted by uracil (U) in RNA), guanine [G] or cytosine (C), a pentose sugar called deoxyribose (ribose in RNA), and a phosphate group. The five nitrogenous bases are depicted in Fig. 1.1. The nucleotides are
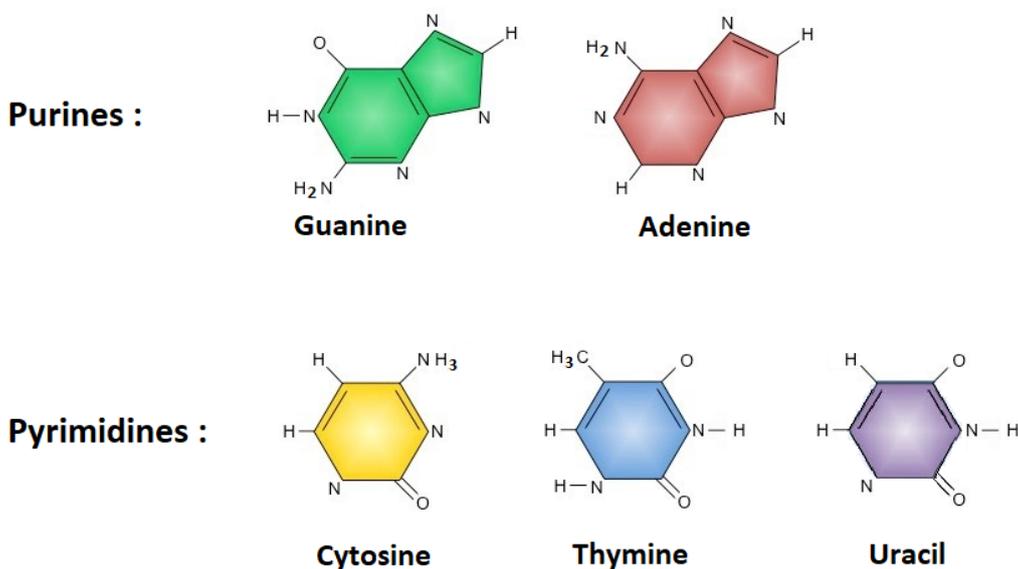
Figure 1.1: The nitrogenous bases of nucleic acids. Purines (G and A) and pyrimidines (C, T and U). Image source: https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/nitrogenous-bases.html.

joined to one another in a chain by covalent bonds (phosphodiester linkage) between the sugar of one nucleotide and the phosphate of the next, resulting in an

alternating sugar-phosphate backbone. The phosphodiester bonds are formed
between the 3' and 5' carbons of the sugars of successive nucleotides. The ori-
entation of the 3' and 5' carbons along the sugar-phosphate backbone confers
directionality to each DNA strand. In a nucleic acid double helix, the direction
of the nucleotides in one strand is opposite to their direction in the other strand:
the strands are antiparallel. In DNA, the nitrogenous bases of the two strands
are bound together with hydrogen bonds, according to base pairing rules, i.e.,
purines (A or G) with pyrimidines (T or C, respectively). This is called comple-
mentary base pairing. Adenine is bonded only to thymine with two hydrogen
bonds, and cytosine is bonded only to guanine with three hydrogen bonds, form-
ing a base pair. Fig. 1.2 depicts the complementary bases and their bonding.

Thymine - Adenine   Cytosine - Guanine
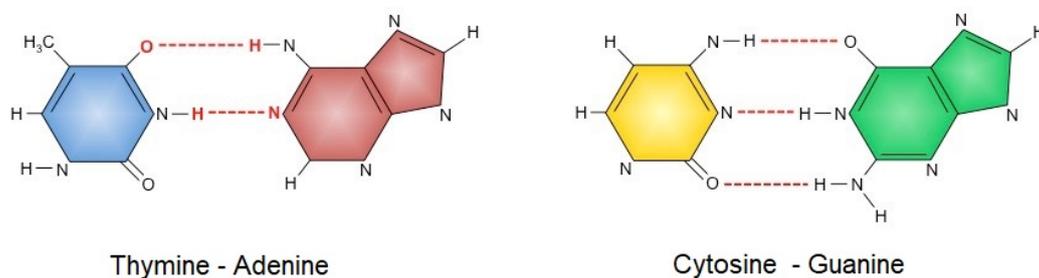
Figure 1.2: Bonded complementary base pairs. Adenine is bonded only to
thymine with two hydrogen bonds, and cytosine is bonded only to guanine
with three hydrogen bonds. Image source: https://ib.bioninja.com.au/
standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/
nitrogenous-bases.html.

In the present Ph.D. Thesis, B-DNA structures have been studied. The
B-DNA is the most common DNA conformation found in nature, under the
conditions found in cells (other conformations are the A-DNA and the Z-DNA).
In this conformation, the double helix is right-handed with about $10 - 10.5$ base
pairs per turn (helix step). The helix step is ≈ 34 Å long and its diameter is $\approx$
20 Å. The distance between two successive base pairs is ≈ 3.4 Å and the related
twist is $\approx 36°$ (Fig. 1.5).

In our calculations within this Ph.D. Thesis, we will employ only valence
electrons of the atoms of each molecule (base or similar). A valence electron is
an electron in the outer shell associated with an atom, and that can participate
in the formation of a chemical bond if the outer shell is not closed. In a single co-
valent bond, a shared pair forms with both atoms in the bond each contributing
one valence electron. Each orbital in an atom is characterized by a set of values
of the three quantum numbers n, ℓ, and m_ℓ, which respectively correspond to
the electron's energy, angular momentum, and magnetic quantum number. The
simple names **s** orbital, **p** orbital, **d** orbital, and **f** orbital refer to orbitals with
angular momentum quantum number $\ell = 0$, 1, 2, and 3 respectively. These
names, together with the value of $n = 0$, 1, 2, ..., are used to describe the
electron configurations of atoms. The organic compounds that we encountered

are composed of the elements Hydrogen (H), Carbon (C), Nitrogen (N), and Oxygen (O). Hydrogen utilizes 1s valence orbitals, while, carbon, nitrogen and oxygen utilize 2s and 2p ($2p_x$, $2p_y$, $2p_z$) orbitals. Fig. 1.3 depicts the shape of these orbitals.
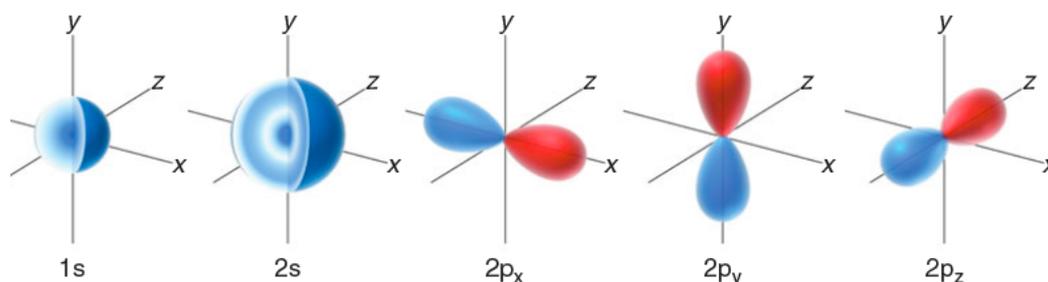
Figure 1.3: 1s, 2s, $2p_x$, $2p_y$, $2p_z$ orbitals employed in the present Ph.D. Thesis' calculations. Image source: https://chemicalalgos.wordpress.com/tag/energy/.

These orbitals form bonds of σ or π type, as depicted in Fig. 1.4.
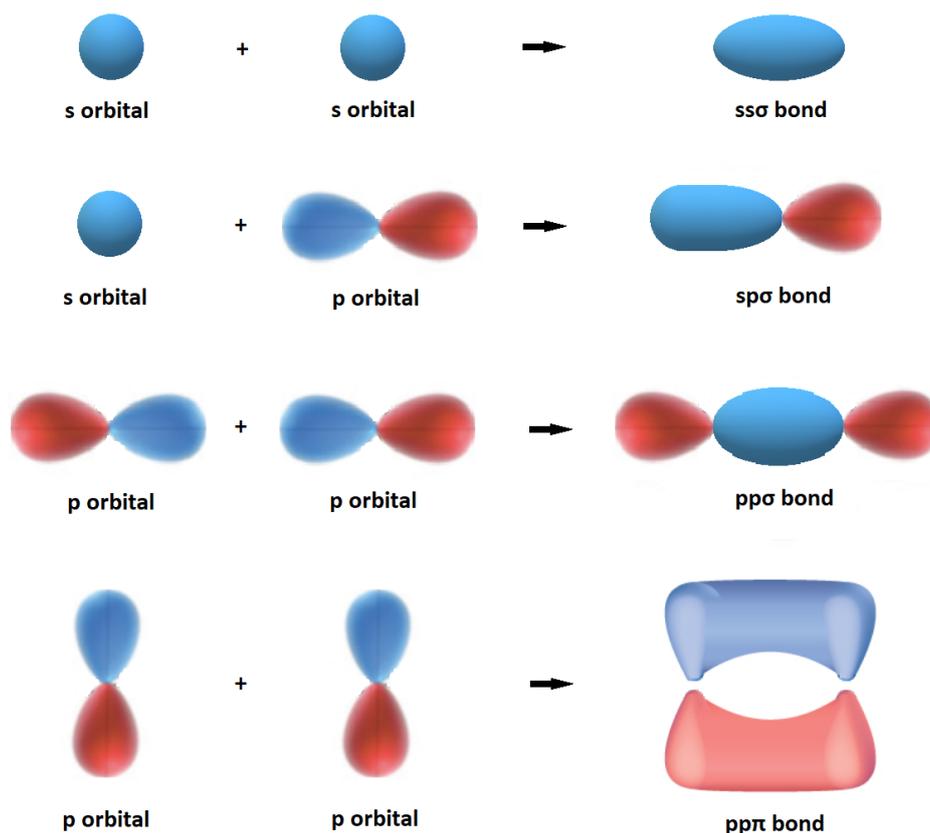
Figure 1.4: Covalent bonds between different types of atomic orbitals. Here are depicted only bonding orbitals.

Except for its biological role, that has been prominent with the previous discussion, the DNA molecule could, as well, be viewed as a molecular wire. DNA

has a persistence length of $\ell_p = 50$ nm or 150 base pairs [6], that makes it hard to bend, unless it is longer than ℓ_p. In addition to the fairly large persistent length, DNA provides the opportunity to construct easily a variety of different polymers by combining a given number of base pairs, each one containing two complementary bases.

The structure of DNA favours the overlap of the electron density of adjacent bases, which, besides stabilising the double helix, creates a nearly one-dimensional π-pathway along which charge transfer and transport are possible. The term *transfer* implies that a carrier, created (e.g. a hole by oxidation or an extra electron by reduction) or injected at a specific location, moves to more favourable sites, without application of external gradient (e.g. temperature gradient or voltage). The term *transport* implies the application of an external gradient.

It is clear from the discussion so far, that the understanding of the electronic structure and charge transfer [7] properties of DNA is a crucial issue in biology, involved in functions like damage and repair [8--10], carcinogenesis and mutagenesis [11, 12]. For example, the rapid hole migration from other bases to guanine is connected to the fact that direct strand breaks occur preferentially at guanines [11]. Charge transfer and transport are relevant in discrimination between pathogenic and nonpathogenic mutations [13]. Charge transport could probe DNA of different origin or organisms [14], mutations and diseases [15, 16]. Distortions of base-pair stack [17, 18] affect charge transfer and transport. Deviations in that stacking, e.g., through base modifications, insertions, or protein binding, can be electrically observed. DNA charge transfer and transport has been used to detect changes in DNA, like lesions, binding proteins, protein activity,
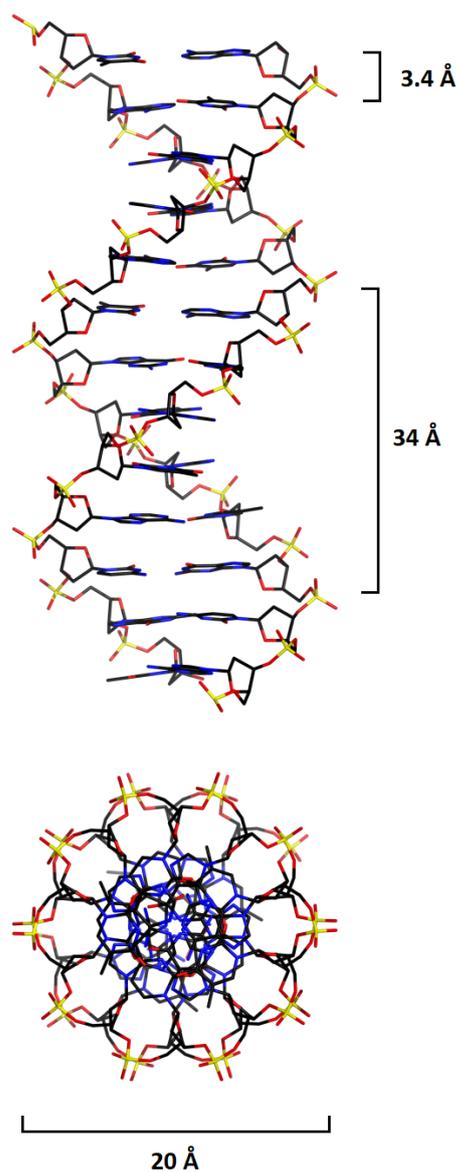
Figure 1.5: The B-DNA conformation. Image source: https://en.wikipedia. org/wiki/Nucleic_acid_double_helix.

even reactions under weak magnetic fields [19]. These properties are also important for novel applications in nanotechnologies [20, 21]. The last two decades, the intense interest in the experimental and theoretical probing of the electronic structure of DNA-based polymers has arise in connection to the quest for the development of molecular electronics [22]. We have witnessed a surge of studies that define DNA polymers as the basis for molecular wires and molecular electronics devices/circuits, based on self-assembly and specific base hybridization [22--26]. The prospect of using DNA in materials science stems from exploiting its properties of molecular recognition, assembly, and processing information [23]. Properties of long-range charge transport in DNA and DNA-mediated charge transfer and mechanisms have been studied a for a long time now [25]. Today DNA plays an increasingly important role in molecular electronics due to its structural and molecular recognition properties [22].

Considering the above, the present Ph.D. Thesis aims to enlighten the mechanisms of correlation of the DNA sequencing and its charge transfer and transport properties, by employing a simple and broadly used theoretical method, i.e., Tight-Binding. In Chapter 2, we introduce the semi-empirical Linear Combination of Atomic Orbitals (LCAO) method to calculate the ionization and excitation energies of nucleic acid bases and similar biologically important molecules as well as assemblies of DNA bases. We set the relevant parameterization. Chapter 3 is dedicated in charge transfer. We outline the TB theory for charge transfer of an extra carrier along DNA, and we also introduce the physical quantities studied. More specifically, we employ a TB wire model, where the base pairs are the sites of the chain, to study the spectral and charge transfer properties of periodic sequences with increasing repetition unit, as well as deterministic aperiodic DNA segments. Next, in Chapter 4, we addressed the impact of structural flexibility (dynamics) on the electronic structure and charge transfer ability of B-DNA. To this end, we applied our LCAO method to 20 AA and GG dimers, extracted from representative structures in a classical MD trajectory of a 20mer, and studied some useful physical quantities. In Chapter 5, we move on to study charge transport along DNA molecules, using the time-independent Schrödinger equation together with the transfer-matrix method in order to finally obtain I-V curves. We examine ideal and natural geometries concerning two categories of mutations: (i) DNA sequences that contain point substitution mutations, and (ii) sequences extracted from segments of human chromosomes, modified by expansion of the CAG triplet to mimic diseases. Finally, in Chapter 6 we state our overall conclusion and set some perspectives for future research.

2 | ELECTRONIC STRUCTURE[1]

The determination and study of the electronic structure of organic heterocyclic molecules have been of crucial interest for the scientific community for decades, especially since the establishment of quantum-mechanics-based methods. This includes the electronic structure and properties of nucleic acid oligomers, polymers, DNA, RNA and similar molecules that are important from a biological point of view. A large motivation for this pursuit is the radical effects that ultraviolet radiation has on the chemical integrity of DNA and RNA nucleic acid monomers and polymers [28, 29] . The π molecular electronic structure of planar molecules with sp^2 hybridization is involved in a variety of biological functions, e.g. in molecules like vitamin A, vitamin D precursors, and carotene, molecules containing polyene chromophores [30], chlorophyll in photosynthesis, and the retinal molecule involved in vision or in photon-driven ion pumps [31]. The electronic structure of single DNA molecules has been resolved by transverse scanning tunnelling spectroscopy and assigned to groups of orbitals originating from the molecular entities, i.e., nucleobases, backbone and counterions [24]. The research for the understanding of the electronic structure of DNA bases, RNA bases and similar molecules has led to the evaluation of the ionization and excitation energies.

Up to now, many calculations of ionization and excitation energies of nucleic acid bases with ab initio methods have been reported. However, theoretical predictions, commonly based on molecular quantum mechanics, calculations such as density functional theory (DFT) or Hartree - Fock, are still somehow not fully conclusive. Specifically, ionization energies of the nucleic acid bases have been calculated using restricted open shell Hartree - Fock (ROHF) [32, 33], restricted open shell Möller - Plesset second-order perturbation theory (MP2) [33], DFT [34--37], semi-empirical multiconfiguration configuration interaction (AM1- MCCI) [38] and partial third-order electron propagator (P3) [39] methods. Besides the above, photoionization and photoelectron spectroscopy and related ab initio quantum calculations have been performed on the gas phase adenine, thymine, uracil, cytosine and guanine [40]. Various other methods have also been employed in order to get DNA's bases ionization energies [41, 42]. As for the excitation energies the resulting spectrum is rather complex due to the presence of several heteroatoms with lone pairs, that leads to the existence of a number of low-lying $n \to \pi^*$ and $n \to \sigma^*$ transitions in addition to the $\pi \to \pi^*$ transitions. Calculations of the vertical excitation spectra of the bases A, T, G and C have been performed by employing the complete active space self consistent field (CASSCF) [43], complete active space second-order perturbation theory (CASPT2) [44--47], configuration interaction single

---

[1]Part of this Chapter can be found published in Ref. [27], under CC BY 4.0

(CIS) [46, 48], random phase approximation (RPA) [43, 49], configuration interaction by perturbation with multiconfigurational zeroth-order wave functions selected by iterative process (CIPSI) [50], time-dependent density functional theory (TDDFT) [51] methods, an asymptotic extrapolation scheme with hybrid CC and molecular dynamics simulations [52] and CR-EOMCCSD(T) [53--55]. Some excited-state geometry optimizations have also been performed for the lowest π^* and n^* states [46, 50, 51]. Generally, the lack of an adequate number of experimental data for single DNA molecules makes the establishment of the theories more complex. Experimental data such as optical spectra, in which a structure of one broad peak of the coupled levels is observed [56, 57], are not refined enough to determine the single-molecule electronic structure [24].

In the present PhD theses, we have calculated the ionization and excitation energies of nucleic acid bases and similar biologically important molecules as well as assemblies of DNA bases using a semi-empirical Linear Combination of Atomic Orbitals (LCAO) method. First, we present past efforts to obtain $\pi \to \pi^*$ excitation and π ionization energies using a semi-empirical Hückel-type model that employs $2\mathrm{p}_z$ valence orbitals [58--61] with the parameterization of Hawke et al. [62, 63] as well as with the parameterization introduced by Mantela et al. [55]. We extend this model in order to include all valence orbitals, i.e., 2s, $2\mathrm{p}_x$, $2\mathrm{p}_y$, $2\mathrm{p}_z$ orbitals for C, N, and O atoms and 1s orbital for H atoms [27]. We investigate the electronic structure of the four DNA bases A, T, G, C and of the two Watson-Crick H-bonded pairs A-T and G-C. We focus on the HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) wave functions and energies. Regarding the diagonal matrix elements (also known as on-site energies), we introduce a novel parameterization. For the non-diagonal matrix elements referring to neighboring atoms, we employ the Slater - Koster two-center interaction transfer integrals. We use Harrison-type expressions with factors slightly modified relative to the original. We calculate the matrix elements for ideal geometries, namely for planar bases and base pairs separated and twisted approximately by 3.4 Å and 36°, respectively, relative to the double helix growth axis. Finally, the deformed base pairs pruned from several snapshots of a 500 ns Molecular Dynamics (MD) trajectory of a 20mer [18] are used in order to address the effects of structural variability on the electronic structure of B-DNA within the LCAO approach. Taking into account all valence orbitals, we are in the position to treat deflection from the planar geometry, e.g., DNA structural variability, a task impossible for the plane Hückel approach (i.e., using only $2\mathrm{p}_z$ orbitals). We compare our LCAO predictions for the ionization and excitation energies of heterocycles with those obtained from Ionization Potential Equation of Motion Coupled Cluster with Singles and Doubles (IP-EOMCCSD)/aug-cc-pVDZ level of theory and Completely Normalized Equation of Motion Coupled Cluster with Singles, Doubles, and non-iterative Triples (CR-EOMCCSD(T))/aug-cc-pVDZ level of theory, respectively (vertical values), as well as with available experimental data, where available [27, 55]. We should state that for almost half of the molecules studied in the present work, we could not find in the literature experimental or theoretical-computational data for the ionization and the excitation energies, which was part of our motivation

for this study. The other part is that we needed a model that could provide reasonable estimates of the lowest $\pi \to \pi^*$ excitation and π ionization energies and be less demanding in CPU cycles and memory compared to coupled cluster (CC) methods in the same time.

This chapter is organized as follows: In Section 2.1 we develop the LCAO method for: a) nucleic acid bases and similar heterocycles using only p_z valence orbitals (Subsection 2.1.1) and, b) nucleic acid bases and similar heterocycles using all valence orbitals (Subsection 2.1.2). In Section 2.2 we describe our method to determine the electronic structure of B-DNA base pairs employing two approximations: a) the Linear Combination of Molecular Orbitals (LCMO) using only the p_z valence orbitals and treating the base pair like a superposition of its two bases (Subsection 2.2.1) and, b) the LCAO using all valence orbitals and treating the base pair as a superposition of all of its atoms (Subsection 2.2.2). In Section 2.3 we present our results on ionization and excitation energies of various heterocyclic planar molecules, isolated DNA bases and base pairs and state our conclusions.

2.1 Description at the Molecular Level

The LCAO method was developed after many efforts of the physics and chemistry scientists to explain and describe the nature of chemical bond. This model provides quantitative and qualitative information about the chemical bonds, the stereochemistry, the dimensions and a variety of spectroscopic properties of molecules. The main concept of the LCAO method is to express the molecular wavefunction $\psi(\vec{r})$ as a linear combination of the atomic wavefunctions. In a defined area around each atomic nucleus that is involved in a bond, the solution of Schrödinger equation is approximately the atomic wavefunction. Thus, within the limits of LCAO method, the wavefunction of a chemical bond is considered a superposition - linear combination - of the corresponding atomic orbitals.

A molecular orbital can be written as:

$$\psi(\vec{r}) = \sum_{\nu=1}^{N} \sum_{i=1}^{I} c_{i\nu} \phi_{i\nu}(\vec{r}). \tag{2.1}$$

The index ν runs among all N atoms of the molecule and the index i runs among all I orbitals of each atom, respectively. $\psi(\vec{r})$ obeys the Schrödinger equation

$$\hat{H}^b \psi(\vec{r}) = E^b \psi(\vec{r}). \tag{2.2}$$

$\hat{H}^b$ is the Hamiltonian of the base (or other molecule), with eigenvalues E_k^b and eigenvectors $c_{i\nu,k}$. Multiplying with $\phi_{j\mu}^*$ and integrating, Equation (2.2) gives:

$$\sum_{\nu} \sum_{i} c_{i\nu} \int d^3r \ \phi_{j\mu}(\vec{r})^* \hat{H}^b \phi_{i\nu}(\vec{r}) = E^b \sum_{\nu} \sum_{i} c_{i\nu} \int d^3r \phi_{j\mu}(\vec{r})^* \phi_{i\nu}(\vec{r}), \tag{2.3}$$

that is the linear system of equations:

$$\sum_{\nu=1}^{N} \sum_{i=1}^{I} \left[\left(H_{j\mu i\nu}^b - E^b S_{j\mu i\nu} \right) c_{i\nu} \right] = 0, \quad \mu = 1, \dots, N, \quad j = 1, \dots, I. \tag{2.4}$$

The Hamiltonian matrix elements $H_{j\mu i\nu}^b$ are given by

$$H_{j\mu i\nu}^b = \langle \phi_{j\mu} | \hat{H}^b | \phi_{i\nu} \rangle \tag{2.5}$$

and the overlap matrix elements (Appendix A) are

$$S_{j\mu i\nu} = \langle \phi_{j\mu} | \phi_{i\nu} \rangle \approx \delta_{j\mu i\nu}. \tag{2.6}$$

We approximate $S_{j\mu i\nu}$ by $\delta_{j\mu i\nu}$, supposing that the atomic orbital wavefunctions are tightly bound to their atoms. The system of Equation (2.4) is solved by numerical diagonalization, giving the eigenenergies E_k^b and eigenvectors $c_{i\nu,k}$.

From this point, we proceed by developing two different rationales on the LCAO method: in the first case, we continue using only p_z valence orbitals in our calculations (Subsection 2.1.1), while in the second case, we include all valence orbitals, i.e., 2s, 2p_x, 2p_y, 2p_z orbitals for C, N, and O atoms and 1s

orbital for H atoms (Subsection 2.1.2).

2.1.1 LCAO with 2p$_z$ Valence Orbitals

Atoms in planar organic molecules are bonded with sp^2 hybridization, while having their 2p$_z$ atomic orbitals perpendicular to the molecular plane where their electrons are eventually delocalized. Thus, in this first approximation, the corresponding molecular wavefunction is a linear combination of the 2p$_z$ atomic orbitals from each atom. A π molecular single-electron wave function can be approximated as:

$$\psi(\vec{r}) = \sum_{\nu=1}^{N} c_\nu \phi_\nu(\vec{r}), \tag{2.7}$$

with ν running among all N atoms of the molecule. Following the same procedure as in Section 2.1, we end up with the linear system of equations:

$$\sum_{\nu=1}^{N} \left[\left(H_{\mu\nu}^b - E^b \, S_{\mu\nu} \right) c_\nu \right] = 0, \quad \mu = 1, \ldots, N. \tag{2.8}$$

The Hamiltonian matrix elements $H_{\mu\nu}^b$ are now given by

$$H_{\mu\nu}^b = \langle \phi_\mu | \, \hat{H}^b \, | \phi_\nu \rangle \tag{2.9}$$

and the overlap matrix elements by

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle \approx \delta_{\mu\nu}. \tag{2.10}$$

Thus, Eq. (2.8) becomes:

$$\sum_{\nu=1}^{N} \left[\left(H_{\mu\nu}^b - E^b \, \delta_{\mu\nu} \right) c_\nu \right] = 0, \quad \mu = 1, \ldots, N, \tag{2.11}$$

that is trivially solved by Hamiltonian diagonalization (by the **eig** command in MATLAB programming language, using LAPACK functions, see Appendix B).

To this end, we need the values of the Hamiltonian matrix elements, $H_{\mu\nu}^b$. Within the limits of LCAO approximation using only the p$_z$ electrons we consider:

$$H_{\mu\nu}^b = \begin{cases} E_x, & \mu = \nu \\ 0, & \mu \neq \nu, \text{not sp}^2 \text{ bonded atoms} \\ V_{pp\pi} & \mu \neq \nu, \text{sp}^2 \text{ bonded atoms} \end{cases} \tag{2.12}$$

The HKS parameterization [62, 63] proposes,

$$V_{pp\pi} = \chi \frac{\hbar^2}{md^2}, \tag{2.13}$$

with m being the electron mass, d being the two-center distance, $\chi = -0.63$ and for the on-site energies E_x, the use of empirical parameters: $E_C = -6.7$ eV

for carbon 2p$_z$ orbital, $E_{\mathrm{N2}} = -7.9$ eV for nitrogen with one electron in the 2p$_z$ atomic orbital (i.e., with coordination number 2), $E_{\mathrm{N3}} = -10.9$ eV for nitrogen with two electrons in the 2p$_z$ atomic orbital (i.e. with coordination number 3) and $E_{\mathrm{O}} = -11.8$ eV for oxygen 2p$_z$ orbital. In a quest to decrease the deviation between the experiment and the Hückel-type model predictions, we use the more recent MMTS parameterization [55]. The optimization introduced there, results in $\chi = -0.77$, $E_{\mathrm{C}} = -6.56$ eV, $E_{\mathrm{N2}} = -9.62$ eV, $E_{\mathrm{N3}} = -11.48$, and $E_{\mathrm{O}} = -10.35$ eV.

2.1.2 LCAO with All Valence Orbitals

In order to treat deflection from the planar geometry, e.g. DNA structural variability, we need to take into account at least all valence orbitals of a molecule. The aforementioned plane Hückel approach (i.e., using only 2p$_z$ orbitals) is inadequate in this case. Thus, we consider the molecular wavefunction as a linear combination of all valence orbitals, i.e., 2s, 2p$_x$, 2p$_y$, 2p$_z$ for C, N and O atoms, and 1s for H atoms, as in Eq. (2.1). Following the concept of Section 2.1, we obtain the system of Eqs. (2.4) which can be solved by numerical diagonalization (by the **eig** command in MATLAB programming language, using LAPACK functions). The occupied and unoccupied orbitals - and thus the HOMO and LUMO - can be found by counting all valence electrons contributed by the atoms of the molecule and arranging them successively in couples of different spin in accordance with the Pauli principle.

In this case, we need a new set of parameters for the Hamiltonian matrix elements, $H^b_{j\mu i\nu}$ [27]. Regarding the diagonal matrix elements $H^b_{i\nu i\nu}$ we utilize a novel parameterization, namely: $E_{\mathrm{H(1s)}} = -13.64$ eV for H 1s orbitals, $E_{\mathrm{C(2s)}} = -13.18$ eV for C 2s orbitals, $E_{\mathrm{C(2p)}} = -6.70$ eV for C 2p orbitals, $E_{\mathrm{N(2s)}} = -14.51$ eV for N 2s orbitals, $E_{\mathrm{N(2p)}} = -9.55$ eV for N 2p orbitals, $E_{\mathrm{O(2s)}} = -15.03$ eV for O 2s orbitals, $E_{\mathrm{O(2p)}} = -11.52$ eV for O 2p orbitals. For the non-diagonal matrix elements $H^b_{j\mu i\nu}$ ($\mu \neq \nu$) referring to neighbouring atoms, we utilize the Slater - Koster two-center interaction integrals [64]

$$V_{\mathrm{ss}} = V_{\mathrm{ss}\sigma}, \tag{2.14}$$

$$V_{\mathrm{sx}} = \xi_1 V_{\mathrm{sp}\sigma}, \tag{2.15}$$

$$V_{\mathrm{xx}} = \xi_1^2 V_{\mathrm{pp}\sigma} + (1 - \xi_1^2) V_{\mathrm{pp}\pi}, \tag{2.16}$$

$$V_{\mathrm{xy}} = \xi_1 \xi_2 (V_{\mathrm{pp}\sigma} - V_{\mathrm{pp}\pi}), \tag{2.17}$$

with ξ_1, ξ_2 being the directional cosines of $\vec{d} = \vec{ji}$ which points from atom i to atom j. Concerning the values of $V_{\mathrm{ss}\sigma}, V_{\mathrm{sp}\sigma}, V_{\mathrm{pp}\sigma}, V_{\mathrm{pp}\pi}$, we use the relevant expressions proposed by Harrison [65, 66], of the form:

$$V_\chi = \chi \frac{\hbar^2}{md^2}, \tag{2.18}$$

with m being the electron mass and d the two-center distance. The χ values we propose, are: $\chi_{\mathrm{ss}\sigma} = -1.32$, $\chi_{\mathrm{sp}\sigma} = -1.42$, $\chi_{\mathrm{pp}\pi} = -0.73$ (slightly modified

relative to the original Harrison constant), $\chi_{\mathrm{pp}\sigma} = 2.22$. For each H orbital the interactions are multiplied by a factor $b = 0.70$ that resulted from the optimization. We arrived at the above parameterization after careful optimization, by fitting the LCAO numerical results with the experimental values for the excitation and the ionization energies of nucleic acid bases A, G, T, C, and U. To do so, we used the Nelder - Mead algorithm as implemented in Matlab software, in order to minimise the sum of the squared differences between the experimental and the predicted values for the excitation and the ionization energies. All other non-diagonal matrix elements, referring to non-neighboring atoms, are assumed equal to zero, $H^b_{j\mu i\nu} = 0$. Tables 2.1 and 2.2 summarize our LCAO parameters.

Table 2.1: Diagonal matrix elements (on-site energies) in our LCAO parameterization (eV).

$E_{\mathbf{H(1s)}}$	$E_{\mathbf{C(2s)}}$	$E_{\mathbf{C(2p)}}$	$E_{\mathbf{N(2s)}}$	$E_{\mathbf{N(2p)}}$	$E_{\mathbf{O(2s)}}$	$E_{\mathbf{O(2p)}}$
-13.64	-13.18	-6.70	-14.51	-9.55	-15.03	-11.52

Table 2.2: χ values of Harrison-type expressions for non-diagonal matrix elements, utilizing Slater - Koster two-center interaction transfer integrals and the correction factor for interactions involving H atoms, in our LCAO parameterization.

$\chi_{\mathbf{ss}\sigma}$	$\chi_{\mathbf{sp}\sigma}$	$\chi_{\mathbf{pp}\pi}$	$\chi_{\mathbf{pp}\sigma}$	b
-1.32	-1.42	-0.73	2.22	0.70

2.2   Description at the Base-pair Level

In this section, we present the procedure to obtain the HOMO and LUMO states of a B-DNA base pair or *monomer*. We employ two different methods: a) in the first case, we treat the base pair like two adjacent molecules (bases) with electronic overlap, employing the Linear Combination of Molecular Orbitals (LCMO) method, and b) in the second case, we consider the base pair as a superposition of its atoms, as we already did in Section 2.1, using LCAO with all valence orbitals but with proper adjustments.

2.2.1   LCMO at B-DNA Base-pair Level

The rationale behind this method is that each DNA base pair has its bases (A and T, or G and C) connected with non-covalent hydrogen bonds. The length of a hydrogen bond is around 3 Å [67], longer than the typical length of the covalent bond connecting neighbouring atoms within a base, which is around 1.3 - 1.5 Å. The terms HOMO and LUMO are still in use for the base-pairs, to describe the single-electron wave functions that represent the highest in energy occupied orbital and the lowest in energy unoccupied orbital, respectively, of the molecular complex.

The base-pair HOMO/LUMO (H/L) wavefunction can be expressed as:

$$\psi^{bp}_{H/L}(\vec{r}) = C_1\,\psi^{b(1)}_{H/L}(\vec{r}) + C_2\,\psi^{b(2)}_{H/L}(\vec{r}), \tag{2.19}$$

where $\psi^{b(1)}_{H/L}(\vec{r})$, $\psi^{b(2)}_{H/L}(\vec{r})$ are the corresponding HOMO/LUMO orbitals of the two bases forming the base pair. Equation (2.19) is inserted into Schrödinger's equation:

$$\hat{H}^{bp}\psi^{bp}_{H/L}(\vec{r}) = E^{bp}_{H/L}\psi^{bp}_{H/L}(\vec{r}), \tag{2.20}$$

where $E^{bp}_{H/L}$ is the HOMO/LUMO on-site energy of the base pair. We get:

$$\hat{H}^{bp}[C_1\,\psi^{b(1)}_{H/L}(\vec{r}) + C_2\,\psi^{b(2)}_{H/L}(\vec{r})] = E^{bp}_{H/L}[C_1\,\psi^{b(1)}_{H/L}(\vec{r}) + C_2\,\psi^{b(2)}_{H/L}(\vec{r})]. \tag{2.21}$$

Multiplying Eq. (2.21) with $\psi^{b(1)*}_{H/L}(\vec{r})$ and integrating all over space, we get:

$$C_1 \int d^3\vec{r}\,\psi^{b(1)*}_{H/L}(\vec{r})\,\hat{H}^{bp}\,\psi^{b(1)}_{H/L}(\vec{r}) + C_2 \int d^3\vec{r}\,\psi^{b(1)*}_{H/L}(\vec{r})\,\hat{H}^{bp}\,\psi^{b(2)}_{H/L}(\vec{r}) =$$

$$E^{bp}_{H/L}\,C_1 \int d^3\vec{r}\,\psi^{b(1)*}_{H/L}(\vec{r})\,\psi^{b(1)}_{H/L}(\vec{r}) + E^{bp}_{H/L}\,C_2 \int d^3\vec{r}\,\psi^{b(1)*}_{H/L}(\vec{r})\,\psi^{b(2)}_{H/L}(\vec{r}). \tag{2.22}$$

In this approximation, we employed only the $2p_z$ atomic orbitals (see Subsection 2.1.1). Supposing that the atomic $2p_z$ orbital wavefunctions are tightly bound to their atoms, we obtain vanishing overlap integral between $\psi^{b(1)*}_{H/L}(\vec{r})$

and $\psi_{H/L}^{b(2)}(\vec{r})$, i.e.,

$$\int d^3\vec{r} \, \psi_{H/L}^{b(1)*}(\vec{r}) \, \psi_{H/L}^{b(2)}(\vec{r}) = \sum_{\nu=1}^{N_1} \sum_{\mu=1}^{N_2} c_{H/L,\nu}^{b(1)*} \, c_{H/L,\mu}^{b(2)} \int d^3\vec{r} \, \phi_\nu^{b(1)*} \, \phi_\mu^{b(2)} = 0. \quad (2.23)$$

According to the basic TB hypothesis, supposing that the Hamiltonian of the base pair is approximated by the Hamiltonian of base1 close to base 1, we obtain:

$$\int d^3\vec{r} \, \psi_{H/L}^{b(1)*}(\vec{r}) \, \hat{H}^{bp} \, \psi_{H/L}^{b(1)}(\vec{r}) \simeq \int d^3\vec{r} \, \psi_{H/L}^{b(1)*}(\vec{r}) \, \hat{H}^{b(1)} \, \psi_{H/L}^{b(1)}(\vec{r}) = E^{b(1)}, \quad (2.24)$$

We define:

$$t_{H/L} = \int d^3\vec{r} \, \psi_{H/L}^{b(1)*}(\vec{r}) \, \hat{H}^{bp} \, \psi_{H/L}^{b(2)}(\vec{r}). \quad (2.25)$$

Substituting Eq. (2.23), Eq. (2.24) and Eq. (2.25) into Eq. (2.22) we get:

$$E_{H/L}^{b(1)} C_1 + t_{H/L} C_2 = E_{H/L}^{bp} C_1. \quad (2.26)$$

In a similar procedure for the base b(2) we get:

$$t_{H/L}^* C_1 + E_{H/L}^{b(2)} C_2 = E_{H/L}^{bp} C_2. \quad (2.27)$$

Also,

$$t_{H/L} = \int d^3\vec{r} \, \psi_{H/L}^{b(1)*}(\vec{r}) \, \hat{H}^{bp} \, \psi_{H/L}^{b(2)}(\vec{r}) \Longrightarrow$$

$$t_{H/L} = \sum_{\nu=1}^{N_1} \sum_{\mu=1}^{N_2} c_{H/L,\nu}^{b(1)*} \, c_{H/L,\mu}^{b(2)} \int d^3\vec{r} \, \phi_\nu^{b(1)*}(\vec{r}) \, H^{bp} \, \phi_\mu^{b(2)}(\vec{r}) \Longrightarrow$$

$$t_{H/L} = \sum_{\nu=1}^{N_1} \sum_{\mu=1}^{N_2} c_{H/L,\nu}^{b(1)*} \, c_{H/L,\mu}^{b(2)} V_{\nu\mu}, \quad (2.28)$$

and $t_{H/L}^* = t_{H/L}$. We arrive at the system of equations:

$$E_{H/L}^{b(1)} C_1 + t_{H/L} C_2 = E_{H/L}^{bp} C_1$$

$$t_{H/L} C_1 + E_{H/L}^{b(2)} C_2 = E_{H/L}^{bp} C_2. \quad (2.29)$$

The solution of this system provides the eigenenergies of the base pair:

$$E_{H/L}^{bp} = \frac{E_{H/L}^{b(1)} + E_{H/L}^{b(2)}}{2} \pm \sqrt{\left(\frac{E_{H/L}^{b(1)} - E_{H/L}^{b(2)}}{2} \right)^2 + t_{H/L}^2}. \quad (2.30)$$

The matrix elements $V_{\nu\mu}$ are provided through the Slater-Koster expression [64, 68]:

$$V_{\nu\mu} = V_{pp\sigma} \sin^2 \phi + V_{pp\pi} \cos^2 \phi, \quad (2.31)$$

where ϕ denotes the angle formed by the line connecting atoms ν and μ and the plane perpendicular to 2p$_z$ orbitals. For intra-base covalently bonded atoms, the matrix elements $V_{\nu\mu} = V_{pp\pi}$ ($\phi = 0$) are given by the Harrison expression, that is valid for interatomic distances of the order of covalent bond. When dealing with distances between atoms belonging to different molecules, the Harrison expression is replaced with the exponentially decaying form:

$$V_{pp\pi} = -0.63 \, \frac{\hbar^2}{md_0^2} \, e^{-\frac{2}{d_0}(d-d_0)}, \tag{2.32}$$

where $d_0 \simeq 1.35 \mathring{A}$ is a typical covalent bond length within a base.

Having determined the matrix elements $V_{\nu\mu}$, then interaction integrals $t_{H/L}$ are calculated through Eq. (2.28). Through Eqs. (2.30) we calculate the HOMO/LUMO eigenenergy of the base pair, while the corresponding wavefunction is determined by Eq.(2.19) after obtaining the coefficients $\mathcal{C}_1$, $\mathcal{C}_2$ by the system of Eqs. (2.29).

2.2.2   LCAO at B-DNA Base-pair Level

Here, we follow a different procedure in order to obtain the HOMO and LUMO states of a B-DNA monomer. We consider the base pair as a superposition of its atoms, just like we did for the molecules in Subsection 2.1.2. Let us call N_1, N_2 the number of atoms making up the two bases of the base pair. We consider the base pair wavefunction $\psi^{bp}(\vec{r})$ as a linear combination of all valence orbital wavefunctions $\phi_{i\nu}(\vec{r})$, i.e., 2s, 2p$_x$, 2p$_y$, 2p$_z$ for C, N and O atoms and 1s for H atoms:

$$\psi^{bp}(\vec{r}) = \sum_{\nu=1}^{N_1+N_2} \sum_{i=1}^{I} c_{i\nu}\phi_{i\nu}(\vec{r}). \tag{2.33}$$

The indexes ν and i run among the $N_1 + N_2$ atoms of the base pair and the I orbitals of each atom, respectively. $\psi^{bp}(\vec{r})$ obeys the Schrödinger equation:

$$\hat{H}^{bp}\psi^{bp}(\vec{r}) = E^{bp}\psi^{bp}(\vec{r}). \tag{2.34}$$

Once again, we employ the concept of Section 2.1 to obtain the system of equations:

$$\sum_{\nu=1}^{N_1+N_2} \sum_{i=1}^{I} \left[\left(H_{j\mu i\nu}^{bp} - E^{bp} S_{j\mu i\nu} \right) c_{i\nu} \right] = 0, \quad \mu = 1, \ldots, N_1 + N_2, \quad j = 1, \ldots, I, \tag{2.35}$$

which can be solved by numerical diagonalization.

In this case, the values of the Hamiltonian matrix elements, $H_{j\mu i\nu}^{bp}$, are expressed slightly differently. The matrix elements with (a) $1 \leq \nu \leq N_1$ and $1 \leq \mu \leq N_1$, and (b) $N_1 + 1 \leq \nu \leq N_1 + N_2$ and $N_1 + 1 \leq \mu \leq N_1 + N_2$, are expressed in the same way as previously described in Subsection 2.1.2. For the remaining matrix elements, we employ the Slater-Koster two-center interaction

transfer integrals of Eqs. (2.14), (2.15), (2.16), (2.17) but in this case, the values of $V_{ss\sigma}, V_{sp\sigma}, V_{pp\sigma}, V_{pp\pi}$ are of the form

$$V_\chi = \chi \, \frac{\hbar^2}{md_0^2} \, e^{-\frac{2}{d_0}(d-d_0)},$$ (2.36)

where $d_0 = 1.35$ Å is a typical covalent bond distance within a base [27]. This difference stems from the fact that, as we already mentioned in Subsection 2.2.1, Harrison's relations are valid for interatomic distances of the size of covalent bonds, so when dealing with distances of the size of hydrogen bonds and longer, we should instead use the appropriate exponentially decaying expressions of the form of Equation (2.36) [69--71].

From the aforementioned diagonalization of the Hamiltonian matrix, we obtain the energy eigenvalues E_k^{bp} - including HOMO and LUMO - of the electronic spectrum, as well as the corresponding base-pair eigenvectors

$$c_k^{bp} = \begin{bmatrix} c_{11,k} \\ c_{12,k} \\ \vdots \\ c_{i\nu,k} \\ \vdots \\ c_{I(N_1+N_2),k} \end{bmatrix}.$$ (2.37)

2.3 Results and Discussion

In this section we present the results on ionization and excitation energies of various heterocyclic planar molecules, isolated DNA bases and base pairs. In the rest of this section, we make the convenient simplifying assumption that the HOMO absolute value expresses the ionization energy, and the HOMO - LUMO gap expresses the excitation energy (in most cases the first $\pi - \pi^*$ transition). In Subsection 2.3.1, are presented our results for a list of heterocyclic planar molecules, including isolated DNA bases, employing the LCAO method using all valence orbitals with our newly introduced parameterization (MSF) of Subsection 2.1.2 [27]. These results are compared to the corresponding results obtained from: a) the LCAO method using only $2p_z$ valence orbitals with the parameterizations of Subsection 2.1.1 introduced in [62, 63] (HKS) and [55] (MMTS), and b) ab-initio methods. Subsection 2.3.2 contains our results for the B-DNA base pairs, using the LCAO theoretical scheme described in Subsection 2.2.2 [27]. Our results are compared to the corresponding obtained from the LCMO method using only $2p_z$ valence orbitals described in Subsection 2.2.1 [63].

2.3.1 Heterocyclic Planar Molecules including Nucleic Acid Bases

The LCAO method using all valence orbitals with our newly introduced parameterization (MSF) of Subsection 2.1.2 was employed to calculate the HOMO and LUMO eigenenergies for a variety of heterocyclic planar organic molecules. Below, the ionization energies refer to π molecular orbital character and the excitation energies correspond to $\pi - \pi^*$ transitions, unless otherwise stated. We studied the following groups of molecules: adenine and isomers, guanine and isomers, purine and isomers, thymine, cytosine, uracil and isomers, pyrimidine and isomers, and other planar heterocyclic molecules. Table 2.3 summarizes our LCAO results using all valence orbitals, along with relevant experimental values (where available). I_{MMTS} and E_{MMTS} are the ionization and excitation energies obtained by the LCAO scheme using only $2p_z$ electrons with the MMTS [55] parameterization and the HKS [62, 63] parameterization (in parentheses) of Subsection 2.1.1. I_{CC} and E_{CC} are calculations of the vertical ionization energies at the Ionization Potential Coupled Cluster with Singles and Doubles (IP-EOMCCSD)/aug-cc-pVDZ level of theory and vertical excitation energies at the Completely Renormalised Equation of Motion Coupled Cluster with Singles, Doubles, and non-iterative Triples (CR-EOMCCSD(T))/aug-cc-pVDZ level of theory, respectively [55]. Table 2.3 also includes transition oscillator strengths f that we calculated in a simplistic approximation, considering point contribution of the corresponding orbitals. The transition dipole moment $\vec{d}$ was approximated as

$$\vec{d} = (-e)\langle L|\, \vec{r}\, |H\rangle = (-e)\left(\sum_{\nu=1}^{N}\sum_{i=1}^{I} c_{i\nu L}^* \langle \phi_{i\nu}|\right)\vec{r}\left(\sum_{\mu=1}^{N}\sum_{j=1}^{I} c_{j\mu H}\, |\phi_{j\mu}\rangle\right)$$

$$= (-e) \sum_{\nu=1}^{N} \sum_{i=1}^{I} \sum_{\mu=1}^{N} \sum_{j=1}^{I} c_{i\nu L}^{*} c_{j\mu H} \langle \phi_{i\nu} | \, \vec{r} \, | \phi_{j\mu} \rangle . \tag{2.38}$$

Since $|\phi_{i\nu}\rangle$, $|\phi_{j\mu}\rangle$ are not exactly known, the previous integral cannot be calculated. To make an estimation of d, we make the crude approximation:

$$\vec{d} \simeq (-e) \sum_{\nu=1}^{N} \sum_{i=1}^{I} c_{i\nu L}^{*} \vec{r_i} c_{i\nu H}, \tag{2.39}$$

where $|L\rangle$ ($|H\rangle$) is the LUMO (HOMO) state. In other words, we replace the integral over all space with a sum at the atom positions approximatively assuming that the wavefunction contribute mainly at this points. The oscillator strength is [72]

$$f = \frac{2}{3} \frac{m}{e^2 \hbar^2} E \, d^2. \tag{2.40}$$

E is the excitation energy. The results are illustrated in Figures 2.1, 2.2 and 2.3.

Table 2.3: Ionization and excitation energies (eV). I_{LCAO} and E_{LCAO} are the ionization and excitation energies obtained by our LCAO scheme, including all valence orbitals. f_{LCAO} is the relevant oscillator strength. I_{MMTS} and E_{MMTS} are the ionization and excitation energies obtained by the LCAO scheme using only $2p_z$ electrons with the MMTS [55] parameterization and the HKS [62, 63] parameterization (in parentheses) of Subsection 2.1.1. I_{CC} and E_{CC} are the energies calculated at the IP-EOMCCSD/aug-cc-pVDZ and CR-EOMCCSD(T)/aug-cc-pVDZ level of theory [55]. I_{exp} and E_{exp} are the experimental data. In parentheses, the character of the transition.

Name Formula	I_{LCAO}	$E_{LCAO}(f_{LCAO})$	I_{MMTS}	E_{MMTS}	I_{CC}	E_{CC}	I_{exp}	E_{exp}
Adenine $C_5H_5N_5$ (Isomer 1)	8.44	4.20 (0.330)	8.74 (8.25)	4.61 (3.79)	8.23	5.04	8.44 [73]	4.84 [74, 75]
2-Aminopurine $C_5H_5N_5$ (Isomer 2)	8.56	3.84 (0.239)	8.79 (8.25)	4.30 (3.61)	7.95	4.27		4.11 [76]
1H-pyrazolo[3,4-d] pyrimidin-4-amine $C_5H_5N_5$ (Isomer 3)	8.78	4.25 (0.328)	9.23 (8.37)	4.80 (3.77)	8.51	4.92		
Pyrimido [5,4-e]-as-triazine, 1,2-dihydro- $C_5H_5N_5$ (Isomer 4)	8.04	3.21 (0.282)	7.76 (7.44)	3.53 (2.77)	7.18	3.16		
Guanine $C_5H_5N_5O$ (Isomer 1)	8.36	4.25 (0.288)	8.29 (8.03)	4.51 (3.54)	7.83	4.77 ($\pi \to \sigma^*$) 4.85	8.24 [73]	4.51 [75]
7-Amino-S-triazolo(1,5-a) pyrimidin-5(4H)-one $C_5H_5N_5O$ (Isomer 2)	8.42	4.37 (0.285)	8.60 (8.52)	4.65 (3.90)	8.60	4.91		
Pyrimido[5,4-e]-as-triazin-5[6h]-one, 1,2-dihydro- $C_5H_5N_5O$ (Isomer 3)	8.19	3.42 (0.198)	7.18 (7.21)	2.83 (2.34)	6.68	2.54		
7H-imidazo[4,5-d]-v triazin-4-one, 6-methyl- $C_5H_5N_5O$ (Isomer 4)	8.93	3.64 (0.302)	8.77 (8.42)	3.46 (3.51)	8.92	4.47 ($n \to \sigma^*$) 4.55		
9H-purine $C_5H_4N_4$ (Isomer 1)	9.20	4.40 (0.313)	9.20 (8.63)	4.63 (3.88)	9.34	4.49 ($n \to \pi^*$) 4.92	9.52 [73]	4.28 [77] ($n \to \pi^*$) 4.68 [77]
7H-purine $C_5H_4N_4$ (Isomer 1 taut.)	9.08	4.26 (0.295)	9.29 (8.86)	4.63 (4.06)	9.34 (n) 9.40	4.36 ($n \to \pi^*$) 4.79		
1H-1,2,3-triazolo [4,5-b]pyridine $C_5H_4N_4$ (Isomer 2)	9.42	4.12 (0.340)	9.50 (8.89)	4.18 (3.91)	9.41	4.49 4.54		
[1,2,4]Triazolo [1,5-a]pyrazine $C_5H_4N_4$ (Isomer 3)	8.95	4.20 (0.230)	8.97 (8.27)	4.16 (3.41)	9.27	4.63		
[1,2,3]Triazolo [1,5-a]pyrazine $C_5H_4N_4$ (Isomer 4)	8.64	3.96 (0.172)	8.63 (7.99)	3.98 (3.19)	8.95	4.31		

Table 2.3: *Cont.*

Name Formula	I_{LCAO}	$E_{\text{LCAO}}(f_{\text{LCAO}})$	I_{MMTS}	E_{MMTS}	I_{CC}	E_{CC}	I_{exp}	E_{exp}
Thymine $C_5H_6N_2O_2$	9.09	4.77 (0.316)	8.99 (9.02)	4.90 (4.13)	9.03	5.07 ($n \to \pi^*$) 5.17	9.14 [73]	4.69 [75]
Cytosine $C_4H_5N_3O$	8.68	4.54 (0.306)	9.06 (8.89)	4.89 (4.20)	8.67	4.64	8.94 [73]	4.64 [75]
Uracil $C_4H_4N_2O_2$ *(Isomer 1)*	8.89	4.70 (0.286)	8.99 (9.02)	4.90 (4.14)	9.44	5.03 ($n \to \pi^*$) 5.27	9.50 [73]	4.79 [75, 77]
Pyrazine, 1,4-dioxide $C_4H_4N_2O_2$ *(Isomer 2)*	8.77	4.28 (0.403)	8.46 (9.26)	3.51 (3.52)	8.11	3.30	8.33 [78]	4.05 [79]
4(1H)-pyrimidinone, 6-hydroxy- $C_4H_4N_2O_2$ *(Isomer 3)*	9.01	4.95 (0.103)	8.60 (8.84)	4.84 (4.65)	9.66	5.29		
Maleic hydrazide $(C_4H_4N_2)_2$ *(Isomer 4)*	8.77	3.34 (0.113)	9.16 (9.47)	4.11 (3.79)	8.77	4.11		
Pyrazine $C_4H_4N_2$ *(Isomer 1)*	9.53	4.39 (0.258)	9.55 (9.17)	4.60 (4.50)	9.49 (n) 10.09	4.07 ($n \to \pi^*$) 4.88	9.63 [80] 10.18 [80]	4.20 [81] 4.79 [82, 83]
Pyrimidine $C_4H_4N_2$ *(Isomer 2)*	9.98	5.28 (0.249)	10.02 (9.40)	5.56 (4.85)	9.56 (n) 10.44	4.41 ($n \to \pi^*$) 4.84 ($n \to \pi^*$) 5.25	9.73 [80] 10.41 [80]	4.35 [81] 4.62 [82] 5.13 [75, 77, 82, 83]
Pyridazine $C_4H_4N_2$ *(Isomer 3)*	9.41 (n) 10.39	4.28 (0.000 ($n \to \pi^*$)) 5.26 (0.253)	10.47 (9.52)	5.54 (4.84)	9.07 (n) 10.59	3.76 ($n \to \pi^*$) 4.47 ($n \to \pi^*$) 5.12	9.31 [80] 10.61 [80]	3.70 [81] 5.00 [83]
1H-imidazole $C_3H_4N_2$ *(Isomer 1)*	8.80	4.97 (0.000 ($\pi \to \sigma^*$)) 5.77 (0.171)	8.83 (8.49)	5.76 (4.76)	8.90	5.50 ($\pi \to \sigma^*$) 6.29	8.96 [84]	5.99 [85]
1H-pyrazole $C_3H_4N_2$ *(Isomer 2)*	9.69 9.48	5.69 (0.000 ($\pi \to \sigma^*$)) 5.90 (0.000 ($\pi \to \sigma^*$)) 5.97 (0.196)	9.62 (8.80)	5.99 (4.84)	9.35	6.11 ($\pi \to \sigma^*$) 6.25	9.38 [86]	5.90 [87]
1H-benzimidazole $C_7H_6N_2$ *(Isomer 1)*	8.84	4.63 (0.245)	8.99 (8.52)	4.98 (3.97)	8.40	4.67	8.44 [84]	4.47 [88]
1H-indazole $C_7H_6N_2$ *(Isomer 2)*	8.41	3.85 (0.217)	8.76 (8.46)	4.32 (3.75)	8.26	4.50	8.35 [89]	4.27 [90]
2H-indazole $C_7H_6N_2$ *(Isomer 2 taut.)*	8.42	3.84 (0.229)	8.41 (7.87)	3.78 (2.87)	7.90	4.54		
1H-pyrrolo[2,3-b] pyridine $C_7H_6N_2$ *(Isomer 3)*	8.47	3.82 (0.184)	8.53 (8.28)	4.12 (3.59)	8.17	4.50	8.11 [91]	4.28 [92]
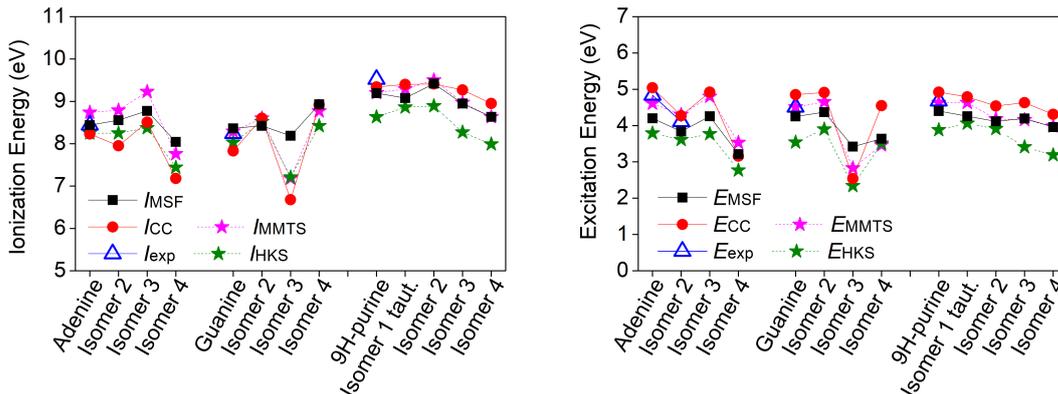
Figure 2.1: First π ionization energy and first $\pi - \pi^*$ excitation energy of purines calculated via our LCAO method using all valence orbitals, along with results at the IP-EOMCCSD/aug-cc-pVDZ (vertical ionization energies) and CR-EOMCCSD(T)/aug-cc-pVDZ (vertical excitation energies) level of theory [55], as well as available experimental data. Results employing Hückel model, using only $2p_z$ orbitals with MMTS [55] parameterization and HKS [62, 63] parameterization, are also included. Different isomers are specified in Table 2.1.
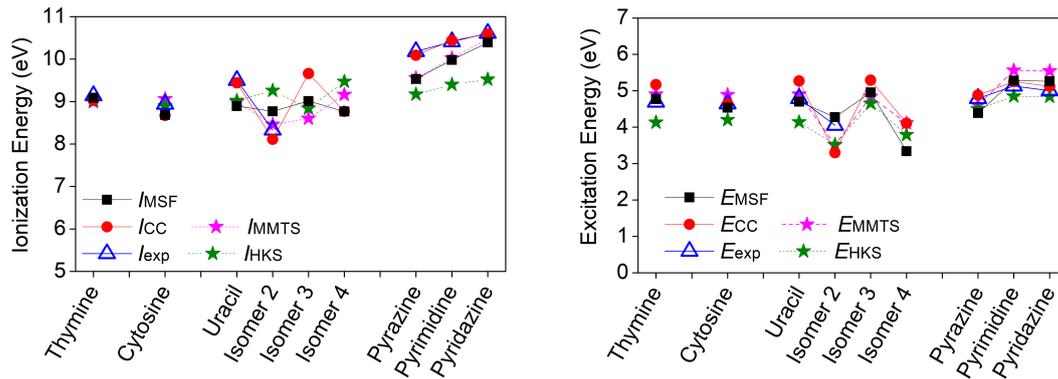
Figure 2.2: First π ionization energy and first $\pi - \pi^*$ excitation energy of pyrimidines calculated via our LCAO method using all valence orbitals, along with results at the IP-EOMCCSD/aug-cc-pVDZ (vertical ionization energies) and CR-EOMCCSD(T)/aug-cc-pVDZ (vertical excitation energies) level of theory [55], as well as available experimental data. Results employing Hückel model, using only $2p_z$ orbitals with MMTS [55] parameterization and HKS [62, 63] parameterization, are also included. Different isomers are specified in Table 2.1.
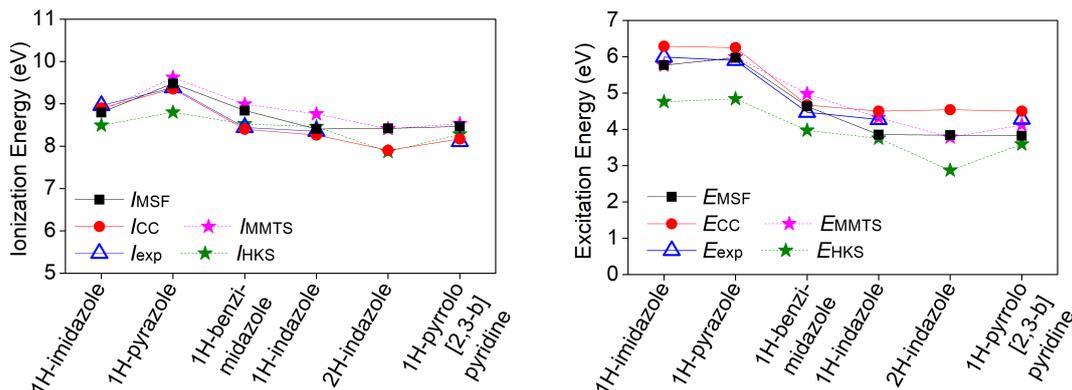
Figure 2.3: First π ionization energy and first $\pi - \pi^*$ excitation energy of other planar heterocyclic molecules calculated via our LCAO method using all valence orbitals, along with results calculated at the IP-EOMCCSD/aug-cc-pVDZ (vertical ionization energies) and CR-EOMCCSD(T)/aug-cc-pVDZ (vertical excitation energies) level of theory [55], as well as available experimental data. Results employing Hückel model, using only $2p_z$ orbitals with MMTS [55] parameterization and HKS [62, 63] parameterization, are also included. Different isomers are specified in Table 2.1.

Regarding the ionization energy, the LCAO (MSF) obtained results are in very good agreement with both the experimental data and the CC results, although there are some deviations. Also, our method using all valence orbitals gives more successful predictions with respect to the experimental results, than the previously used LCAO method using only $2p_z$ electrons, especially the one using the HKS [62, 63] parameterization. The Root Mean Square Percentage Error (RMSPE), with respect to the experimental values, is 3.65%. Differences in tautomer ionization energies are as expected negligible, that is 0.12 eV for purine tautomers and 0.01 eV for indazole tautomers. As for the excitation energies of the $\pi - \pi^*$ transition, the RMSPE, with respect to the experimental values, is 6.49%. Both purine and indazole tautomers have a negligible 0.03 eV difference in their excitation energies. Based on the presented data and reported comments about individual bases, we note that the LCAO method using all valence orbitals, though not exact, is capable of producing results in a good agreement with experimental data, when choosing the suitable set of parameters. Vertical ionization energies of nucleic acid bases in the gas phase with different electronic structure methods are, generally, in agreement with our results, cf. Ref. [93] and references therein.

2.3.2 B-DNA Base Pairs

In this subsection we present our results for the B-DNA base pairs, that are Adenine (A) - Thymine (T) and Guanine (G) - Cytosine (C). In Table 2.4 we show the HOMO, LUMO and HOMO - LUMO gap energies of these base pairs, according to the procedure described in Section 2.2.2 using LCAO with

all valence orbitals, along with the corresponding energies found in Ref. [63] using only $2p_z$ orbitals. At this point, we should state that the bases making up the base pairs are slightly deformed in comparison to their structure when isolated (cf. Section 2.3.1, Table 2.3), so the corresponding HOMO and LUMO energies for these two cases may differ. Thus, Table 2.4 contains additionally the HOMO, LUMO and HOMO - LUMO gap energies of the distorted bases. The HOMO (LUMO) energies are of π (π^*) molecular orbital character and the HOMO - LUMO gap energies are $\pi - \pi^*$ transitions, unless otherwise stated.

Table 2.4: HOMO ($E_{\mathrm{LCAO,H}}$) and LUMO ($E_{\mathrm{LCAO,L}}$) eigenenergies of the base pairs A-T and G-C, obtained in this work using LCAO with all valence orbitals, along with the corresponding HOMO-LUMO energy gaps ($E_{\mathrm{LCAO,g}}$) in eV (rows 6 and 7). Rows 2 - 5 contain the calculated HOMO and LUMO energies of each distorted base making up these base pairs [27]. The third, the fifth and the seventh column list the corresponding energies from Ref. [63] where only p_z orbitals had been used.

Base or Base Pair	$E_{\mathrm{LCAO,H}}$	E_{H} [63]	$E_{\mathrm{LCAO,L}}$	E_{L} [63]	$E_{\mathrm{LCAO,g}}$	E_{g} [63]
A	-8.50	-8.30	-4.19	-4.40	4.31	3.90
T	-9.12	-9.00	-4.30	-4.90	4.82	4.10
G	-8.31	-8.00	-4.12	-4.50	4.19	3.50
			-4.43 (σ^*)		4.24 ($\pi \to \sigma^*$)	
C	-8.67	-8.80	-4.11	-4.30	4.56	4.50
A-T	-8.49	-8.30	-4.31	-4.90	4.18	3.40
			-4.43 (σ^*)		3.87 ($\pi \to \sigma^*$)	
G-C	-8.30	-8.00	-4.14	-4.50	4.16	3.50

As expected, the energy values for the bases are slightly different compared to these of Table 2.3. Besides that, we observe that the HOMO energy of a particular base pair is very close to the highest of the HOMO energies of the two bases of the base pair, while the LUMO energy of the base pair is very close to the lowest of the two LUMO energies.

In order to have a more consistent view of the above observation, in Figures 2.4 and 2.5 we illustrate the occupation probabilities of holes (HOMO state) and electrons (LUMO state) on each atomic orbital of bases and base pairs, calculating the squared coefficients $|c_{i\nu}|^2$ (cf. Equations (2.1) and (2.33)) of the corresponding states. We observe that our calculated HOMO state for the base pair A-T (G-C) is localized almost totally in A (G), while the corresponding LUMO wave function is localized in T (C), in accordance to results from ab-initio techniques of Refs. [94, 95], which locate the HOMO of a base pair in purine and the LUMO in pyrimidine. This is due to the higher HOMO energy of A (G) and lower LUMO energy of T (C) and the large values of these differences.
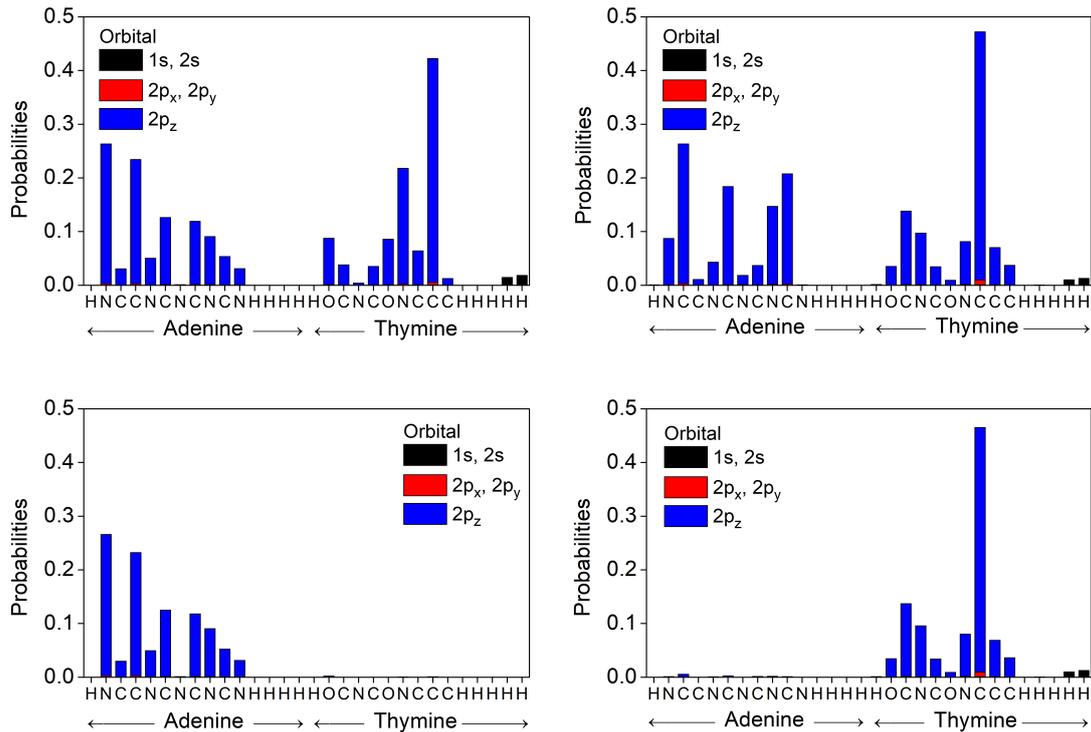
Figure 2.4: Occupation probabilities of each atomic orbital, $|c_{i\nu}|^2$ (cf. Equation (2.1)), for the HOMO (**left**) and LUMO (**right**) states of A and T bases within an A-T base pair (**top**), along with the corresponding probabilities (cf. Equation (2.33)) for the HOMO and LUMO states of the A-T base pair (**bottom**).
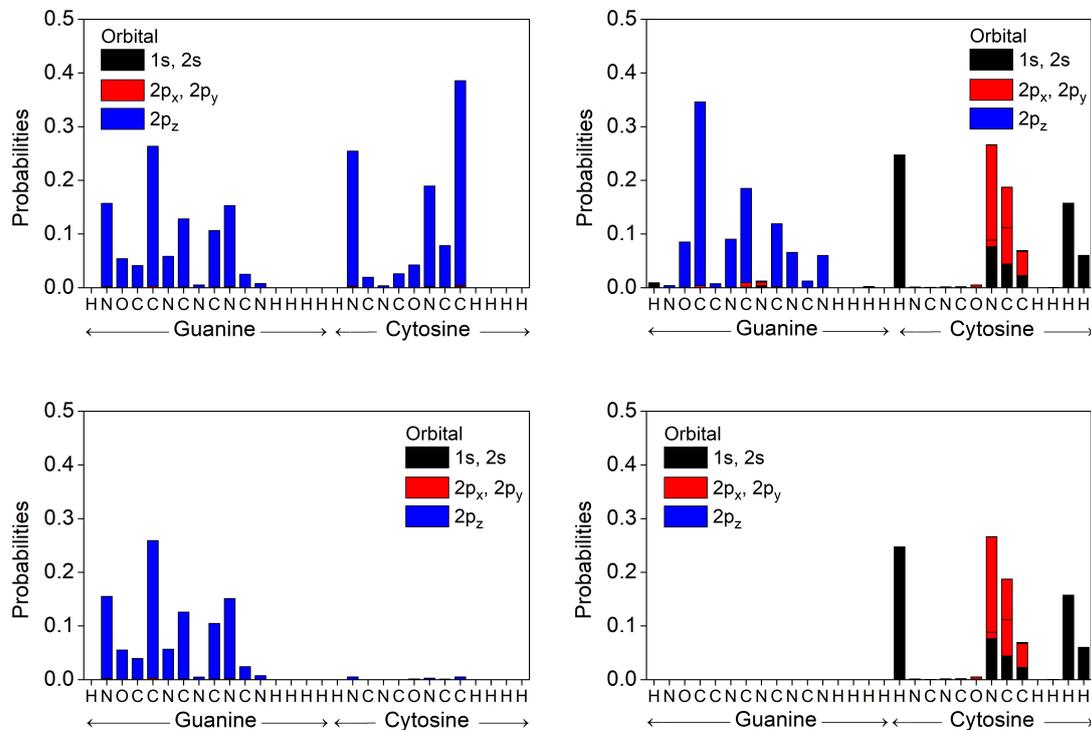
Figure 2.5: Occupation probabilities of each atomic orbital, $|c_{i\nu}|^2$ (cf. Equation (2.1)), for the HOMO (**left**) and LUMO (**right**) states of G and C bases within a G-C base pair (**top**), along with the corresponding probabilities (cf. Equation (2.33)) for the HOMO and LUMO states of the G-C base pair (**bottom**).

3 | Charge Transfer[1]

Besides the electronic structure of biological molecules, their charge transfer properties - especially these of DNA and RNA - have attracted considerable interest of a broad interdisciplinary scientific community [10, 23, 68, 98--103]. As already mentioned in Chapter 1, by the term *transfer* we imply that a carrier (electron or hole), created (e.g. by oxidation or reduction) or injected at a specific site, moves to a more favorable location, without application of an external gradient, e.g., of potential energy or temperature.

Charge movement is usually attributed to two types of mechanisms [104, 105]: (i) *incoherent* or *thermal* hopping between nearest neighboring or more distant sites and (ii) *coherent* hopping or *tunneling* or *superexchange*. The term *tunneling* implies quantum mechanical tunneling, between two sites, e.g., the carrier donor and the carrier acceptor, through a bridge. The term *superexchange*, not to be confused with the similar term in magnetism, emanates from the distant interaction between the two sites, e.g. the donor and the acceptor, through a bridge. The coherent mechanism is expected to dominate carrier movement in the low temperature regime. In natural DNA, it is more likely that a hole will be created at a guanine which has the highest HOMO of all bases and an electron will be created at a thymine which has the lowest LUMO of all bases [63]. However, coherently, if e.g. the hole is initially created or injected at an adenine, charge transfer will mainly be accomplished through adenines and similarly for other initial conditions [106]. Typically, in *coherent* transfer, charge is never exactly localized but there is a mean over time occupation probability to find it at each site, the carrier does not exchange energy with the environment during its transfer and this way it can travel short distances; strictly quantum mechanically, just a percentage of the carrier reaches the last site. Commonly, in *thermal* hopping, charge is localized. The carrier exchanges energy with the environment during its transfer and this way it can travel far longer than via the coherent mechanism.

Many mechanisms have been suggested to explain the charge movement along DNA. The base-pair stack of the DNA double helix creates a nearly one-dimensional π-pathway, formed by the overlap of π molecular orbitals of the stacked aromatic bases of DNA, that favors charge transfer. This π-pathway transfer can lead to charge propagation even at long distances [8, 68, 107, 108].

After many years of research [109--114], we realize that many factors influence carrier motion along DNA, including aqueousness, counterions, extraction process, electrodes, purity, substrate, structural fluctuations and geometry. These factors are either intrinsic or extrinsic. In the present Thesis, we focus on the

---

[1]The content of this Chapter can be found published in Refs. [27, 96], under CC BY 4.0, and in Ref. [97], ©2018 American Physical Society.

most important of the intrinsic factors, i.e. the effect of alternating the base-pair sequence, which affects the overlaps across the π-stack. We study rather long sequences, so we employ a Tight-Binding (TB) model which allows to address systems of realistic length [16, 106, 115--132]. Still, we aim to understand how the base-pair sequence affects carrier motion. Up to now, ab initio calculations [133--141], used to explore experimental results and the underlying mechanisms, are currently limited to short segments for computational reasons. Furthermore, research has shown that carrier movement through B-DNA can be manipulated. The carrier transfer rate through DNA can be tuned by chemical modification, using various natural and artificial nucleobases with different HOMO and LUMO levels [142]. This way it can be increased by many orders of magnitude with appropriate sequence choice [97, 106, 116, 143]. Furthermore, structural fluctuations is another factor which influences quantum transport through DNA molecular wires [144].

Up to now, there are several studies concerning charge and energy transfer in polymer systems [145--147]. Besides that, there are several works devoted to the study of transfer and transport in specific DNA structures using variants of the TB method [14, 104, 106, 116, 118, 119, 130, 148--150].

Here, we study the coherent regime for several cases of periodic and aperiodic polymers. More specifically, we employ a TB wire model, where the base pairs (monomers) are the sites of the chain, to study the spectral and charge transfer properties of periodic sequences with increasing repetition unit, as well as deterministic aperiodic [Thue-Morse (TM), Fibonacci (F), Double-Period (DP), Rudin-Shapiro (RS), Cantor Set (CS), Asymmetric Cantor Set (ACS)] DNA segments. The relevant parameters are the on-site energies of base pairs and the interaction integrals between successive base pairs. We also assume that the state or movement of an extra hole or electron in the polymer can be expressed through a combination of the HOMO or LUMO, respectively, of all monomers. This way, we define the HOMO regime and the LUMO regime. We have to solve a system of N coupled equations for the time-independent problem, and a system of N coupled first order differential equations for the time-dependent problem. We study HOMO and LUMO eigenspectra, HOMO-LUMO gaps and the relevant density of states, as well as the mean over time probabilities to find the carrier at each site. We are also interested in the frequency content of carrier movement, hence, we analyze the Fourier spectra of the time-dependent probability to find the carrier at each site. We calculate the weighted mean frequency of each monomer and the total weighted mean frequency of the polymer, as a measure of the overall transfer frequency content. Finally, we study the pure mean transfer rate from a certain site to another, which describes the easiness of charge transfer and gives us a measure of how much of the carrier is transferred and how fast this process is.

The rest of this chapter is organized as follows: In Section 3.1, we delineate the theory behind the time-independent (Subsection 3.1.1) and the time-dependent (Subsection 3.1.2) problem, as well as introduce the studied physical quantities (Subection 3.1.3). Section 3.2 provides some details on the studied periodic sequences with increasing repetition unit (Subsection 3.2.1) and the

studied deterministic aperiodic sequences (Subsection 3.2.2) and we outline our notation. In Section 3.3, after we present the result TB parameterization, we discuss our results for the aforementioned categories of studied polymers and state our conclusions. The content of this chapter can be found published in refs. [27, 96, 97].

3.1 The Tight-Binding Wire Model for Charge Transfer

Within the TB wire model description for charge transfer that we employ in this work, a site is a monomer (e.g., in DNA, a base pair). We assume that the state or movement of an inserted hole can be expressed through the monomer HOMOs, while these of an electron through LUMOs. We call λ the monomer index, $\lambda = 1, 2, \ldots, N$. In this approximation, the single carrier (hole/electron) wave function of the whole macromolecule (e.g. DNA) is considered as a linear combination of base-pair wave functions. In the following subsections, we will treat the time-independent and the time-dependent problem of the coherent charge transfer.

3.1.1 Stationary States - Time-independent Problem

The time-independent single carrier DNA wavefunction can be then expressed as a linear combination of base-pair wave functions with time-independent coefficients, i.e.,:

$$\psi^{DNA}(\vec{r}) = \sum_{\lambda} A_{\lambda} \psi_{\lambda}^{bp}(\vec{r}), \tag{3.1}$$

where $\psi_{\lambda}^{bp}(\vec{r})$ is the HOMO/LUMO wavefunction of the λ-th base pair.

Substituting Eq. (3.1) into the time-independent Schrödinger's equation

$$\hat{H}^{DNA} \psi^{DNA}(\vec{r}) = E^{DNA} \psi^{DNA}(\vec{r}), \tag{3.2}$$

we get:

$$\sum_{\lambda} A_{\lambda} \hat{H}^{DNA} \psi_{\lambda}^{bp}(\vec{r}) = E^{DNA} \sum_{\lambda} A_{\lambda} \psi_{\lambda}^{bp}(\vec{r}) \Longrightarrow$$

$$\sum_{\lambda} A_{\lambda} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \hat{H}^{DNA} \psi_{\lambda}^{bp}(\vec{r}) = E^{DNA} \sum_{\lambda} A_{\lambda} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \psi_{\lambda}^{bp}(\vec{r}) \Longrightarrow$$

$$A_{\lambda'-1} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \hat{H}^{DNA} \psi_{\lambda'-1}^{bp}(\vec{r}) + A_{\lambda'} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \hat{H}^{DNA} \psi_{\lambda'}^{bp}(\vec{r}) +$$

$$A_{\lambda'+1} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \hat{H}^{DNA} \psi_{\lambda'+1}^{bp}(\vec{r}) = E^{DNA} A_{\lambda'-1} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \psi_{\lambda'-1}^{bp}(\vec{r}) +$$

$$E^{DNA} A_{\lambda'} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \psi_{\lambda'}^{bp}(\vec{r}) + E^{DNA} A_{\lambda'+1} \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r}) \psi_{\lambda'+1}^{bp}(\vec{r}). \tag{3.3}$$

Within the TB method, we assume:

$$\int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r})\hat{H}^{DNA}\psi_{\lambda'}^{bp}(\vec{r}) \approx \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}(\vec{r})\hat{H}_{\lambda'}^{bp}\psi_{\lambda'}^{bp}(\vec{r}) = E_{\lambda'}^{bp}, \qquad (3.4)$$

with $E_{\lambda'}^{bp}$ the on-site energy of base pair λ' (see Table 2.4). Additionally:

$$t_{\lambda',\lambda}^{bp} = \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}\hat{H}^{DNA}\psi_{\lambda}^{bp}(\vec{r}) \qquad (3.5)$$

and

$$\int d^3\vec{r}\, \psi_{\lambda'}^{bp*}\psi_{\lambda}^{bp}(\vec{r}) = \delta_{\lambda'\lambda}. \qquad (3.6)$$

Substituting Eq. (3.4), Eq. (3.5) and Eq. (3.6) into Eq. (3.3), we arrive to a system of N coupled algebraic equations

$$E^{DNA}A_\lambda = E_\lambda^{bp}A_\lambda + t_{\lambda,\lambda-1}^{bp}A_{\lambda-1} + t_{\lambda,\lambda+1}^{bp}A_{\lambda+1}. \qquad (3.7)$$

Assuming that the base-pair wavefunction can be expressed as $\psi^{bp}(\vec{r}) = \sum_{\nu=1}^{N}\sum_{i=1}^{I}c_{i\nu}\phi_{i\nu}(\vec{r})$, then:

$$t_{\lambda',\lambda}^{bp} = \int d^3\vec{r}\, \psi_{\lambda'}^{bp*}\hat{H}^{DNA}\psi_{\lambda}^{bp}(\vec{r}) \implies$$

$$t_{\lambda,\lambda'}^{bp} = \sum_{\nu=1}^{N_\lambda}\sum_{i=1}^{I_\lambda}\sum_{\mu=1}^{N_{\lambda'}}\sum_{j=1}^{I_{\lambda'}} c_{i\nu(\lambda)}^* \, V_{i\nu j\mu}\, c_{j\mu(\lambda')}, \qquad (3.8)$$

where

$$V_{i\nu j\mu} = \langle\phi_{i\nu(\lambda)}|\,\hat{H}_{\text{DNA}}\,|\phi_{j\mu(\lambda')}\rangle. \qquad (3.9)$$

λ and λ' denote neighbouring base pairs, while ν, μ extend up to the total number of atoms N_λ and $N_{\lambda'}$ respectively, and i, j are running among all the orbitals I_λ and $I_{\lambda'}$ of each atom of base pair λ and λ', respectively. The matrix elements $V_{i\nu j\mu}$ are, once again, given by the Slater - Koster two-center interaction transfer integrals of Eqs. (2.14), (2.15), (2.16), (2.17) with the values of $V_{ss\sigma}, V_{sp\sigma}, V_{pp\sigma}, V_{pp\pi}$ being of the form of Eq. (2.36). Alternatively, for near-to-planar molecules and near-to-ideal geometries, we can employ the approximation that uses only $2p_z$ electrons, and transfer integrals will be

$$t_{bp(\lambda,\lambda')}^{H/L} = \sum_{\nu=1}^{N_\lambda}\sum_{\mu=1}^{N_{\lambda'}} C_{\nu(\lambda)}^{H/L*}\, V_{\nu\mu}\, C_{\mu(\lambda')}^{H/L}, \qquad (3.10)$$

where $C_{\nu(\lambda)}^{H/L*}$, $C_{\mu(\lambda')}^{H/L}$ are defined through the system of Eqs. (2.29). The matrix elements in this case will be of the form of Eq. (2.31), but with

$$V_{pp\pi} = 2.22\,\frac{\hbar^2}{md_0^2}\,e^{-\frac{2}{d_0}(d-d_0)}. \qquad (3.11)$$

We, also, obtain the maximum transfer percentage of the carrier from one base pair to another. This refers to the maximum probability to find the extra

hole or electron at the site where it was not placed at initially. The maximum transfer percentage reads

$$p = \frac{(2t_{\lambda,\lambda'}^{bp})^2}{(2t_{\lambda,\lambda'}^{bp})^2 + \Delta_{\lambda,\lambda'}^2} \tag{3.12}$$

where $t_{\lambda,\lambda'}^{bp}$ is the transfer parameter between the two base pairs and $\Delta_{\lambda,\lambda'}$ is the difference between the HOMO or LUMO energies of the two base pairs.

Eq. (3.7) is equivalent to the eigenvalue-eigenvector problem

$$\boldsymbol{H}\vec{v} = E\vec{v}, \tag{3.13}$$

where $\boldsymbol{H}$ is the hamiltonian matrix of order N, composed of the TB parameters E_λ and $t_{\lambda,\lambda'}$, and $\vec{v}$ is the vector matrix composed of the coefficients v_λ (which can be chosen to be real). The diagonalization of $\boldsymbol{H}$ leads to the determination of the eigenenergy spectrum (*eigenspectrum*), $\{E_k\}$, $k = 1, 2, \ldots, N$, for which we suppose that $E_1 < E_2 < \cdots < E_N$, as well as to the determination of the occupation probabilities for each eigenstate, $|v_{\lambda k}|^2$, where $v_{\lambda k}$ is the λ-th component of the k-th eigenvector. $\{v_{\lambda k}\}$ are normalized, and their linear independence is checked in all cases.

3.1.2 Time-dependent Problem

To describe the spatiotemporal evolution of an extra carrier created at a particular site of the polymer, we consider the state of the polymer as a linear combination of base-pair wave functions with time-dependent coefficients, i.e.,:

$$\psi^{DNA}(\vec{r}, t) = \sum_\lambda A_\lambda(t)\psi_\lambda^{bp}(\vec{r}), \tag{3.14}$$

where $\psi_\lambda^{bp}(\vec{r})$ is the wavefunction of the λ-th base pair.

Substituting Eq. (3.14) into the time-dependent Schrödinger's equation

$$i\hbar\frac{\partial\psi^{DNA}}{\partial t} = \hat{H}^{DNA}\psi^{DNA}, \tag{3.15}$$

we get:

$$i\hbar\frac{dA_\lambda}{dt} = E_\lambda^{bp}A_\lambda + t_{\lambda,\lambda-1}^{bp}A_{\lambda-1} + t_{\lambda,\lambda+1}^{bp}A_{\lambda+1}, \tag{3.16}$$

following the same procedure as in the previous Subsection 3.1.1. Hence, having determined E_λ^{bp} and $t_{\lambda',\lambda}^{bp}$, we can numerically solve the system of Eqs. (3.16), and obtain, through $A_\lambda(t)$, the time evolution of a charge transfer along the DNA segment of interest.

Eq. (3.16) is equivalent to a 1st order matrix differential equation of the form

$$\dot{\vec{A}}(t) = -\frac{i}{\hbar}\boldsymbol{H}\vec{A}(t), \tag{3.17}$$

where $\vec{A}(t)$ is a vector matrix composed of the coefficients $A_\lambda(t)$, $\lambda = 1, 2, \ldots, N$. Eq. (3.17) can be solved with the eigenvalue method, i.e., by looking for solutions of the form $\vec{A}(t) = \vec{v}e^{-\frac{i}{\hbar}Et} \Rightarrow \dot{\vec{A}}(t) = -\frac{i}{\hbar}E\vec{v}e^{-\frac{i}{\hbar}Et}$. Hence, Eq. (3.17) leads to the eigenvalue problem $\boldsymbol{H}\vec{v} = E\vec{v}$. Having determined the eigenvalues and eigenvectors of $\boldsymbol{H}$, the general solution of Eq. (3.17) is

$$\vec{A}(t) = \sum_{k=1}^{N} \sigma_k \vec{v}_k e^{-\frac{i}{\hbar}E_k t}. \tag{3.18}$$

In other words, the coefficients $A_\lambda(t), \lambda = 1, 2, \ldots, N$, are given by a superposition of the time evolution of the stationary states with time-independent coefficients σ_k. Hence, this is a coherent phenomenon. The coefficients σ_k are determined from the initial conditions. In particular, if we define the $N \times N$ eigenvector matrix $\boldsymbol{V}$, with elements $v_{\lambda k}$, then it can be shown that the vector matrix $\vec{\sigma}$, composed of the coefficients σ_k, $k = 1, 2, \ldots, N$, is given by the expression

$$\vec{\sigma} = \boldsymbol{V}^T \vec{A}(0). \tag{3.19}$$

Suppose that initially the extra carrier is placed at the l-th monomer, i.e., $A_l(0) = 1$, $A_\lambda(0) = 0, \forall \lambda \neq l$. Then,

$$\vec{\sigma} = \begin{bmatrix} v_{l1} \\ \vdots \\ v_{lk} \\ \vdots \\ v_{lN} \end{bmatrix}. \tag{3.20}$$

In other words, the coefficients σ_k are given by the row of the eigenvector matrix which corresponds to the monomer the carrier is initially placed at.

3.1.3 Physical Quantities

Having determined the eigenspectrum, we can compute the density of states ($\boldsymbol{DOS}$), generally given by

$$g(E) = \sum_{k=1}^{N} \delta(E - E_k). \tag{3.21}$$

Changing the view of a polymer from one (e.g. top) to the other (e.g. bottom) side of the growth axis, reflects the hamiltonian matrix $\boldsymbol{H}$ of the polymer on its main antidiagonal. This reflected Hamiltonian, $\boldsymbol{H}^{\mathrm{equiv}}$, describes the *equivalent polymer* [106]. $\boldsymbol{H}$ and $\boldsymbol{H}^{\mathrm{equiv}}$ are connected by the similarity transformation $\boldsymbol{H}^{\mathrm{equiv}} = \boldsymbol{L}^{-1}\boldsymbol{H}\boldsymbol{L}$, where $\boldsymbol{L}(= \boldsymbol{L}^{-1})$ is the unit antidiagonal matrix of order N. Therefore, $\boldsymbol{H}$ and $\boldsymbol{H}^{\mathrm{equiv}}$ have identical eigenspectra (hence the equivalent polymers' DOS is identical) and their eigenvectors are connected by $v_{\lambda k} = v^{\mathrm{equiv}}_{(N-\lambda+1)k}$. Generally,

$$\mathrm{equiv(YX...Z)} = \mathrm{Z_{compl}} \ldots \mathrm{Y_{compl}} \mathrm{X_{compl}}. \tag{3.22}$$

From Eq. (3.18) it follows that the probability to find the extra carrier at the λ-th monomer is

$$|A_\lambda(t)|^2 = \left| \sum_{k=1}^{N} \sigma_k v_{\lambda k} e^{-\frac{i}{\hbar} E_k t} \right|^2 \implies$$

$$|A_\lambda(t)|^2 = \sum_{k=1}^{N} \sum_{k'=1}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} e^{-\frac{i}{\hbar} E_k t} e^{\frac{i}{\hbar} E_{k'} t} \implies$$

$$|A_\lambda(t)|^2 = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 + 2 \sum_{k>k'=1}^{N} \sum_{k'=1}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \cos\left(\frac{E_k - E_{k'}}{\hbar} t \right), \qquad (3.23)$$

provided that σ_k, $\sigma_{k'}$, $v_{\lambda k}$, $v_{\lambda k'}$ are real. The mean over time probability to find the extra carrier at the λ-th in a time period $\mathcal{T}$ is:

$$\langle |A_\lambda(t)|^2 \rangle = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} |A_\lambda(t)|^2 dt. \qquad (3.24)$$

Substituting Eq. (3.23) into Eq. (3.24), we get:

$$\langle |A_\lambda(t)|^2 \rangle = \frac{1}{\mathcal{T}} \left[\sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \int_0^{\mathcal{T}} dt + 2 \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \int_0^{\mathcal{T}} dt \cos\left(\frac{E_k - E_{k'}}{\hbar} t \right) \right.$$

$$\implies \langle |A_\lambda(t)|^2 \rangle = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 + 2 \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \frac{\sin\left(\frac{E_k - E_{k'}}{\hbar} \mathcal{T} \right)}{\frac{E_k - E_{k'}}{\hbar} \mathcal{T}}.$$

For $\mathcal{T} \to \infty$, in the absence of degeneracy, we obtain the **mean over time probability** to find the extra carrier at the λ-th monomer:

$$\langle |A_\lambda(t)|^2 \rangle = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2. \qquad (3.25)$$

Periods and **frequencies** involved in charge transfer through a polymer of N monomers length are:

$$T_{kk'} = \frac{h}{E_k - E_{k'}}, \ \forall k > k' \qquad (3.26)$$

and

$$f_{kk'} = \frac{E_k - E_{k'}}{h}, \ \forall k > k', \qquad (3.27)$$

If M stands for the number of discrete Hamiltonian eigenvalues, then the different periods or frequencies involved in charge transfer are

$$\mathcal{S} = \binom{M}{2} = \frac{M!}{2(M-2)!} = \frac{M(M-1)}{2}. \qquad (3.28)$$

If there are no degenerated eigenenergies (which holds for all cases studied here, but e.g. does not hold for *cyclic* homopolymers [116]), then $M = N$. If eigenenergies are symmetric relative to some central value, then, S decreases (there exist degenerate $f_{kk'}$ or $T_{kk'}$). Specifically, in that case, $S = \frac{M^2}{4}$, for even M and $S = \frac{M^2-1}{4}$ for odd M. Fourier transform of a function $x(t)$ (as is $|A_\lambda(t)|^2$) can be expressed as:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{i\omega t} d\omega, \tag{3.29}$$

with Fourier coefficients

$$\mathcal{F}(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt, \tag{3.30}$$

or else, using frequencies:

$$x(t) = \int_{-\infty}^{\infty} \mathcal{F}(f) e^{i2\pi f t} df, \tag{3.31}$$

$$\mathcal{F}(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt. \tag{3.32}$$

Applying Eq. (3.32) to $|A_\lambda(t)|^2$ we obtain the Fourier coefficients $\mathcal{F}_\lambda(f)$, that determine the amplitude of each frequency:

$$\mathcal{F}_\lambda(f) = \int_{-\infty}^{\infty} |A_\lambda(t)|^2 e^{-i2\pi f t} dt \implies$$

$$\mathcal{F}_\lambda(f) = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \int_{-\infty}^{\infty} e^{-i2\pi f t} dt + 2 \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \int_{-\infty}^{\infty} e^{-i2\pi f t} \cos(2\pi f_{kk'} t) dt$$

$$= \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \int_{-\infty}^{\infty} e^{-i2\pi f t} dt + \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \int_{-\infty}^{\infty} e^{-i2\pi f t} \left(e^{i2\pi f_{kk'} t} + e^{-i2\pi f_{kk'} t} \right) dt. \tag{3.33}$$

It is known that

$$\delta(x - x_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(x-x_0)t} dt,$$

and

$$\delta(ax) = \frac{\delta(x)}{|\alpha|}, \; \forall \alpha \in \mathbb{R}.$$

Thus,

$$\mathcal{F}_\lambda(f) = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \delta(f) + \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} [\delta(f - f_{kk'}) + \delta(f + f_{kk'})] \tag{3.34}$$

Due to the function symmetry, the terms of $f < 0$ are cancelled and so we duplicate the terms of $f > 0$. Finally, the Fourier coefficients are:

$$\mathscr{F}_\lambda(f) = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \delta(f) + 2 \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} \sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'} \delta(f - f_{kk'}). \tag{3.35}$$

The frequencies involved in charge transfer are uniquely defined by the Hamiltonian eigenspectrum (see Eq. (3.27)). The frequency content is defined by the amplitudes that correspond to each frequency. Hence, the one-sided Fourier amplitude spectrum that corresponds to the probability $|A_\lambda(t)|^2|$ is given by:

$$|\mathscr{F}_\lambda(f)| = \sum_{k=1}^{N} \sigma_k^2 v_{\lambda k}^2 \delta(f) + 2 \sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} |\sigma_k \sigma_{k'} v_{\lambda k} v_{\lambda k'}| \delta(f - f_{kk'}). \tag{3.36}$$

Apart from the analytical calculation shown above, Fourier spectra can also be calculated using the Fast Fourier Transform (FFT) algorithm. These FFT spectra are in agreement with the ones computed in Eq. (3.36).

We can further define the **weighted mean frequency** (**WMF**) of monomer λ as

$$f_{WM}^\lambda = \frac{\sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} |\sigma_k v_{\lambda k} \sigma_{k'} v_{\lambda k'}| \cdot f_{kk'}}{\sum_{k=1}^{N} \sum_{\substack{k'=1 \\ k>k'}}^{N} |\sigma_k v_{\lambda k} \sigma_{k'} v_{\lambda k'}|}. \tag{3.37}$$

WMF expresses the mean frequency content of the extra carrier oscillation at monomer λ. Having determined the WMF for all monomers, we can now obtain a measure of the overall frequency content of carrier oscillations in the polymer. Since f_{WM}^λ is the weighted mean frequency of monomer λ and $\langle|A_\lambda(t)|^2\rangle$ is the mean probability of finding the extra carrier at monomer λ, we define the **total weighted mean frequency** (**TWMF**) as

$$f_{TWM} = \sum_{\lambda=1}^{N} f_{WM}^\lambda \langle|A_\lambda(t)|^2\rangle. \tag{3.38}$$

A quantity that evaluates simultaneously the magnitude of coherent charge transfer and the time scale of the phenomenon, is the **pure mean transfer rate** [115]

$$k_{l\lambda} = \frac{\left\langle|A_\lambda(t)|^2\right\rangle}{t_{l\lambda}}. \tag{3.39}$$

$t_{l\lambda}$ is the *mean transfer time*, i.e., having placed the carrier initially at monomer l, the time it takes for the probability to find the extra carrier at monomer λ, $\left|A_\lambda(t)\right|^2$, to become equal to its mean value, $\left\langle\left|A_\lambda(t)\right|^2\right\rangle$, for the first time. For the *pure* mean transfer rates,

$$k_{l\lambda} = k_{\lambda l} = k_{(N-l+1)(N-\lambda+1)}^{\text{equiv}} = k_{(N-\lambda+1)(N-l+1)'}^{\text{equiv}}, \tag{3.40}$$

where the superscript "equiv" refers to the equivalent polymer in the sense of Eq.(3.22).

3.2 Studied Polymers: Structures and Notation

In this section, we will present the structure of the studied polymers and set the relevant used notation. In our prototype system, B-DNA, we mention only the base sequence of the $5' - 3'$ strand. For example, we denote two successive monomers by YX, meaning that the base pair X-X$_{\text{compl}}$ is separated and twisted by 3.4 Å and 36°, respectively, relatively to the base pair Y-Y$_{\text{compl}}$, around the B-DNA growth axis. X$_{\text{compl}}$ (Y$_{\text{compl}}$) is the complementary base of X (Y).

3.2.1 Periodic Polymers with Increasing Repetition Unit

One could think of many types of periodic polymers, some of which are shown synoptically in Table 3.1. We studied cases of polymers made of identical monomers (denoted as I type) or different monomers (denoted as D type) [97]. We just give an example of the sequence, e.g., for type I4 we give the example GGCC..., but there are obviously other similar sequences: CCGG..., AATT..., TTAA.... P is the number of monomers in the repetition unit, e.g., for type I4, $P = 4$.

Table 3.1: The types of polymers mentioned in this work. I (D) denotes polymers made of the identical (different) monomers. P is the number of monomers in the repetition unit. We only mention the $5' - 3'$ base sequence.

(I,D)P	Sequence Example
I1	G... or A...
I2	GC...
I3	GGC...
I4	GGCC...
I6	GGGCCC...
I8	GGGGCCCC...
I10	GGGGGCCCCC...
I20	GGGGGGGGGGCCCCCCCCCC...
D2	GA...
D4	GGAA...
D6	GGGAAA...
D8	GGGGAAAA...
D10	GGGGGAAAAA...
D20	GGGGGGGGGGAAAAAAAAAA...

3.2.2   Quasi-Periodic and Fractal Polymers

The deterministic aperiodic sequences considered in this Ph.D. Thesis are either quasi-periodic or fractal [96]. Such structures are generally known as binary substitutional sequences, i.e., based on a binary alphabet, like {0, 1} and generated using appropriate substitution rules.

Fibonacci: The Fibonacci (F) sequence is named after the Italian mathematician Leonardo Pisano (Fibonacci) who introduced it to Western European mathematics in his 1202 book Liber Abaci, in a study of the population growth of rabbits [151]. However, this sequence appears many centuries before in Indian mathematics [152]. Fibonacci considers the growth of an idealized rabbit population, assuming that a single newly born pair of rabbits (N) are put in a field, and rabbits are able to mate at the age of one month so that at the end of its second month a mature pair (M) can produce another pair of rabbits. Rabbits never die and a mating pair always produces one new pair every month from the second month on. The puzzle that Fibonacci posed was: how many pairs will exist in one year? The collection of every month's population is: $\mathcal{F}_0 = $ N, $\mathcal{F}_1 = $ M, $\mathcal{F}_2 = $ MN, $\mathcal{F}_3 = $ MNM, $\mathcal{F}_4 = $ MNMMN, etc. Using e.g. the two-letter alphabet {G, A}, we can define the Fibonacci generation $\mathcal{F}_g$ by the substitution rules A $\rightarrow$ G, G $\rightarrow$ GA, starting with $\mathcal{F}_0 = $ A. Hence, $\mathcal{F}_0 = $ A, $\mathcal{F}_1 = $ G, $\mathcal{F}_2 = $ GA, $\mathcal{F}_3 = $ GAG, $\mathcal{F}_4 = $ GAGGA, etc. If $\mathcal{N}_g$ is the Fibonacci number of generation g, and we set $\mathcal{N}_0 = \mathcal{N}_1 = 1$, the recurrence relation $\mathcal{N}_g = \mathcal{N}_{g-1} + \mathcal{N}_{g-2}$ produces the number sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, ... Figure 3.1 depicts the 0th, 1st, 2nd, 3rd and 4th generation of the Fibonacci sequence, using a two-colour rule.
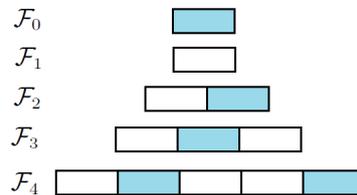
Figure 3.1: Fibonacci sequence, 0th, 1st, 2nd, 3rd and 4th generation, depicted by two-colour rule.

Thue-Morse: The Thue-Morse (TM) or Prouhet-Thue-Morse sequence was first studied by Eugène Prouhet in 1851, who applied it to number theory [153]. The systematic study was left to Axel Thue who, in 1906, applied it on his study of words combinatorics [154]. The most important contribution to the sequence was made in 1921 by Marston Morse in the context of differential geometry and topological dynamics [155], which brought the sequence to worldwide attention. In its simplest form, the TM sequence can be defined by the recursive relations $S_n = \{S_{n-1}S_{n-1}^+\}$ and $S_n^+ = \{S_{n-1}^+S_{n-1}\}$ (for $n \geq 1$), with $S_0 = 0$ and $S_0^+ = 1$ [156]. Using e.g. the two-letter alphabet {G, A} we can build up the sequence using the substitution rules G$\rightarrow$GA and A$\rightarrow$AG. Hence, $\mathcal{TM}_0 = $ G, $\mathcal{TM}_1 = $ GA, $\mathcal{TM}_2 = $ GAAG, $\mathcal{TM}_3 = $ GAAGAGGA, etc. Figure 3.2 depicts the 0th, 1st, 2nd, 3rd and 4th generation of the Thue-Morse sequence, using a two-colour rule.
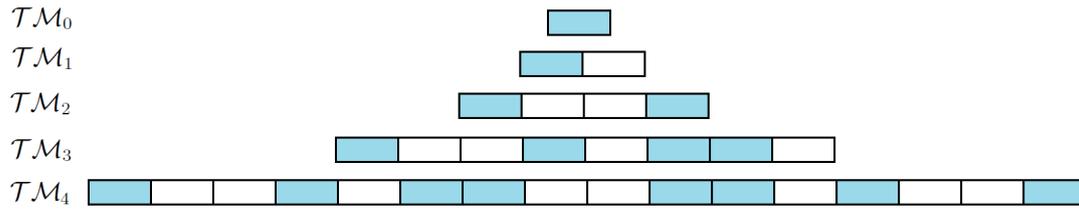
Figure 3.2: Thue-Morse sequence, 0th, 1st, 2nd, 3rd and 4th generation, depicted by two-colour rule.

Double-Period: The double-period (DP) sequence has its origin in the study of system dynamics and laser applications to nonlinear optical fibers [157]. It is closely connected with the TM sequence: the n-th stage is $S_n = \{S_{n-1}S_{n-1}^+\}$ and $S_n^+ = \{S_{n-1}S_{n-1}\}$ (for $n \geq 1$), with $S_0 = 0$ and $S_0^+ = 1$. Using e.g. the two-letter alphabet {G, A}, we can define the n-th generation by the substitution rules G→GA, A→GG. Hence, starting with $\mathscr{DP}_0 =$ G, then $\mathscr{DP}_1 =$ GA, $\mathscr{DP}_2 =$ GAGG, $\mathscr{DP}_3 =$ GAGGGAGA, etc. Figure 3.3 depicts the 0th, 1st, 2nd, 3rd and 4th generation of the Double-Period sequence, using a two-colour rule.
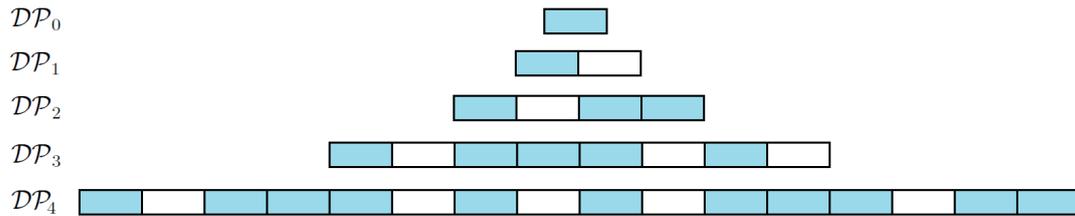
Figure 3.3: Double-Period sequence, 0th, 1st, 2nd, 3rd and 4th generation, depicted by two-colour rule.

Rudin-Shapiro: The Rudin-Shapiro (RS) aka Golay-Rudin-Shapiro sequence is named after Marcel Golay, Walter Rudin and Harold S. Shapiro, who independently investigated its properties [158--160]. It is generated starting with $+1, +1$ and employing the rules:

$$+1, +1 \rightarrow +1, +1, +1, -1$$
$$+1, -1 \rightarrow +1, +1, -1, +1$$
$$-1, +1 \rightarrow -1, -1, +1, -1$$
$$-1, -1 \rightarrow -1, -1, -1, +1 \qquad .$$

Using e.g. the two-letter alphabet {G, A} and employing the inflation rule: GG→GGGA, GA→GGAG, AG→AAGA, AA→AAAG, the first generations are $\mathscr{RS}_1 =$ GG, $\mathscr{RS}_2 =$ GGGA, $\mathscr{RS}_3 =$ GGGAGGAG, etc. Figure 3.4 depicts the 0th, 1st, 2nd, 3rd and 4th generation of the Rudin-Shapiro sequence, using a two-colour rule.

Cantor Set: The Cantor Set (CS), introduced by mathematician Georg Cantor, is one of the most well-known deterministic fractals [161]. It is built by splitting a straight line segment in three, removing the middle third, then removing the middle third of each of the two new straight line segments and the process is repeated *ad infinitum*. Using e.g. the two-letter alphabet {G, A} and
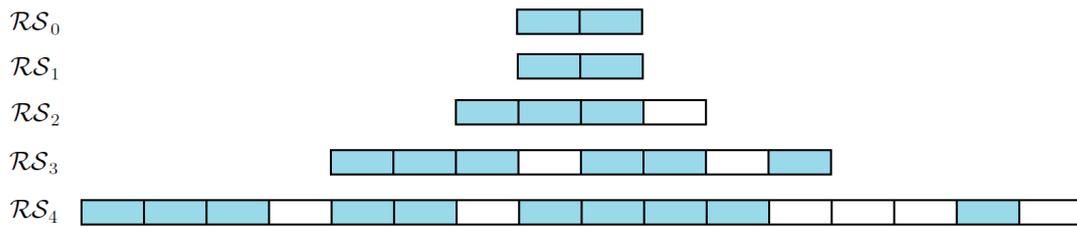
Figure 3.4: Rudin-Shapiro sequence, 0th, 1st, 2nd, 3rd and 4th generation, depicted by two-colour rule.

the substitution rules G→GAG, A→AAA, we can define the n-th generation ($n = 0, 1, 2, ...$) as follows: $\mathcal{CS}_0 =$ G, $\mathcal{CS}_1 =$ GAG, $\mathcal{CS}_2 =$ GAGAAAGAG, etc. Figure 3.5 depicts the 0th, 1st, 2nd and 3rd generation of the Cantor Set sequence, using a two-colour rule.
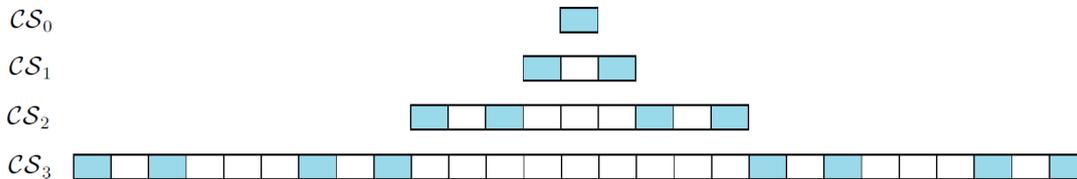
Figure 3.5: Cantor Set sequence, 0th, 1st, 2nd and 3rd generation, depicted by two-colour rule.

Asymmetric Cantor Set: The Asymmetric Cantor Set (ACS), is built by splitting a straight line segment in four, removing the second quarter, then removing the second quarter of each of the three new straight line segments and the process is repeated *ad infinitum*. Using e.g. the two-letter alphabet {G, A} and the substitution rules G→GAGG, A→AAAA, we can define the n-th generation ($n = 0, 1, 2, ...$) as follows: $\mathcal{ACS}_0 =$ G, $\mathcal{ACS}_1 =$ GAGG, $\mathcal{ACS}_2 =$ GAGGAAAAGAGGGAGG, etc. Figure 3.6 depicts the 0th, 1st and 2nd generation of the Asymmetric Cantor Set sequence, using a two-colour rule.
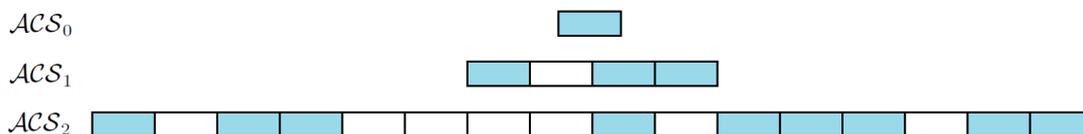
Figure 3.6: Asymmetric Cantor Set sequence, 0th, 1st and 2nd generation, depicted by two-colour rule.

One could think of many types of aperiodic polymers, some of which are shown synoptically in Table 3.2. We just give an example of each type, e.g., for Fibonacci I sequences we give the example G, C, CG, CGC, CGCCG, CGC-CGCGC, ..., but there are obviously other similar sequences e.g. C, G, GC,

GCG, GCGGC, GCGGCGCG, ..., A, T, TA, TAT, TATTA, TATTATAT, ...,
T, A, AT, ATA, ATAAT, ATAATATA, ...

Table 3.2: Examples of the types of polymers studied in this work. I (D) denotes polymers made of identical (different) monomers. We only mention the $5' - 3'$ base sequence along one of the two strands.

Type	Sequence Example	Notation
Fibonacci I	G, C, CG, CGC, CGCCG, ...	F G(C)
Fibonacci D	G, A, AG, AGA, AGAAG, ...	F G(A)
Thue-Morse I	G, GC, GCCG, GCCGCGGC, ...	TM G(C)
Thue-Morse D	A, AG, AGGA, AGGAGAAG, ...	TM A(G)
Double Period I	T, TA, TATT, TATTTATA, ...	DP T(A)
Double Period D	A, AG, AGAA, AGAAAGAG, ...	DP A(G)
Rudin-Shapiro I	AA, AAAT, AAATAATA, ...	RS A(T)
Rudin-Shapiro D	AA, AAAG, AAAGAAGA, ...	RS A(G)
Cantor Set I	T, TAT, TATAAATAT, ...	CS T(A)
Cantor Set D	A, AGA, AGAGGGAGA, ...	CS A(G)
Asymmetric Cantor Set I	C, CGCC, CGCCGGGGCGCCCGCC, ...	ACS C(G)
Asymmetric Cantor Set D	A, AGAA, AGAAGGGGAGAAAGAA, ...	ACS A(G)

3.3   Results and Discussion

First, we obtain the charge transfer parameters between two successive base pairs by calculating the corresponding overlap integrals from Eq. (3.8). Table 3.3 summarizes our LCAO results using all valence orbitals for the transfer parameters, for all possible combinations of successive base pairs and close-to-ideal geometrical conformations [27]. The Table also contains hole transfer parameters of Ref. [115], an estimation from various articles found in bibliography, the electron transfer parameters of Ref. [63], where only $2p_z$ orbitals had been used and the transfer parameters with the parameterization of Ref. [55], where only $2p_z$ orbitals had been used.

Table 3.3: Close-to-ideal geometrical conformations. The absolute values of transfer parameters for all possible combinations of successive base pairs. $|t_H|$ ($|t_L|$) of the second (fifth) column refer to hole (electron) transfer parameters obtained from our LCAO calculations using all valence orbitals [27]. The third column lists hole transfer parameters of Ref. [115], an estimation from various articles found in bibliography (in parentheses the parameters obtained from the HKS parameterization [63]). The sixth column lists the electron transfer parameters of Ref. [63], where only $2p_z$ orbitals had been used. The fourth and seventh column list the transfer parameters with the parameterization of Ref. [55], where only $2p_z$ orbitals had been used. All transfer parameters are given in meV.

XY	$|t_H|$ [27]	$|t_H|$ [115] ([63])	$|t_H|$ [55]	$|t_L|$ [27]	$|t_L|$ [63]	$|t_L|$ [55]
GG, CC	116	100 (62)	51	92 (σ^*) 2	20	8
AG, CT	37	30 (5)	32	11 (σ^*) 11	3	10
TG, CA	28	10 (4)	4	2 (σ^*) 9	17	10
AC, GT	16	10 (2)	3	1 (σ^*) 1	32	23
TC, GA	142	110 (79)	57	3 (σ^*) 6	1	7
AA, TT	38	20 (8)	32	22	29	17
AT	50	35 (20)	6	1	1	1
TA	37	50 (47)	10	2	2	1
GC	10	10 (1)	10	2 (σ^*) 19	10	19
CG	75	50 (44)	13	1 (σ^*) 9	8	13

In Figure 3.7 we illustrate the absolute values of transfer parameters for all possible combinations of successive base pairs for holes and for electrons. The figure contains the transfer parameters obtained from our LCAO calculations using all valence orbitals [27], along with the corresponding parameters found in ref. [115] (where various estimations from bibliography had been taken into

account). Also, those from ref. [55], where only $2p_z$ orbitals had been used, and finally, electron transfer parameters from ref. [63], where only $2p_z$ orbitals had been used. In Figure 3.8 we depict the maximum transfer percentage of Equa-
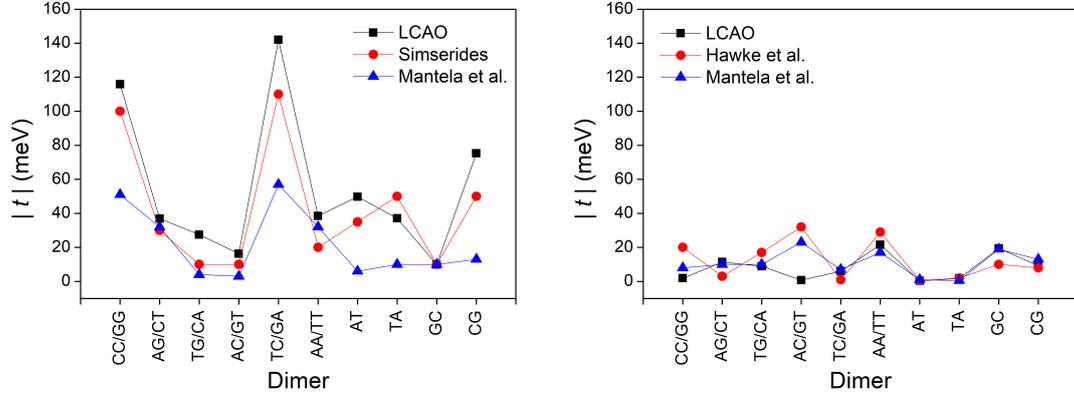
Figure 3.7: The absolute values of transfer parameters for all possible combinations of successive base pairs for holes (left) and for electrons (right). We show the transfer parameters obtained from our LCAO calculations using all valence orbitals [27], as well as the corresponding transfer parameters found in ref. [115] (for holes, estimation from various articles in bibliography), in ref. [55] (using only $2p_z$ orbitals) and in ref. [63] (for electrons, using only $2p_z$ orbitals).

tion (3.12) obtained by our LCAO calculations using all valence orbitals [27], compared to the values using parameters from ref. [115] for holes. Also, from ref. [55] for electrons and holes as well as from ref. [63] for electrons (where only $2p_z$ orbitals had been used). For ideal B-DNA geometries and for dimers made
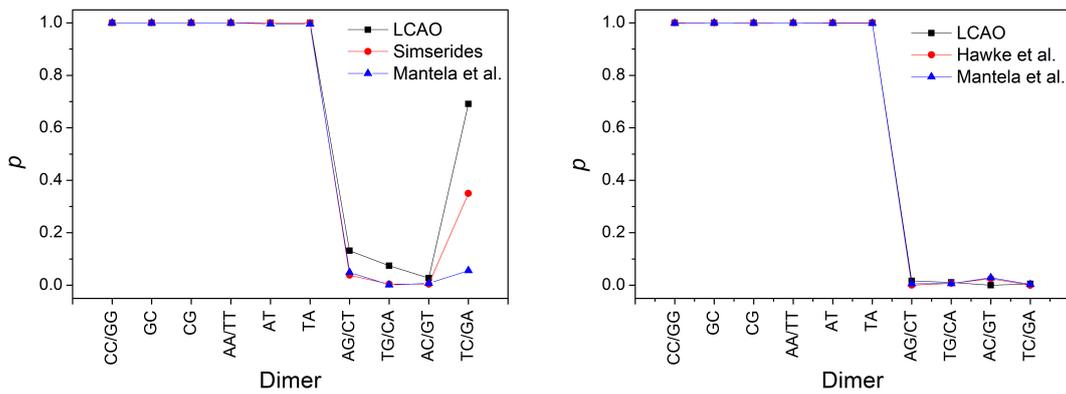
Figure 3.8: Comparison of the maximum transfer percentage p obtained by our LCAO method using all valence orbitals [27], with the p values extracted from other sources: obtained from parameters found in ref. [115] (for holes, estimation from various articles in bibliography), in ref. [55] (using only $2p_z$ orbitals) and in ref. [63] (for electrons, using only $2p_z$ orbitals).

of identical monomers, the maximum transfer percentage is 1, while in the case of different monomers, p is smaller than 1, both for holes and for electrons. Both for t and p, we observe that the LCAO using all valence orbitals is closer to the

results from Ref. [115] for holes (where various estimations from bibliography of different origin had been taken into account). For electrons, as far as we know this current LCAO calculation is the only one beyond simple Hückel models, using only $2p_z$ orbitals.

3.3.1   Periodic Polymers with Increasing Repetition Unit

This Subsection is devoted to charge transfer in several categories of periodic polymers made of N monomers, with a repetition unit made of P monomers, using the TB wire model, where a site is a monomer (e.g., in DNA, a base pair). P is even, and we deal with two categories of such polymers: made of the same monomer (GC..., GGCC..., etc.) and made of different monomers (GA..., GGAA..., etc.), as described in Subsection 3.2.1. The content of this Subsection can be found in ref. [97]. Since the author of the present Ph.D. Thesis is not the first author of ref. [97], the presentation of the associated results will be limited, along with some additional remarks that will be given, comparatively to other categories of polymers in Chapter 6. The TB parameters (HOMO and LUMO interaction integrals) used in this Subsection can be found in Table 3.3, columns 3 and 6, unless otherwise stated. The HOMO and LUMO base pair on-site energies can be found in Table 2.4, columns 3 and 5.

A. Eigenspectra and Density of States.
In Figs. 1 and 2 of ref. [97], are depicted the HOMO and LUMO eigen-spectra of [I2, I4, I6, I8, I10, I20 and I1 polymers] and [D2, D4, D6, D8, D10, D20 and I1 (G...), I1 (A...) polymers]. In Figs. 3 and 4 are plotted the corresponding DOS. The HOMO and LUMO bands of each polymer consist of P subbands. Some eigenenergies protrude periodically from the subbands at certain relationships between N and P. The subbands are separated by small energy gaps, which, increasing P, decrease.

For polymers made of identical monomers all eigenvalues are symmetric around the monomer on-site energy and for N odd the trivial eigenvalue, equal to the monomer on-site energy, exists. Increasing P, the eigenspectra tend to the eigenspectra of I1 polymers, and the DOS tends to the DOS of I1 polymers. For polymers made of different monomers, increasing P, the eigenenergies gather around the two monomers' on-site energies. Increasing P, the eigenspectra gather within the limits defined by the union of eigenspectra of I1 (G...) and I2 (A...) polymers.

The energy gap of a polymer is the difference between the lowest level of the LUMO regime and the highest level of the HOMO regime, because we assume that the orbitals - one per site - which contribute to the HOMO (LUMO) band are occupied (empty), since in both possible monomers there is an even number of $2p_z$ electrons contributing to the π stack [63].
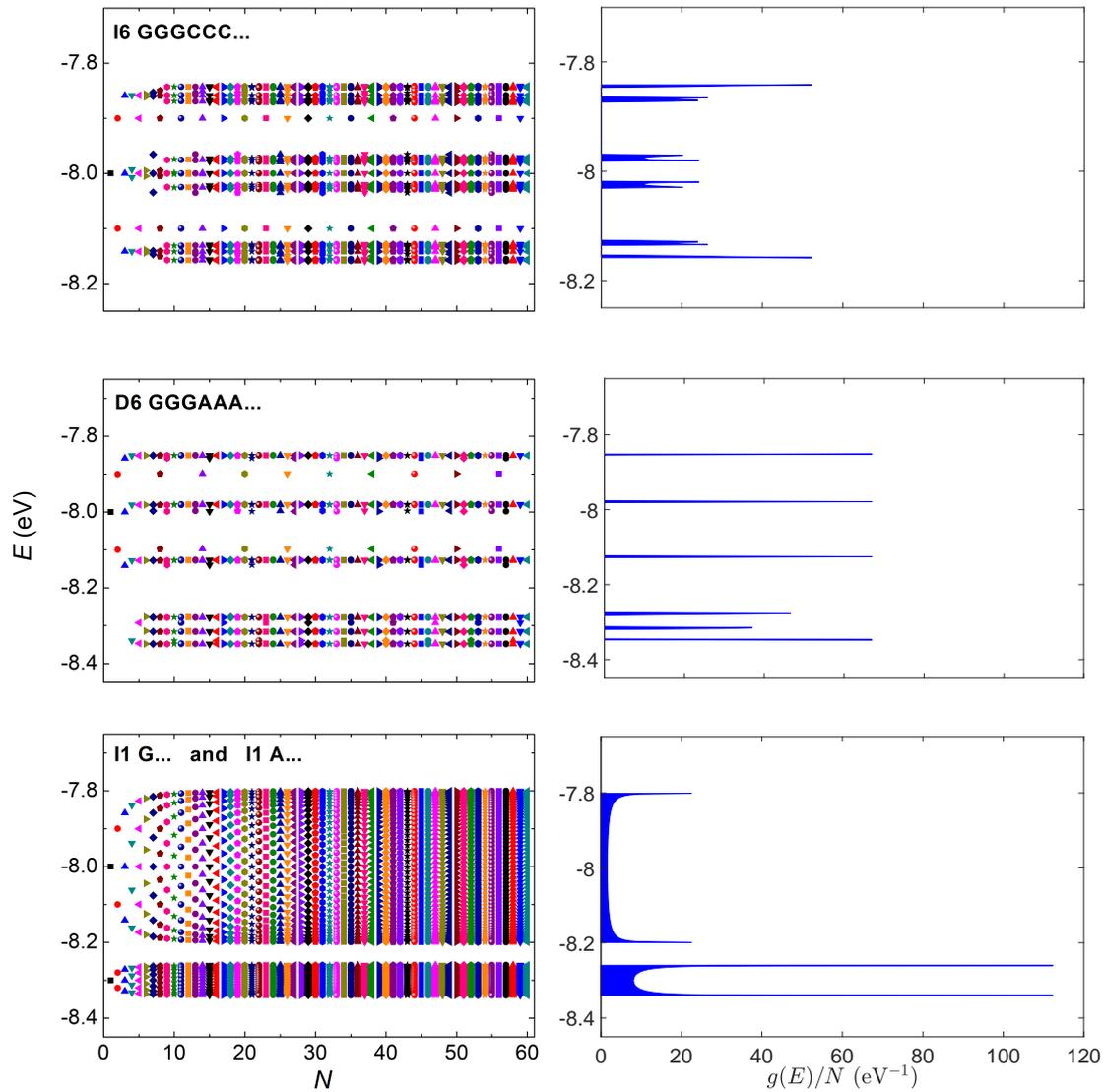
Figure 3.9: Eigenspectra (left) and DOS (right) of I6 polymers (upper panels), D6 polymers (middle panels) and I1(G...) and I1(A...) (lower panels), HOMO regime.

At the large-N limit, increasing P, the gaps of I2, I4, I6, ... polymers approach the gap of I1 polymer. Increasing the repetition unit in the mode GC, GGCC, GGGCCC, ..., finally results in a G...GC...C polymer which is almost G... with just a switch from G to C at the middle of the polymer. Hence, at the large-N limit, the energy gap of I1 polymers is the smallest of these series of polymers. For the same reason, increasing P, the eigenspectra and the DOS of I2, I4, I6, ... polymers tend to the eigenspectra and the DOS of I1 polymers. At the large-N limit, increasing P, the gaps of D2, D4, D6, ... polymers approach the gap of the union of I1 (G...) and I1 (A...) polymers. Increasing the repetition unit in the mode GA..., GGAA..., GGGAAA... and so on, finally results in a G...GA...A polymer which is energetically almost a union of separated G... and A... polymers. This happens due to the large difference of G-C
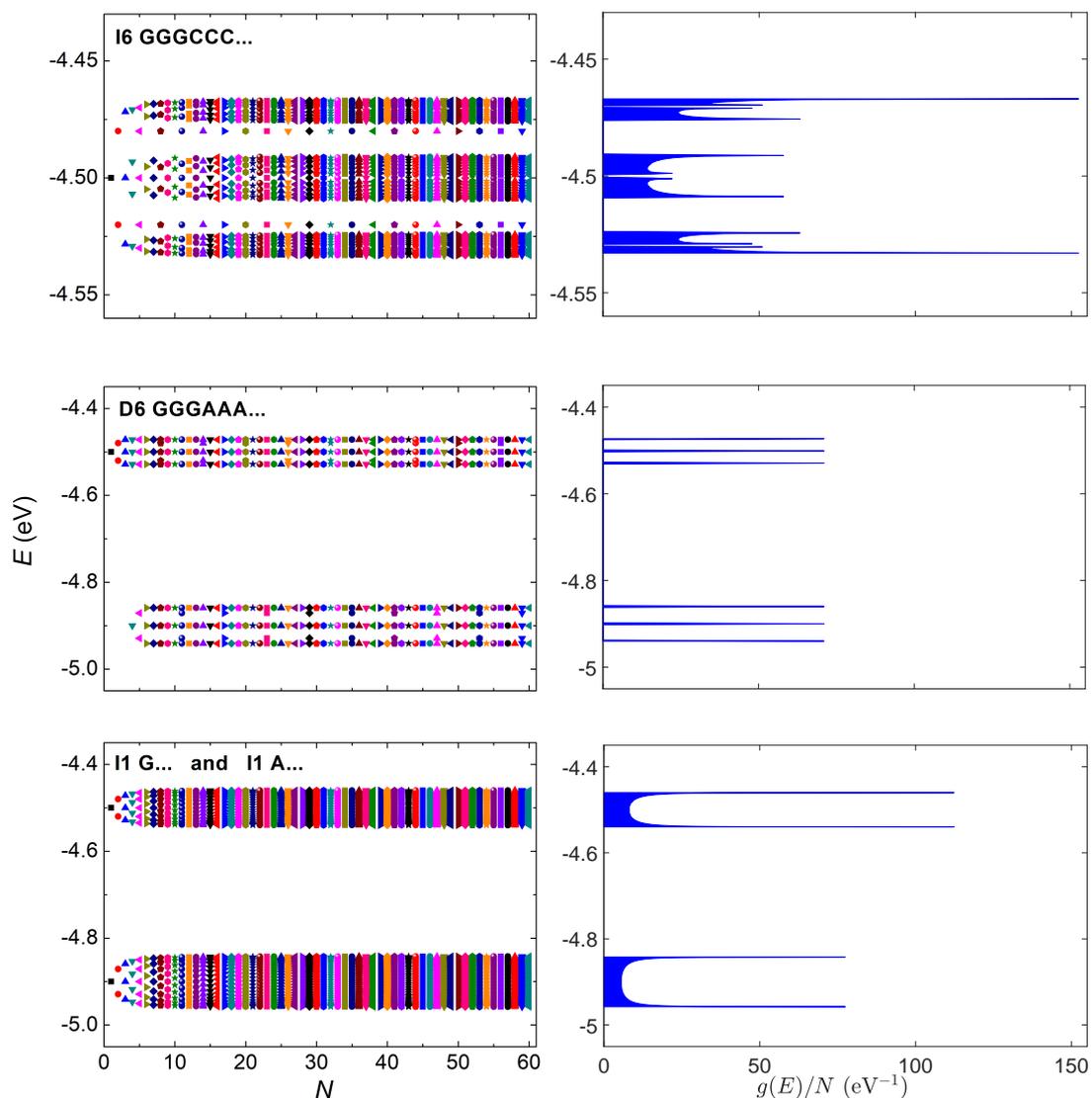
Figure 3.10: Eigenspectra (left) and DOS (right) of I6 polymers (upper panels), D6 polymers (middle panels) and I1(G...) and I1(A...) (lower panels), LUMO regime.

and A-T on-site energies in comparison with the t_{GA} interaction integral. Increasing P, the lowering of the energy gap in the case of D polymers [$\approx$ 0.6 (0.7) eV relative to the A-T (G-C) monomer gap] is much bigger than in the case of I polymers [$\approx$ 0.25 eV relative to the G-C monomer gap]. Figs. 3.9 and 3.10 display some characteristic cases of HOMO and LUMO eigenspectra and the corresponding DOS.

B. Mean over time Probabilities.

The main results for the mean over time probabilities for I2, I4, ... polymers are summarized in Figs. 6 and 7 of ref. [97]. For N equal to natural multiples of P ($N = Pn, n \in \mathcal{n}^*$), *palindromicity* holds, i.e., the probabilities are palindromic due to the fact that the corresponding hamiltonian matrices are palindromic, i.e. reading them from top left to bottom right and vice versa gives the same result. The palindromicity holds for all initial conditions. Hence, in these polymer cases, the appropriate choice of the monomer the carrier is injected to, can lead to enhanced presence at specific sites at its other end, leading to more efficient transfer. For I1 polymers and initial placement of the carrier at the first monomer, the mean over time probability to find the carrier at the first or at the last monomer is ψ and at any other monomer is χ, [116] where

$$\psi = \frac{3}{2(N+1)}, \quad \chi = \frac{1}{N+1}. \tag{3.41}$$

Increasing P, the relevant probabilities of I2, I4, I6, ... polymers tend to the I1 probabilities ψ and χ of Eq. (3.41).

For $N \neq Pn$, palindromicity is lost. In the HOMO regime, all studied polymers with $N \neq Pn$, show increased mean (over time) probabilities at the $\frac{P}{2}$ initial monomers. In the LUMO regime, this property cannot be clearly seen, because t_{GG} is the greater of all, but t_{CG} and t_{GC} have similar values.

The results for the mean (over time) probabilities for D2, D4, ... polymers are summarized in Fig. 8 of ref. [97]. A basic observation for polymers made of different monomers is that if we initially place the carrier at a G-C monomer the probability to find it at an A-T monomer is small, and vice versa.

C. Frequency Content.

The Fourier spectra of the time-dependent probability to find the carrier at each monomer, are, generally, in the THz regime. A general remark is that when the dominant frequencies i.e. those with the greater Fourier amplitudes are smaller (bigger), the carrier transfer from the first to the last monomer is slower (faster).

For $N = Pn$, $n \in \mathcal{n}$, for I1, I2, I4, I6, ... polymers, the Fourier spectra of the time-dependent probability to find an extra carrier at the various monomers, either for the HOMO or the LUMO regime, are palindromic, i.e., they are identical for the μ-th and $(N - \mu + 1)$-th monomer. Since for $N \neq Pn$, $n \in \mathcal{n}$ the hamiltonian matrices are not palindromic, the Fourier spectra are also not palindromic.

As for the TWMF as a function of N, the reader can check out Figs. 9 and 10 of ref. [97]. In the cases of I2 (D2) polymers, only two interaction integrals are involved: t_{GC} and t_{CG} (t_{GA} and t_{AG}). In I4, I6, ... polymers, three interaction integrals are involved: $t_{\mathrm{GG}}, t_{\mathrm{GC}}, t_{\mathrm{CG}}$, while in the cases of D4,

D6, ... polymers, four interaction integrals are involved: $t_{GG}, t_{GA}, t_{AG}, t_{AA}$. This is the reason that in the limit of large N, the TWMF for I2 (D2) polymers tends to a different frequency region than for I4, I6, ... (D4, D6, ...) polymers. In particular, the TWMF of I4, I6, ... polymers, in the limit of large N, tends to the TWMF of I1 polymers. For D4, D6, ... polymers, if we place the carrier initially at a G-C monomer, the TWMF of D4, D6, ... polymers, in the limit of large N, tends to the TWMF of I1 polymers, where only one interaction integral is involved: t_{GG}.

D. Pure Mean Transfer Rates.
An impressive case where appropriate sequence choice can increase k by many orders of magnitude is shown in Fig. 11 of ref. [97]. Type I1, I2, I4, I6, I8 and I10 polymers, for $N = nP$, are palindromic, hence there is enhanced presence of the extra carrier at the last monomer. In all cases, $k(N)$ is a decreasing function. The electron k range is many orders of magnitude narrower than the hole k range, due to the much smaller difference between the interaction integrals (t_{GG}, t_{GC}, t_{CG}) involved. For increasing P, k takes increasingly larger values. In other words, the degree of transfer difficulty is greater for type I2 polymers and decreases gradually for types I4, I6, ... polymers. However, $k(N)$ has an upper limit which is $k(N)$ of type I1 polymers. The latter polymers are structurally simpler (more precisely, they have the simplest possible structure), a fact that favors charge transfer along them, so their transfer rates are higher than those of the other polymer types.

For polymers made of different monomers $k(N)$, is depicted in Fig. 12 of ref. [97]. While for type I1 polymers (G... and A...) k drops $\approx$ by only 2 to 3 orders of magnitude, increasing N, as the number of A in the repetition unit increases, $k(N)$ drops dramatically by many more orders of magnitude. Overall, the above results suggest that type I1 polymers are the best for electron or hole transfer.

3.3.2 Quasi-Periodic and Fractal Polymers

In this Subsection, the TB parameters for B-DNA are taken from ref. [63]. The interaction integrals can be found in Table 3.3, columns 3 (parentheses) and 6. The HOMO and LUMO base pair on-site energies can be found in Table 2.4, columns 3 and 5.

A. Eigenspectra and Density of States.
In Figs. 3.11, 3.12, 3.13, we present the HOMO and LUMO eigenspectra, for increasing N, of I polymers, along with the corresponding DOS for large N. In Figs. 3.14, 3.15, 3.16 , we present the HOMO and LUMO eigenspectra, for increasing N, of I polymers, along with the corresponding DOS for large N. The DOS has been calculated for polymers made of a very big number of monomers N, for illustration purposes. This N value is shown in each panel. Of course, the persistence length of DNA is around

50 nm or 150 base pairs [6]. On the other hand, if we stretch and join the DNA of all chromosomes of a single cell, that would give us a length of the order of a meter and would consist of billions of base pairs. For both I and D polymers, we notice that in quasi-periodic polymers (Fibonacci, Thue-Morse, Double Period and Rudin-Shapiro, Figs. 3.11, 3.12, 3.14, 3.15) the DOS has rather acute subbands, while in fractal polymers (Cantor Set, Asymmetric Cantor Set, Figs. 3.13, 3.16) the DOS is fragmented and spiky.

For I polymers, i.e., polymers made of identical monomers (cf. Figs. 3.11, 3.12, 3.13), we observe that all eigenvalues are symmetric relative to the monomer's on-site energy (this, obviously, also holds for the DOS). This observation can be mathematically proven as follows: For N even, the hamiltonian matrix of a generic I polymer is $\boldsymbol{H} = E_\mu \boldsymbol{I} + \boldsymbol{T}_{GK}$, where E_μ is the (constant) on-site energy, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{T}_{GK}$ is the Golub-Kahan matrix, containing only the non-diagonal elements of $\boldsymbol{H}$, i.e., the HOMO or LUMO interaction integrals $t_{\mu,\lambda}$. It can easily be shown that $\boldsymbol{T}_{GK} = \boldsymbol{P}^T \boldsymbol{B} \boldsymbol{P}$, where $\boldsymbol{P}$ is the perfect shuffle matrix and

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{O} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} t_{1,2} & t_{2,3} & & \\ & t_{3,4} & t_{4,5} & \\ & & \ddots & \ddots \\ & & & t_{N-1,N} \end{pmatrix}. \tag{3.42}$$

By performing the Singular Value Decomposition of the upper bidiagonal matrix $\boldsymbol{A}$, i.e., by writing it as $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{S} \boldsymbol{W}^T$, we obtain

$$\boldsymbol{B} = \boldsymbol{J} \begin{pmatrix} -\boldsymbol{S} & 0 \\ 0 & \boldsymbol{D} \end{pmatrix} \boldsymbol{J}^T, \quad \boldsymbol{J} = \frac{1}{\sqrt{2}} \begin{pmatrix} \boldsymbol{U} & \boldsymbol{U} \\ -\boldsymbol{W} & \boldsymbol{W} \end{pmatrix}. \tag{3.43}$$

So, finally,

$$\boldsymbol{T}_{GK} = \boldsymbol{P}^T \boldsymbol{J} \begin{pmatrix} -\boldsymbol{S} & 0 \\ 0 & \boldsymbol{S} \end{pmatrix} \boldsymbol{J}^T \boldsymbol{P}. \tag{3.44}$$

Hence, the eigenvalues of $\boldsymbol{T}_{GK}$ are given by the positive and negative values of the diagonal matrix $\boldsymbol{S}$, i.e., they are symmetric around zero [162]. Hence, since, $\boldsymbol{H} = E_\mu \boldsymbol{I} + \boldsymbol{T}_{GK}$, the eigenvalues of $\boldsymbol{H}$ are symmetric around E_μ. For N odd, we can add a zero row and a zero column to $\boldsymbol{T}_{GK}$ so that it is again of even order and follow the aforementioned procedure. Then, two degenerate trivial eigenvalues will appear apart from the symmetric ones [163]. So, the eigenvalues of $\boldsymbol{H}$ occur by omitting the zero row and column, hence they are symmetric around E_μ, which is also an eigenvalue.

For D polymers, i.e., polymers made of different monomers (cf. Figs. 3.14, 3.15, 3.16), the eigenenergies and the DOS gather around the two monomers' on-site energies.
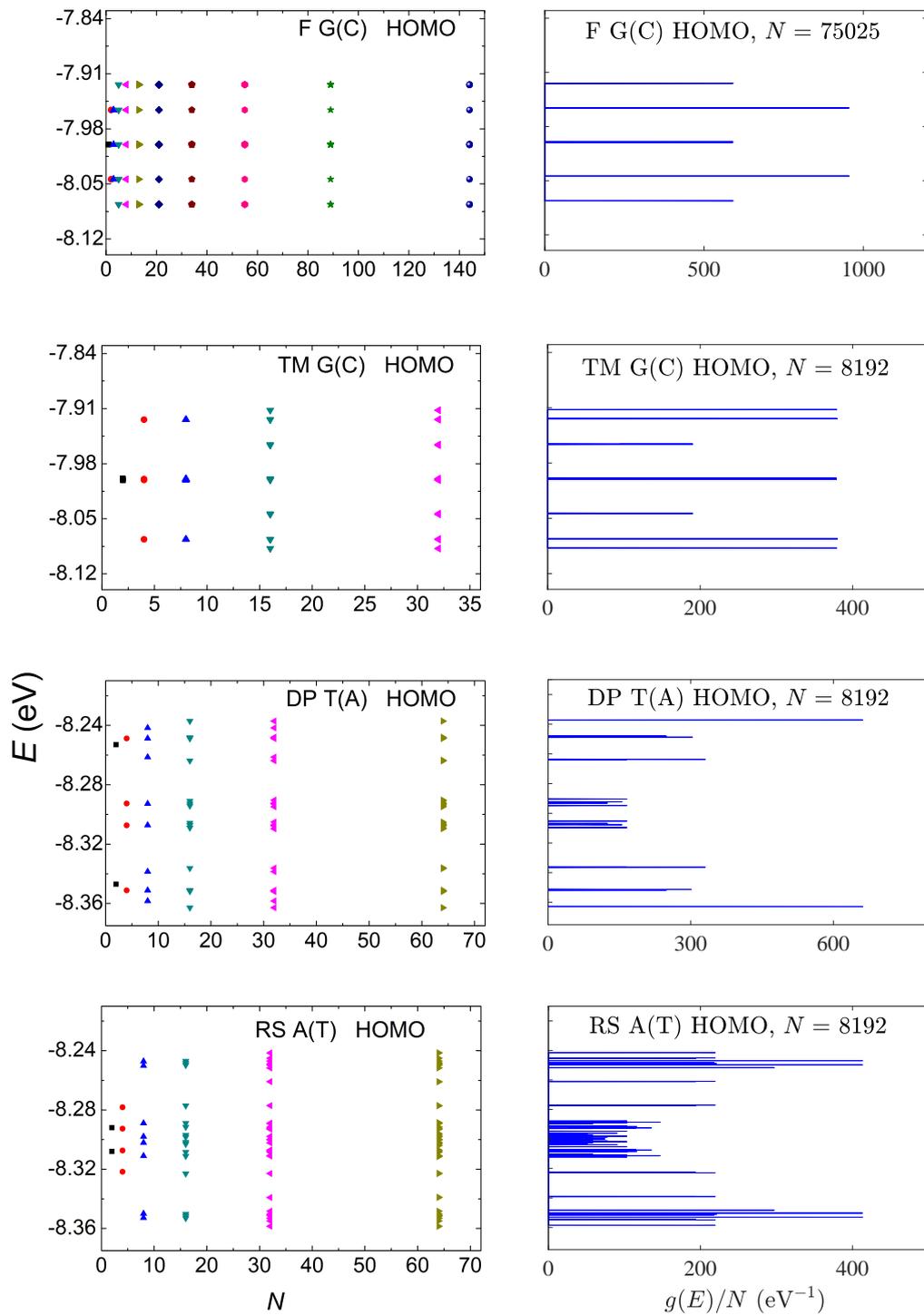
Figure 3.11: Eigenspectra of F G(C), TM G(C), DP T(A), RS A(T) polymers, for the HOMO regime, for a few generations, along with the corresponding DOS for large N.
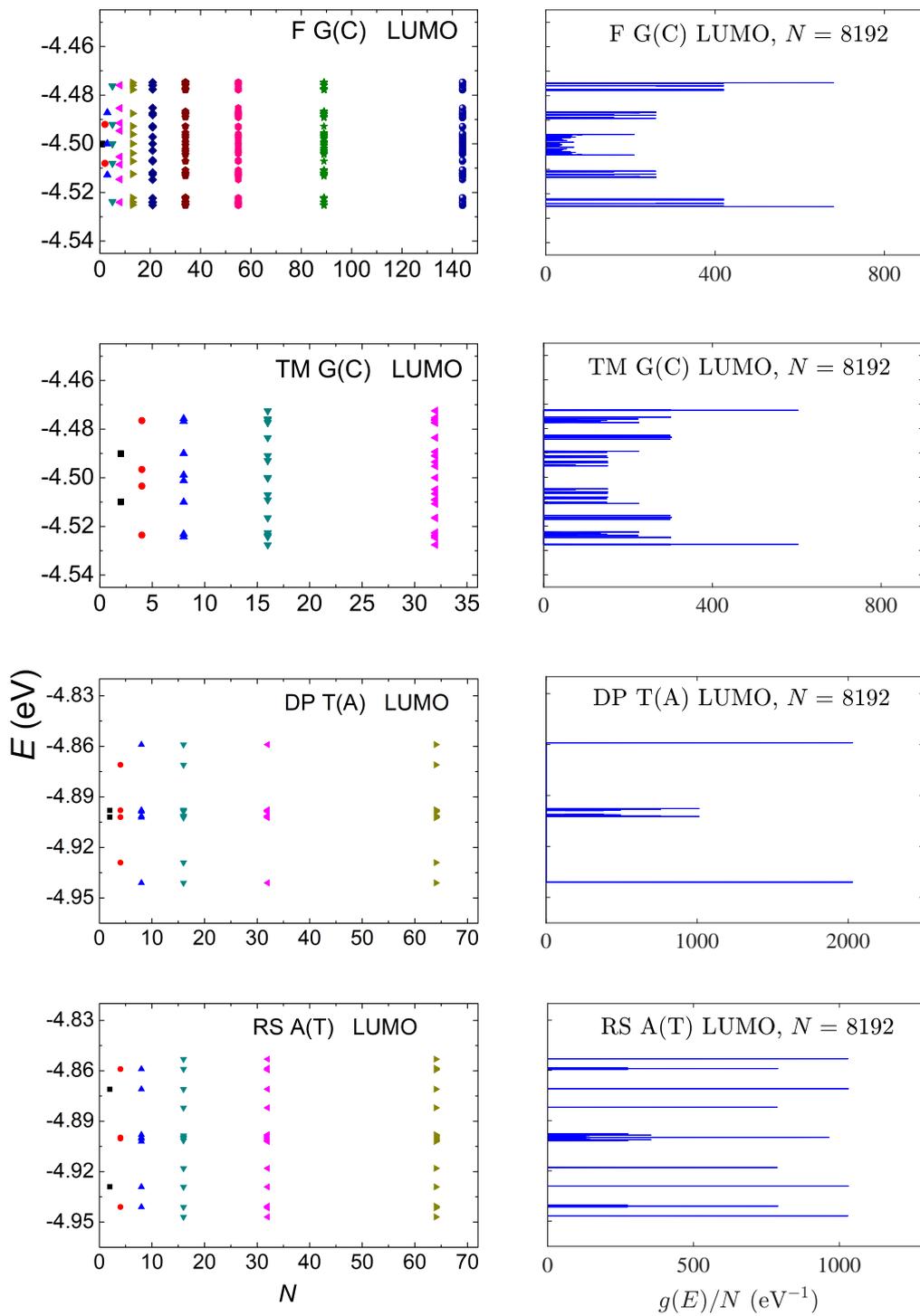
Figure 3.12: Eigenspectra of F G(C), TM G(C), DP T(A), RS A(T) polymers, for the LUMO regime, for a few generations, along with the corresponding DOS for large N.
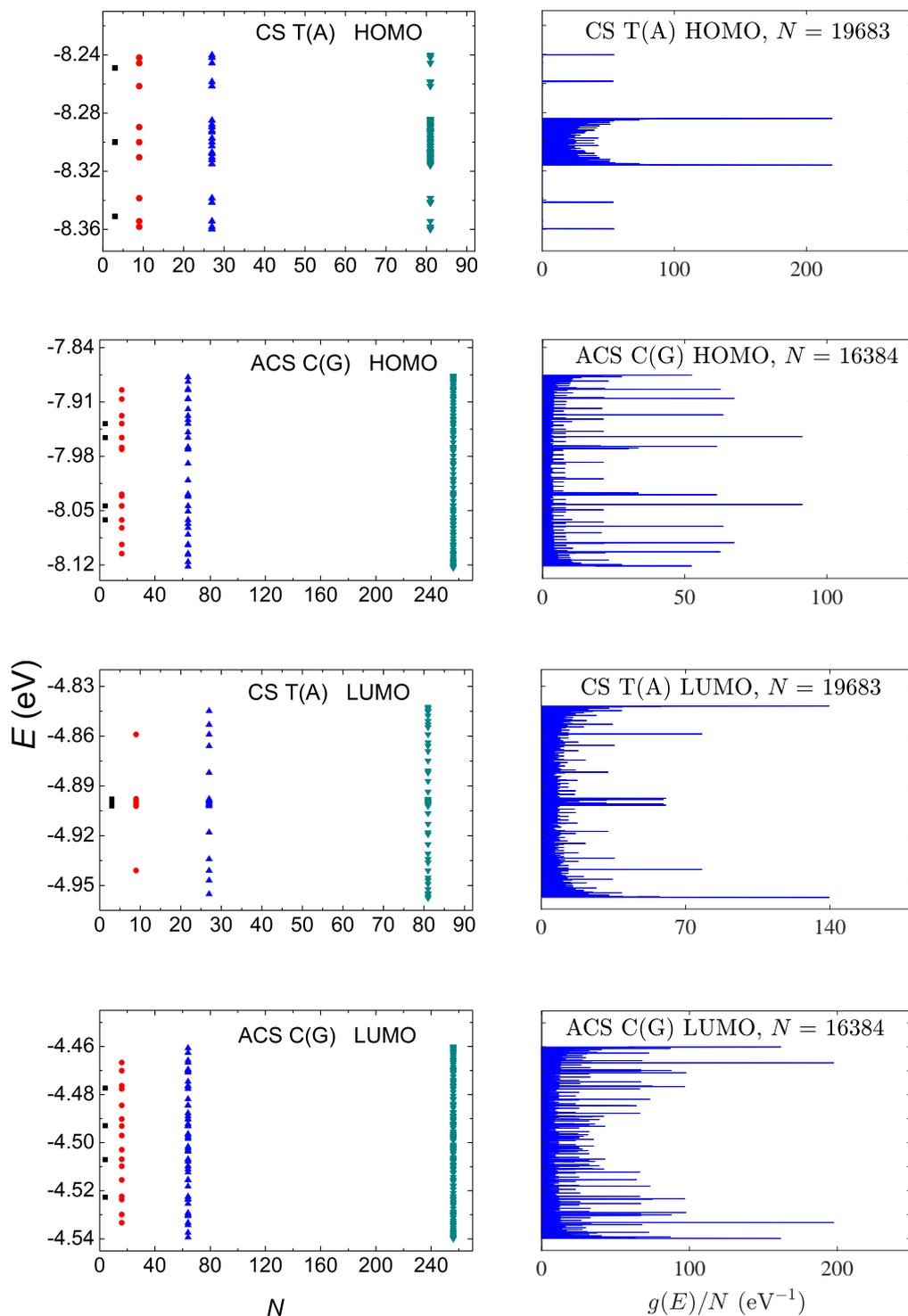
Figure 3.13: Eigenspectra of CS T(A) and ACS C(G) polymers, for the HOMO regime (upper half) and the LUMO regime (lower half), for a few generations, along with the corresponding DOS for large N.
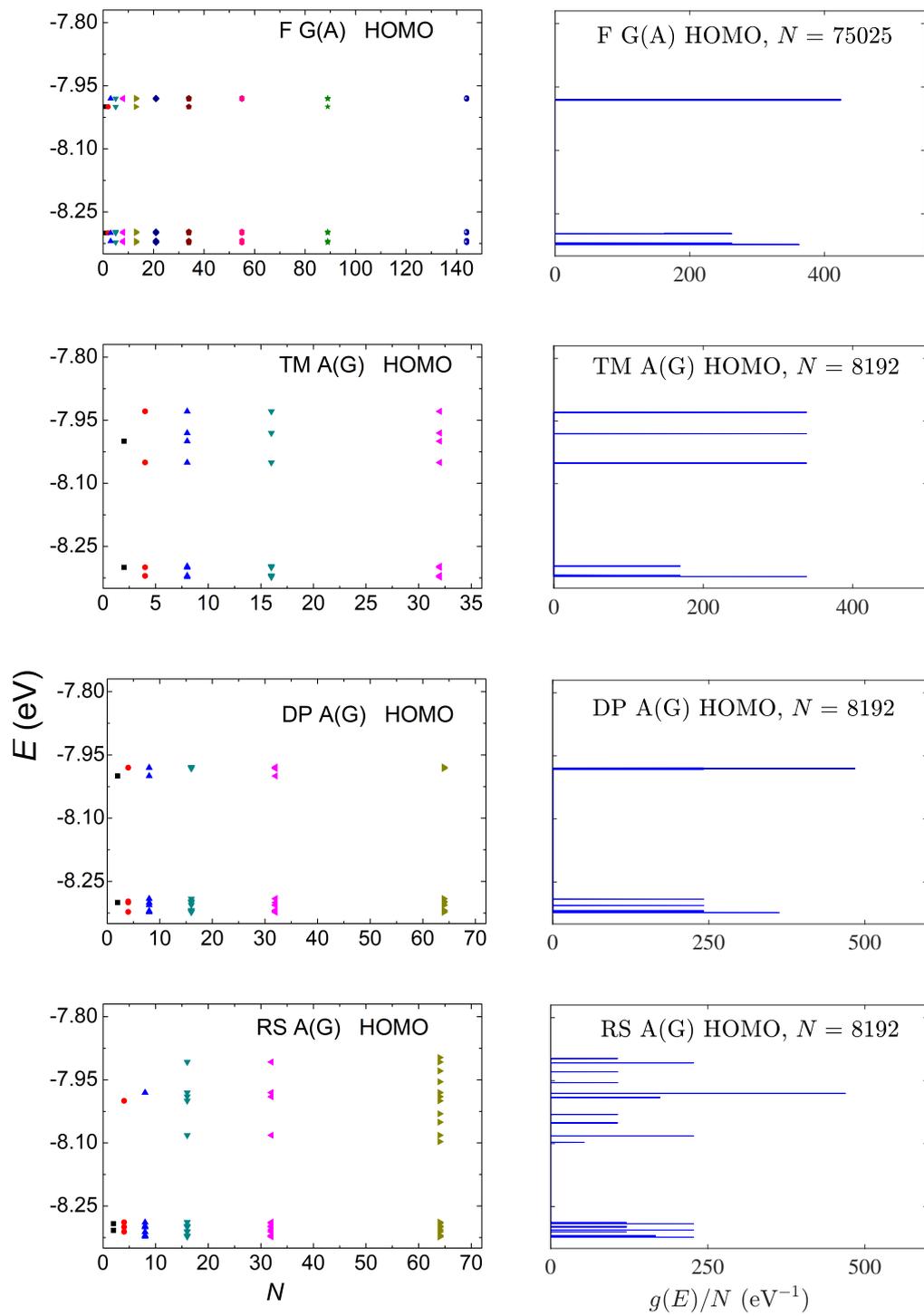
Figure 3.14: Eigenspectra of F A(G), TM A(G), DP A(G), RS A(G) polymers, for the HOMO regime, for a few generations, along with the corresponding DOS for large N.
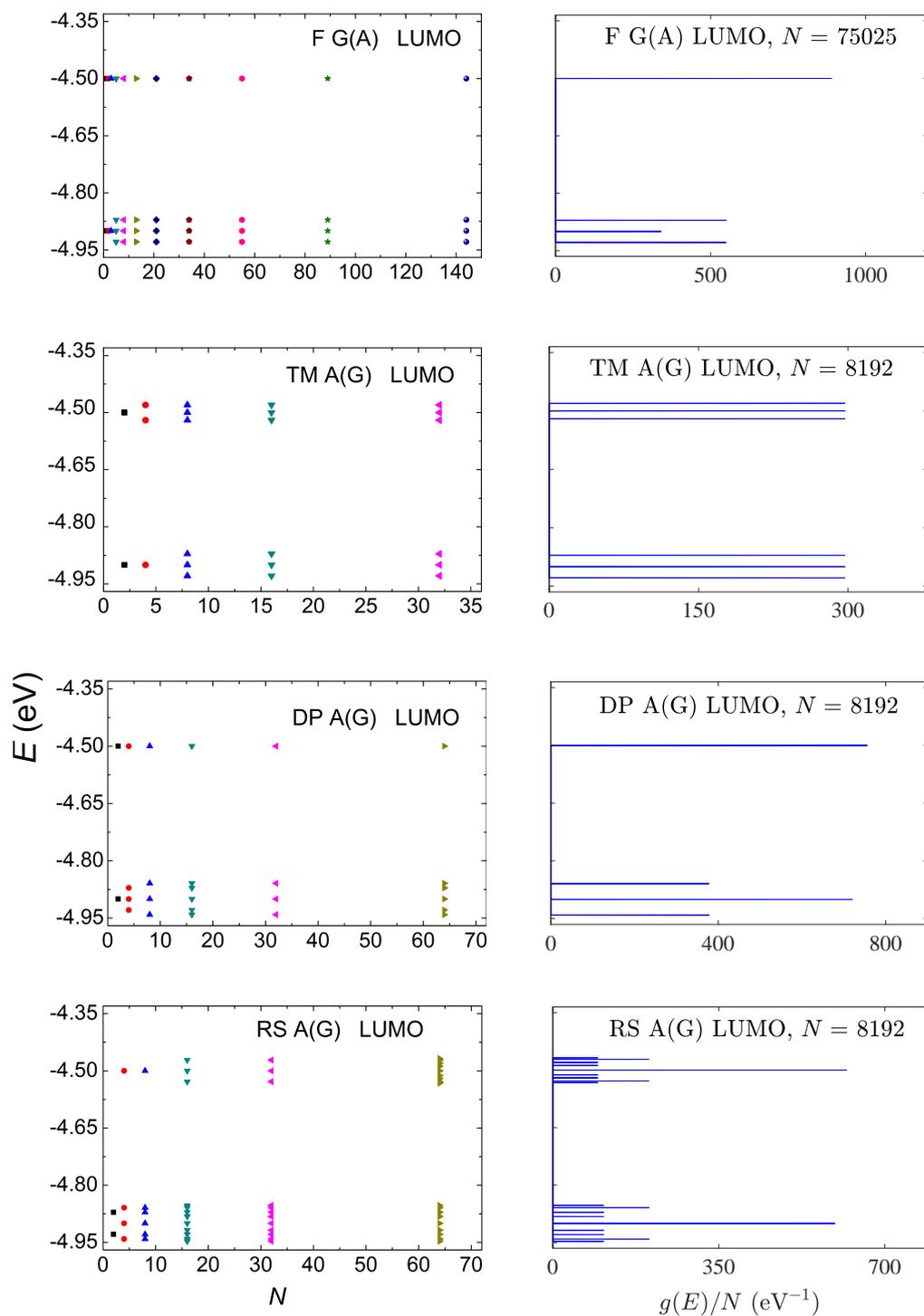
Figure 3.15: Eigenspectra of F A(G), TM A(G), DP A(G), RS A(G) polymers, for the LUMO regime, for a few generations, along with the corresponding DOS for large N.
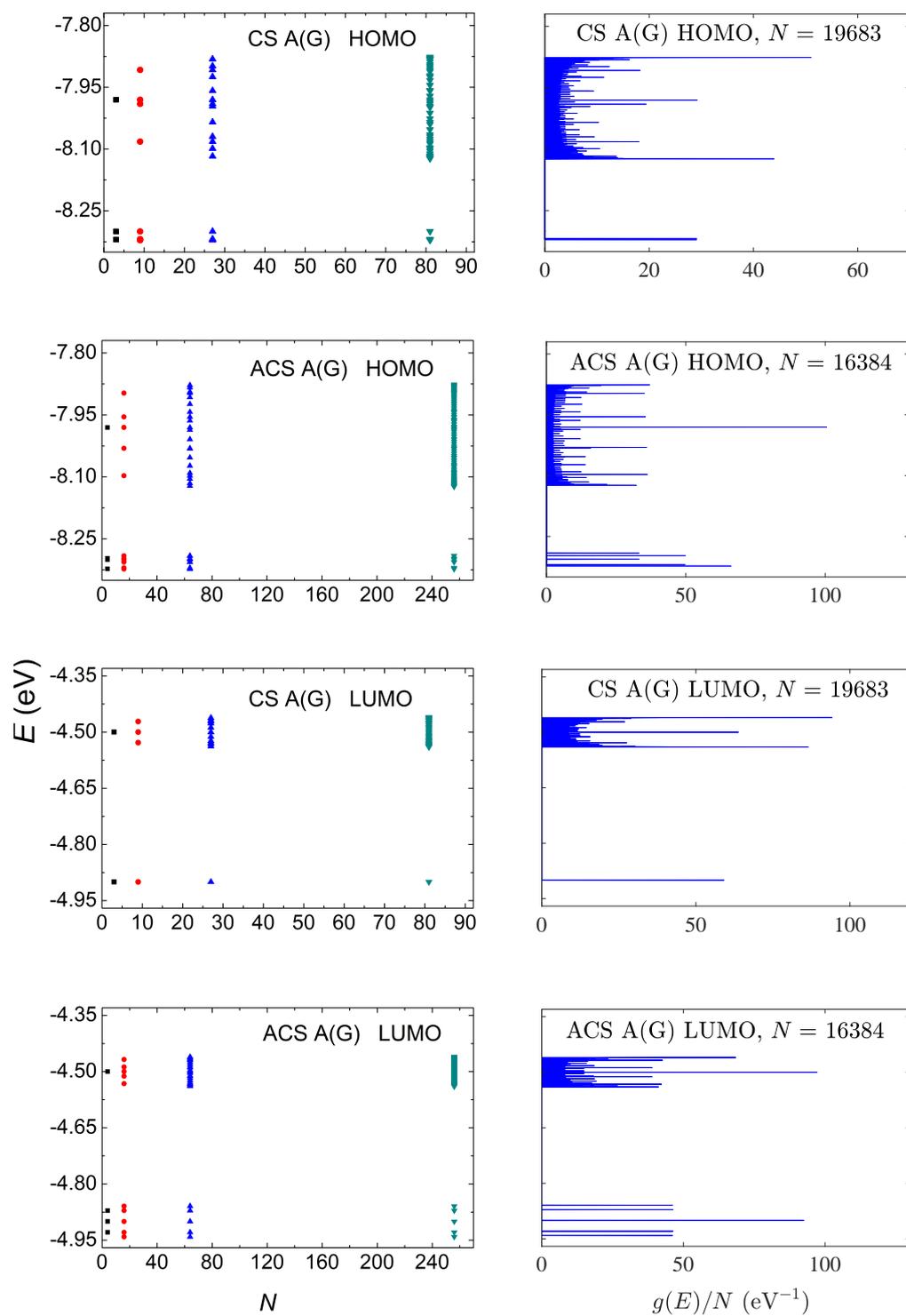
Figure 3.16: Eigenspectra of CS A(G) and ACS A(G) polymers, for the HOMO regime (upper half) and the LUMO regime (lower half), along with the corresponding DOS for large N.

The energy gap of a monomer is the difference between its LUMO and
HOMO levels. The energy gap of a polymer is the difference between
the lowest level of the LUMO regime and the highest level of the HOMO
regime, because we assume that the orbitals - one per site - which con-
tribute to the HOMO (LUMO) band are occupied (empty), since in both
possible monomers there is an even number of $2p_z$ electrons contributing
to the π stack. In Fig. 3.17 we present the energy gaps (calculated for
large N as in DOS) and the HOMO and LUMO band limits of all ape-
riodic polymers examined in this work. The G-C (A-T) monomer gap is
always greater than the gaps of I polymers made of G and C or A and
T. D polymers have smaller HOMO-LUMO gaps than I polymers (cf. left
panel of Fig. 3.17). Furthermore, the lower HOMO (LUMO) band limit
of D polymers is always between the lower and upper HOMO (LUMO)
band limit of I polymers consisted of A and T, while the upper HOMO
(LUMO) band limit of D polymers is always between the lower and upper
HOMO (LUMO) band limit of I polymers consisted of G and C (cf. right
panel of Fig. 3.17).
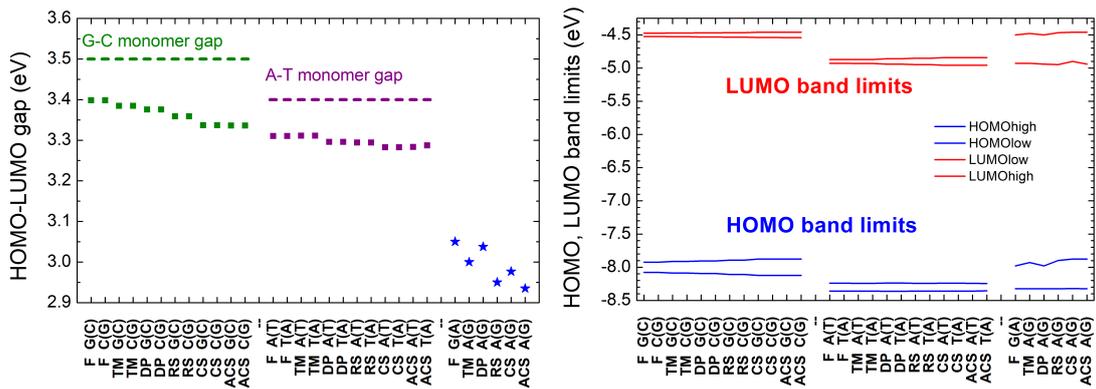
Figure 3.17: Energy gaps (left) as well as HOMO and LUMO band limits (right),
at the large N limit, for all aperiodic polymers considered in this work. Squares: I
polymers, i.e., made of the same monomer. Blue stars: D polymers, i.e., made of
different monomers. The green (purple) dashed line shows the energy gap of the G-C
(A-T) base pair.

B. Mean over time Probabilities.

The main aspects of our results for the mean over time probabilities for I
and D polymers are summarized in Figs. 3.18, 3.19, 3.20 and 3.21 (where
we display only two consecutive generations) and in Figs. 2 and 4 in Ap-
pendix C (where we show many consecutive generations), for some exam-
ple cases. We suppose that the extra carrier is initially placed at the first
monomer. A general observation is that usually these probabilities are
distributed to monomers close to the one the carrier was initially placed
at.

The mean over time probabilities of finding the extra carrier at each
monomer of a polymer depends on the sequence on-site energies and mag-
nitude of interaction parameters between successive monomers. In I poly-
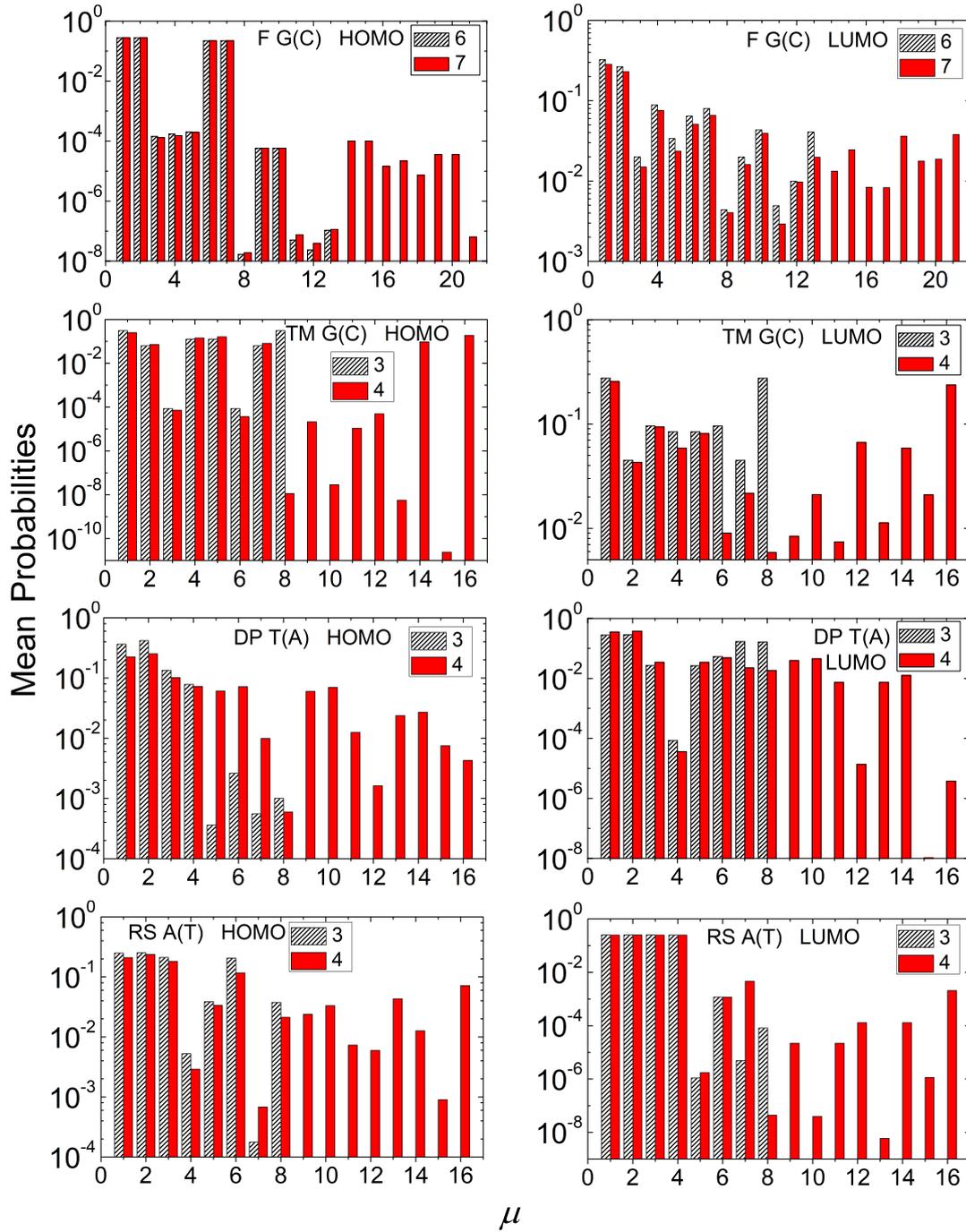
Figure 3.18: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for two consecutive generations (the number of which is denoted at each panel's legend,) for F G(C), TM G(C), DP T(A), RS A(T) polymers (quasi-periodic), for HOMO (left) and LUMO (right).

mers (cf. Figs. 3.18 and 3.19), only the interaction integrals affect the electronic structure. For the Thue-Morse G(C) polymers, the probabilities are palindromic for odd generation numbers. This is due to the fact that the Hamiltonian matrices of these polymers are palindromic, i.e., reading them from top left to bottom right and vice versa gives the same result [97]. This
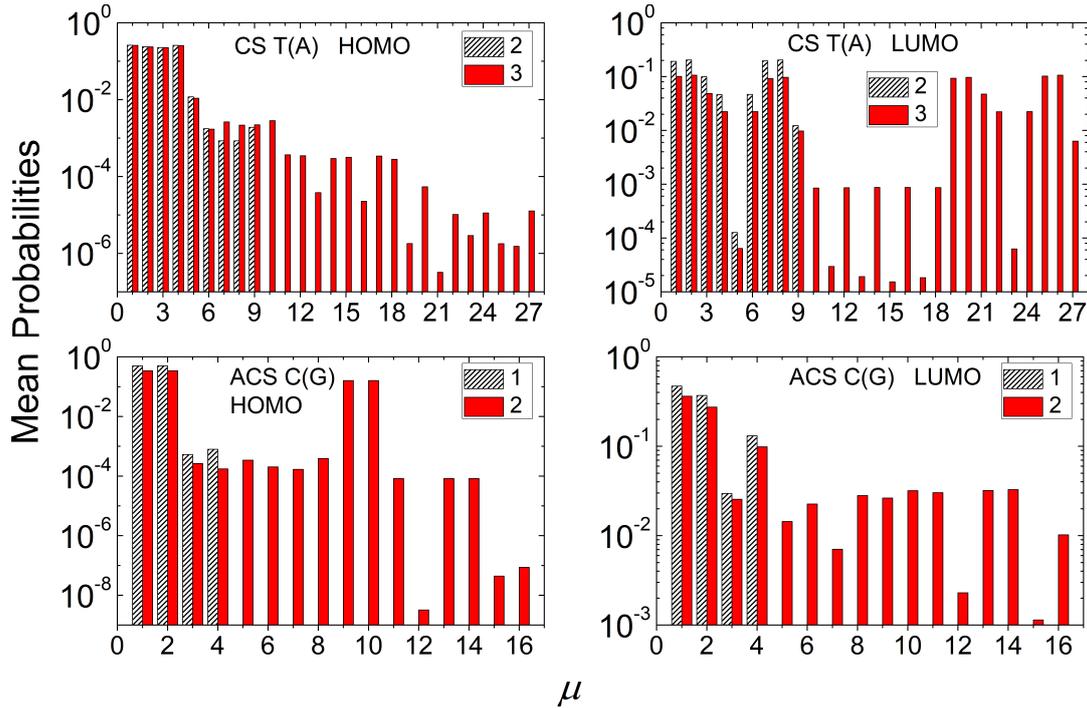
Figure 3.19: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for two consecutive generations (the number of which is denoted at each panel's legend,) for CS T(A), ACS C(G) polymers (fractal), for HOMO (left) and LUMO (right).

property stems directly from the sequence structure. For Cantor Set T(A) polymers, the mean over time probability for an extra hole is almost totally distributed at the four (or three for generation 1) starting monomers, regardless of N, while for an extra electron the probabilities are almost semi-palindromic, i.e. $\left\langle \left| A_\mu(t) \right|^2 \right\rangle = \left\langle \left| A_{N-\mu+1}(t) \right|^2 \right\rangle, \mu = 2, 4, ..., N-1$. In this case, even if the sequence structure is the same for HOMO and LUMO, the magnitude of interaction integrals has a stronger effect on the results. Another example is the Rudin-Shapiro A(T) sequence where the mean over time probability for an extra electron is almost totally distributed at the four starting monomers, regardless of N, while for holes it is basically distributed at monomers 1, 2, 3 and 6. Regarding the extra hole in Asymmetric Cantor C(G) polymers, the probability is much higher for monomers 1, 2, 9, 10 of every 32-monomer period. Generally, for I polymers, the mean over time probabilities are significant only rather close to the first monomer, although in some cases we observe non-negligible probabilities at more distant monomers.

Generally, for D polymers, the mean over time probabilities are almost negligible further than the first monomer. An exception is the Rudin-Shapiro A(G) sequence where the probabilities for both HOMO and LUMO are almost totally distributed at the three starting monomers of each polymer, regardless its length. Likewise, the mean over time probability for the extra
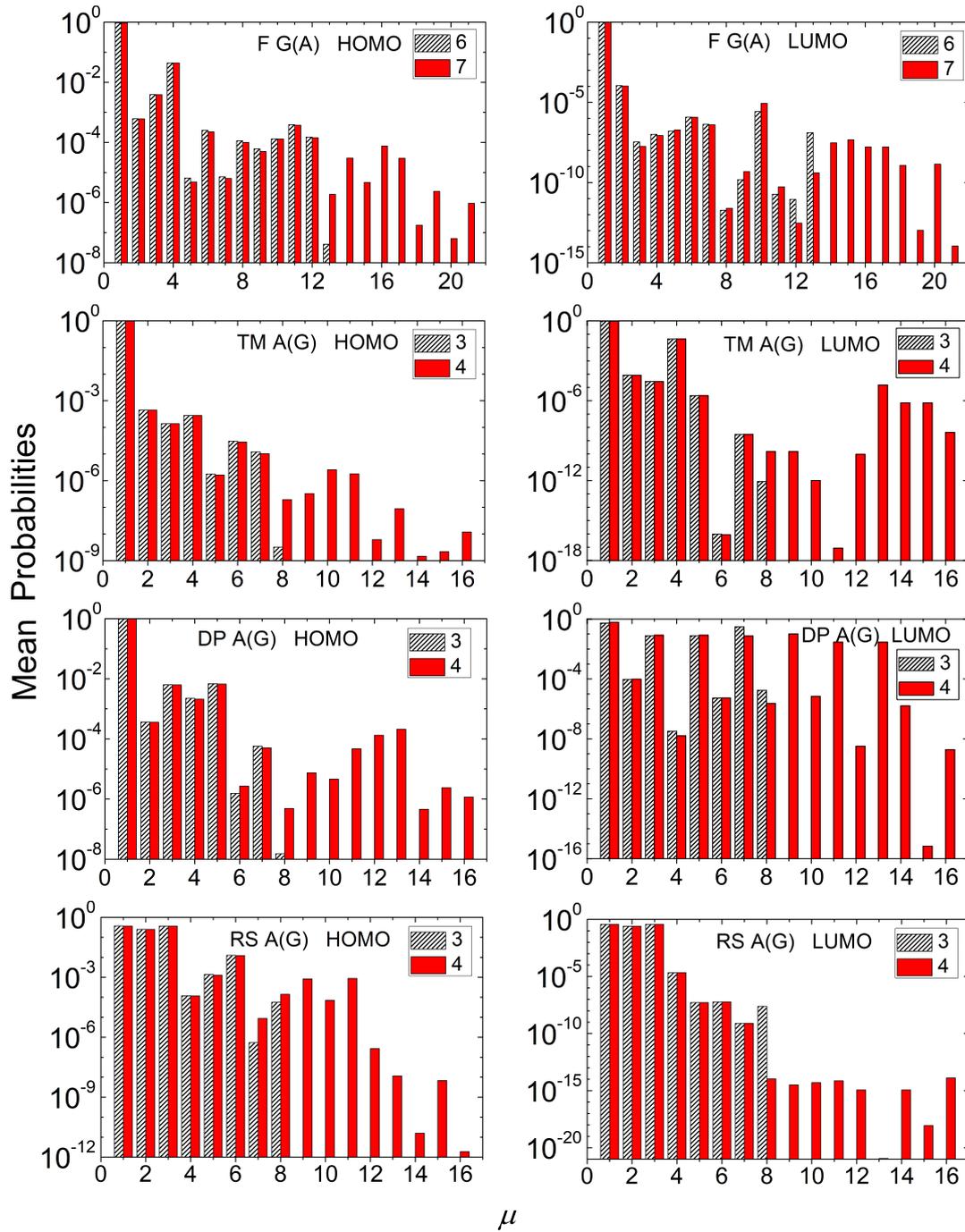
Figure 3.20: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for two consecutive generations (the number of which is denoted at each panel's legend,) for F G(A), TM A(G), DP A(G), RS A(G) polymers (quasi-periodic), for HOMO (left) and LUMO (right).

electron in Cantor Set A(G) polymers is almost totally distributed at the first and third monomer of each polymer, regardless its length. An extra electron in Double-Period A(G) reaches somehow more distant monomers.
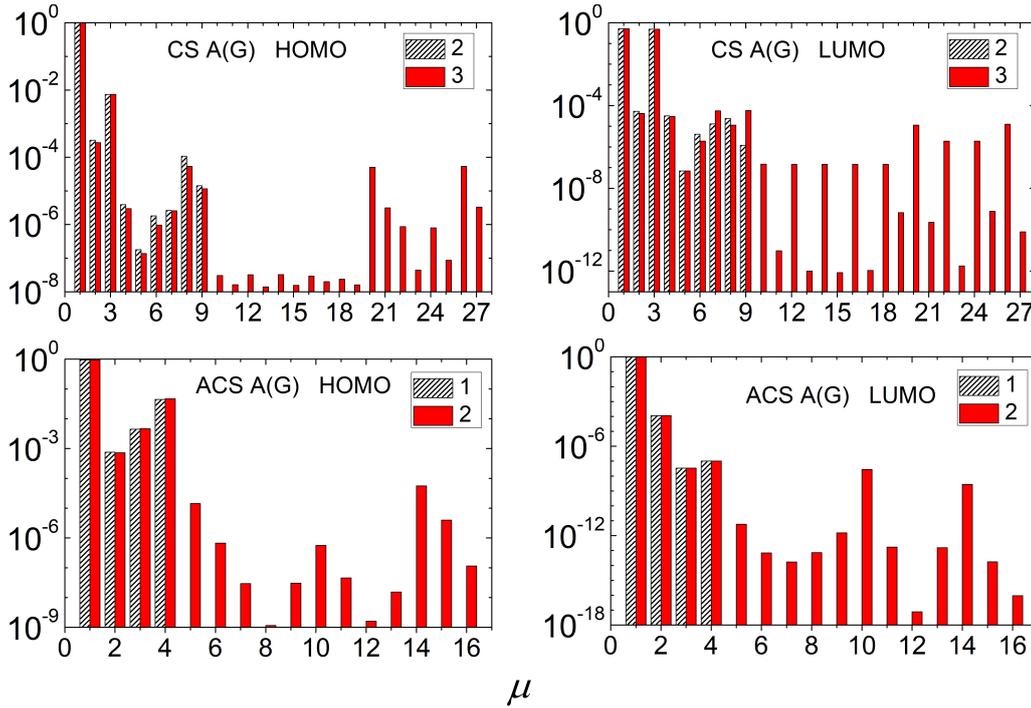
Figure 3.21: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for two consecutive generations (the number of which is denoted at each panel's legend,) for CS A(G), ACS A(G) polymers (fractal), for HOMO (left) and LUMO (right).

C. Frequency Content.

The frequencies involved in charge transfer are given by Eq. (3.27). Hence, the maximum frequency is determined by the maximum difference of eigenenergies, i.e., by the upper and lower limits of the HOMO or LUMO band (calculated for large N as in DOS). These maximum frequencies for all studied polymers are shown in Fig. 3.22.

The Fourier spectra of the time-dependent probability to find an extra electron or hole at each monomer are generally in the THz regime, mainly in the FIR and MIR part of the electromagnetic spectrum. When the dominant frequencies, i.e. those with greater Fourier amplitudes, are smaller (bigger), the carrier transfer - from the first to the last monomer - is slower (faster).

In Figs. 3.24 and 3.23 we depict the TWMF as a function of N for the various types of aperiodic polymers. We notice that the TWMF generally stabilizes as the generation number increases. In all cases, TWMF are in the region $\approx 10^{-2} - 10^2$ THz.
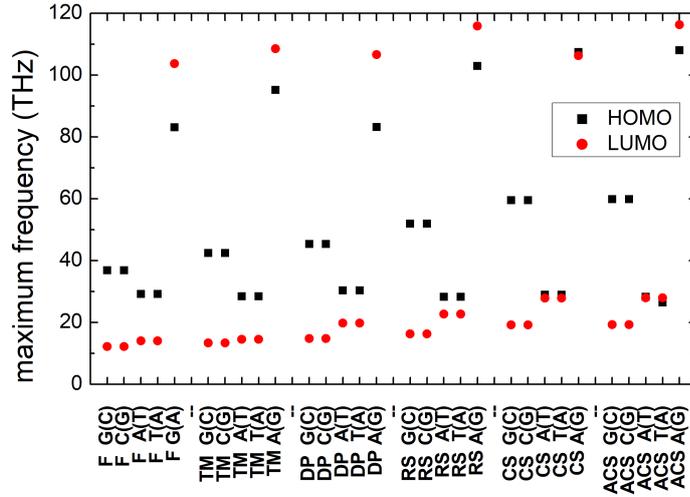
Figure 3.22: The maximum frequency of the Fourier spectrum, for the HOMO and the LUMO regime of Fibonacci, Thue-Morse, Double Period, Rudin-Shapiro, Cantor Set, Asymmetric Cantor Set polymers, at the large N limit
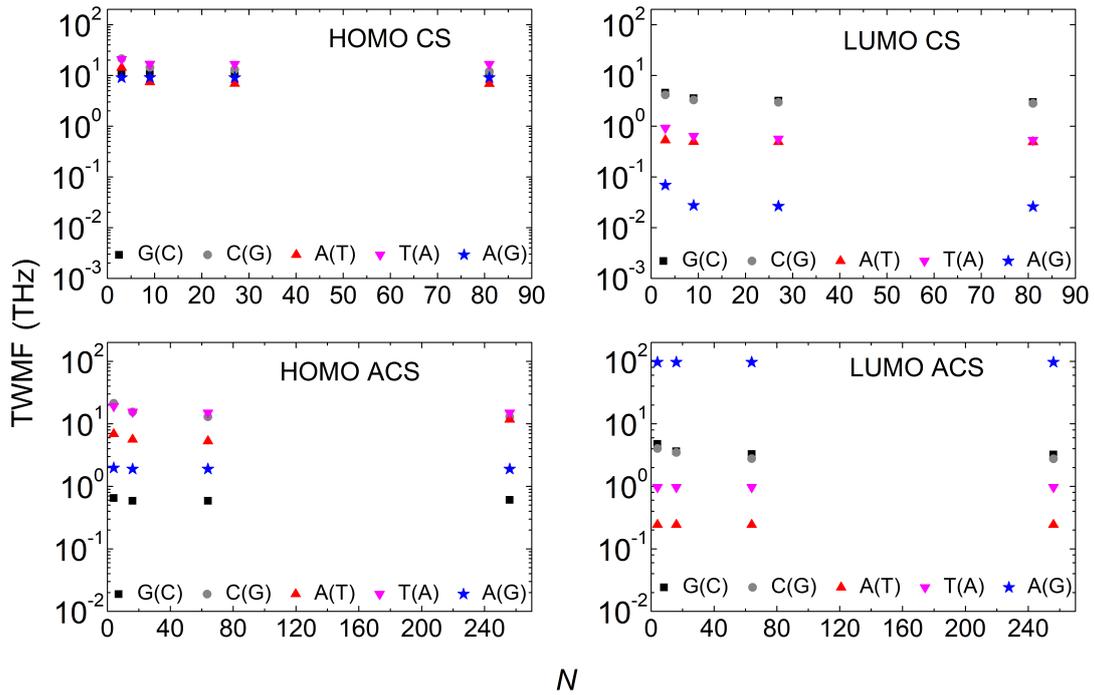
Figure 3.23: Total Weighted Mean Frequency (TWMF) as a function of the number of monomers N in the polymer, having placed the carrier initially at the first monomer, for Cantor Set, Asymmetric Cantor Set polymers (fractal), for the HOMO (left) and the LUMO (right) regime. D polymers, i.e., made of different monomers, are denoted by blue stars.
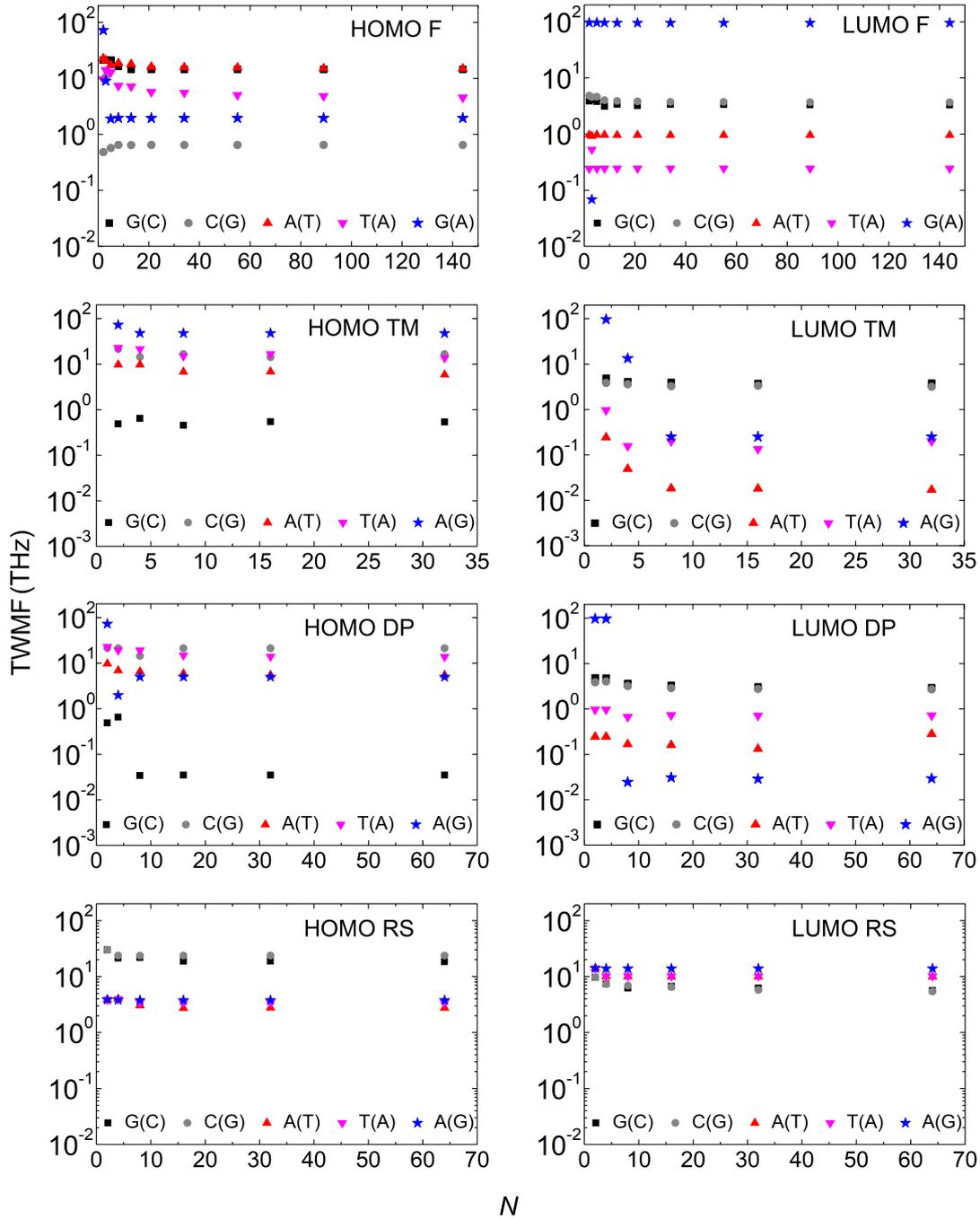
Figure 3.24: Total Weighted Mean Frequency (TWMF) as a function of the number of monomers N in the polymer, having placed the carrier initially at the first monomer, for Fibonacci, Double Period, Rudin-Shapiro polymers (quasi-periodic), for the HOMO (left) and the LUMO (right) regime. D polymers, i.e., made of different monomers, are denoted by blue stars.

D. Pure Mean Transfer Rates.

Next, we study the pure mean transfer rates from the first to the last monomer, $k_{1,N}$, or from now on, just k. We depict $k(N)$ either for HOMO or for LUMO, for I and D polymers in Figs. 3.25 and 3.26. In all cases,

$k(N)$ is a decreasing function. Generally, the degree of coherent transfer difficulty is greater for D polymers. Overall, our results suggest that I polymers, which are simpler cases in terms of energy intricacy, are more efficient regarding coherent hole and electron transfer.
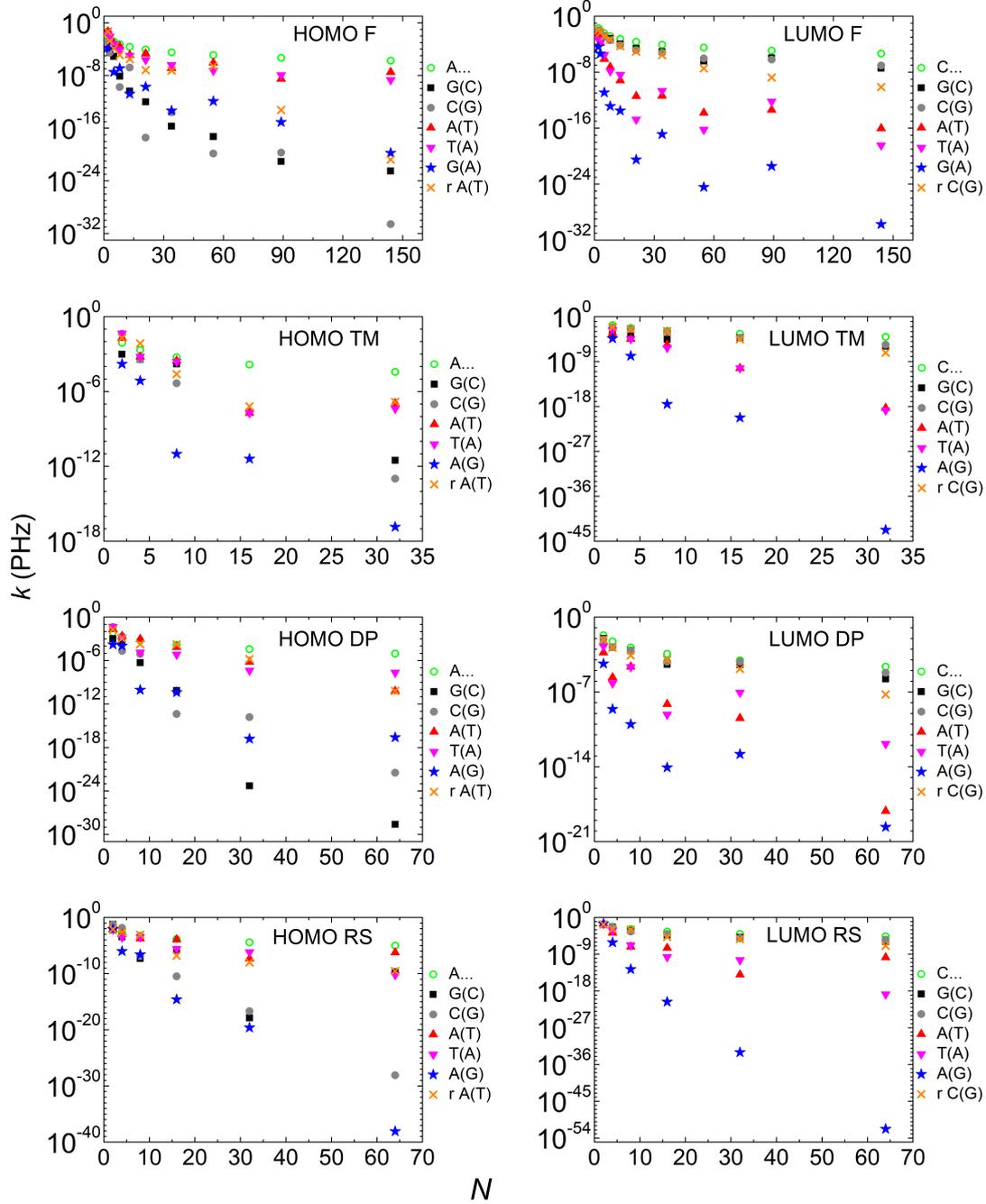
Figure 3.25: *Pure* mean transfer rates k of Fibonacci, Thue-Morse, Double Period, Rudin-Shapiro polymers (quasi-periodic), homopolymers and randomly shuffled aperiodic polymers as a function of the number of monomers N in the polymer, for the HOMO (left) and the LUMO (right) regime. By blue stars we denote D polymers, i.e., made of different monomers.

We include in each panel of Figs. 3.25 and 3.26, $k(N)$ of homopolymers
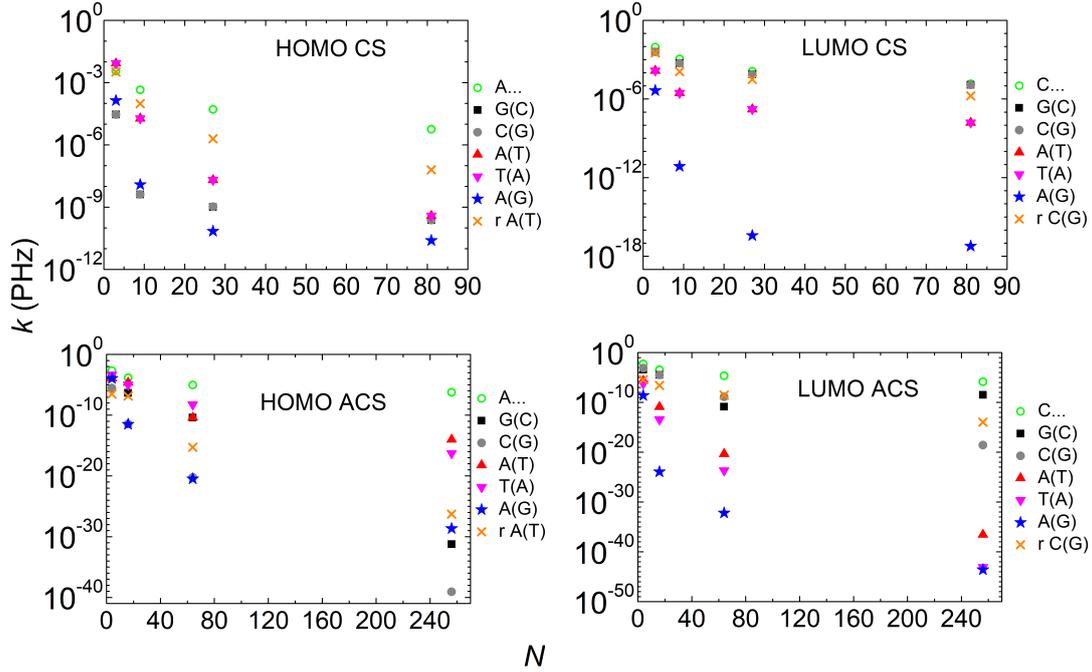
Figure 3.26: *Pure* mean transfer rates k of Cantor Set, Asymmetric Cantor Set polymers (fractal), homopolymers and randomly shuffled aperiodic polymers as a function of the number of monomers N in the polymer, for the HOMO (left) and the LUMO (right) regime. By blue stars we denote D polymers, i.e., made of different monomers.

(e.g., A...) which are the "champions" among periodic polymers in terms of efficiency of coherent carrier transfer [97] (see Subsection 3.3.1), i.e., in terms of magnitude of k and of slower decrease of $k(N)$. It seems that $k(N)$ of homopolymers is an unreachable limit for aperiodic polymers. Comparing periodic polymers [97] (Subsec. 3.3.1) with aperiodic polymers in terms of $k(N)$, we realize that although generally periodic polymers are more efficient, specific aperiodic polymers can be better than specific periodic ones.

In each panel of Fig. 3.25 and 3.26, we also take the best of aperiodic polymers in terms of $k(N)$ and shuffle randomly the sequence of its monomers. In all cases, except for Cantor Set HOMO, this random shuffle deteriorates severely $k(N)$. For Cantor Set, A(T) and T(A) have identical $k(N)$ because the Cantor Set rules for A(T) and T(A) produce equivalent polymers, cf. Eq. (3.22). For equivalent polymers, $k(N)$ from the first to the last monomer are identical, cf. Eq. (3.40). For example, TAT ≡ ATA, TATAAATAT ≡ ATATTTATA, TATAAATATAAAAAAAATATAAAT-AT ≡ ATATTTATATTTTTTTTATATTTATA and so on. Similarly, the Cantor Set rules for G(C) and C(G) produce equivalent polymers, which have identical $k(N)$. In Cantor Set HOMO, the best sequences in terms of $k(N)$ are A(T) and T(A), where the interaction integrals involved are $t_{AA} = t_{TT} = -8$ meV, $t_{AT} = 20$ meV, $t_{TA} = 47$ meV, and we have just one on-site energy, that of A-T. From these interaction integrals, t_{AA} has the smallest absolute value. Given the structure of the Cantor Set sequences,

making the random shuffle, the number of t_{AA} decreases, while the numbers of the bigger interaction integrals, t_{AT} and t_{TA} increase. For this reason, in Cantor Set HOMO, the random shuffle increases $k(N)$. In Cantor Set LUMO, this argument is inverted because now the best sequences in terms of $k(N)$ are G(C) and C(G), where the interaction integrals involved are $t_{GG} = t_{CC} = 20$ meV, $t_{GC} = -10$ meV, $t_{CG} = -8$ meV, and we have just one on-site energy, that of G-C. In this case, the random shuffle decreases the number of the bigger interaction integrals $t_{GG} = t_{CC}$ and decreases the numbers of the smaller interaction integrals t_{GC} and t_{CG}. However, apart from the exception of the Cantor set HOMO, generally speaking, the conclusion is that aperiodic polymers posses some kind of order, i.e., a well-defined construction rule that makes them more efficient than random polymers in terms of $k(N)$; therefore, when this rule is destroyed, the transfer efficiency diminishes.

E. Transfer Rates in Experiments.
Comparison of the coherent pure mean transfer rates k of our prototype system, B-DNA, with experimentally obtained transfer rates is a rather complicated issue. In the past, the experimental transfer rates in donor - bridge (DNA) - acceptor systems were obtained using the concentrations of different products generated e.g. when a hole is (PY) or is not (PN) transferred. The concentrations of PY and PN were indirectly measured by methods like polyacrylamide gel electrophoresis and piperidine treatment [164, 165]. Although these methods revealed some aspects of hole transfer like the sequence dependence and the ability of transfer, they do not provide the kinetics of hole transfer in DNA [7]. Although, generally, greater concentration of PY implies greater charge transfer, there is no proof that the concentrations of PN and PY are proportional to the degree of transfer.

Quantum mechanically, only a fraction of the carrier reaches the acceptor through the bridge. For the same reason, the definition of transfer time is problematic. The transfer rate should depend both on the amount and the speed of transfer. However, the concentration of PY is not strictly proportional to the amount of carrier transfer and not strictly inversely proportional to the time of transfer. A more direct experimental approach is time-resolved spectroscopy, e.g. transient absorption, to observe the products of charge transfer [7, 166, 167].

Our point of view is different, since the quantity we use, the *pure* mean transfer rate [115], given by Eq. 3.39, uses simultaneously the magnitude of coherent charge transfer and the time scale of the phenomenon. However, our method applies to coherent transfer only and cannot cover incoherent mechanisms like thermal hopping.

It is a common assertion in the literature that when the fall of the transfer rate with respect to the length of a given DNA segment is described by an exponential fit, the mechanism of transfer is superexchange, whereas when it is described by a power law fit, the mechanism of transfer is multi-step

hopping. However, we stress that the fitted parameters produced this way should be treated with care, especially when it comes to attributing them to specific mechanisms. For example, in Ref. [7], where the hole transfer kinetics of various short DNA segments were experimentally investigated with time-resolved spectroscopy, the authors present an exponential decay length $\beta = 1.6$ Å^{-1} by fitting the experimental hole transfer rates of $G(A)_nG$ DNA oligomers ($n = 0, 1, 2$) to the exponential law $K = K_0 e^{-\beta d}$, where d is the charge transfer distance, i.e., $d = 3.4 \times (N-1)$ Å. Using the transfer rate values of Ref. [7], we observed that, although β, determined as the slope of the linear fit $\ln(K) = \ln(K_0) - \beta d$ is indeed $\cong 1.6$ Å^{-1}, a direct exponential fit gives $\beta \cong 1.3$ Å^{-1}, suggesting that the law of decay is not exactly exponential. On the contrary, the fits of our theoretically obtained *pure* mean transfer rates, k, for the same system, give $\beta \cong 1.84$ Å^{-1} for β determined as the slope of the linear fit $\ln(k) = \ln(k_0) - \beta d$, and $\beta \cong 1.79$ Å^{-1} for a direct exponential fit $k = k_0 e^{-\beta d}$, suggesting closer convergence to an exponential decay. Similarly, in Ref. [168], the authors experimentally study, with time-resolved spectroscopy, hole transfer through $(GA)_n$ and $(GT)_n$ sequences, where $n = 2$-12 is the number of repetition units. The authors fitted the obtained transfer rates to the power law $K = K_0' \mathcal{N}^{-\eta}$, where $\mathcal{N}$ is the number of hopping steps between guanines (in our notation, $\mathcal{N} = \frac{N}{2} - 1$), reported the same exponent for both sequences, i.e. $\eta = 2$, and suggested that this value provides evidence that the long-distance hole transfer occurs by multi-step hopping between guanines. From the rate values provided in Table I of Ref. [168], we observed that, although η as a slope of the linear fit $\ln(K) = \ln\left(K_0'\right) - \eta \ln(\mathcal{N})$ is indeed 2 for both sequences, a direct power law fit yields $\eta \cong 1.4$ for $(GA)_n$ and $\eta \cong 1.3$ for $(GT)_n$, suggesting that the rate decay does not follow exactly a power law. On the contrary, the fits of our theoretically obtained *pure* mean transfer rates, k, for $(GA)_n$, give $\eta \cong 1.40$ for η determined as the slope of the linear fit $\ln(k) = \ln\left(k_0'\right) - \eta \ln(\mathcal{N})$, and $\eta \cong 1.56$ Å^{-1} for a direct power law fit $k = k_0' \mathcal{N}^{-\eta}$. The respective values for $(GT)_n$ are $\eta \cong 2$ for both fits. Hence, our theoretical results suggest that the fall of k, as the length of the bridge increases, convergences to a power law and that the fall of the transfer rate is less steep when purines are on the same strand compared to the case when purines are crosswise.

DNA is a dynamical structure, i.e., the geometry is not fixed. Large variations of the TB parameters are expected in real situations and also, large variations of the TB parameters have been obtained by different theoretical methods by different authors, cf. e.g. Ref. [115] and references therein. Hence, the parameters any TB model uses have to be utilized with care. In Ref. [169], the authors report experimentally deduced (by transient absorption spectroscopy) charge separation rates, in capped A_n ($n = $1-7) and A_3G_n ($n = $1-19) DNA hairpins with a stilbenedicarboxamide hole donor and a stilbenediether hole acceptor. We computed our theoretical coherent pure mean transfer rates, k, for the same systems with a modified

parametrization: $t_{AA} \to 1.6 t_{AA}$, $t_{AG} \to 2.1 t_{AG}$, $t_{GG} \to 2.25 t_{AG}$ (cf. Table 3.3). In order to mimic the donor and the acceptor, we added two sites at the ends of the TB chain, with on-site energies $E_{\mathrm{don}} = E_{A-T} - 0.1$ eV, $E_{\mathrm{ac}} = E_{G-C} + 0.1$ eV. We used for the interaction integral from the donor (last base pair) to the first base pair (acceptor) 100 meV (250 meV). Our results, along with the experimental ones, are depicted in Fig. 3.27. Apart from the A_1 and A_2 systems, for which we find much larger rates, the pure mean transfer rates k are of the same order of magnitude, in good quantitative agreement with the experimental transfer rates K. Actually, the same sequences A_n (n =1-7) and A_3G_n (n =1-19) analyzed in Ref. [169] had also been analyzed by the same group in Ref. [170]. In Ref. [170], the authors mention a time resolution of ca. 180 fs. Hence, roughly, only transfer rates $K < (1/180)$ PHz $\approx (1/200)$ PHz $= 5 \times 10^{-3}$ PHz can be detected by this technique.
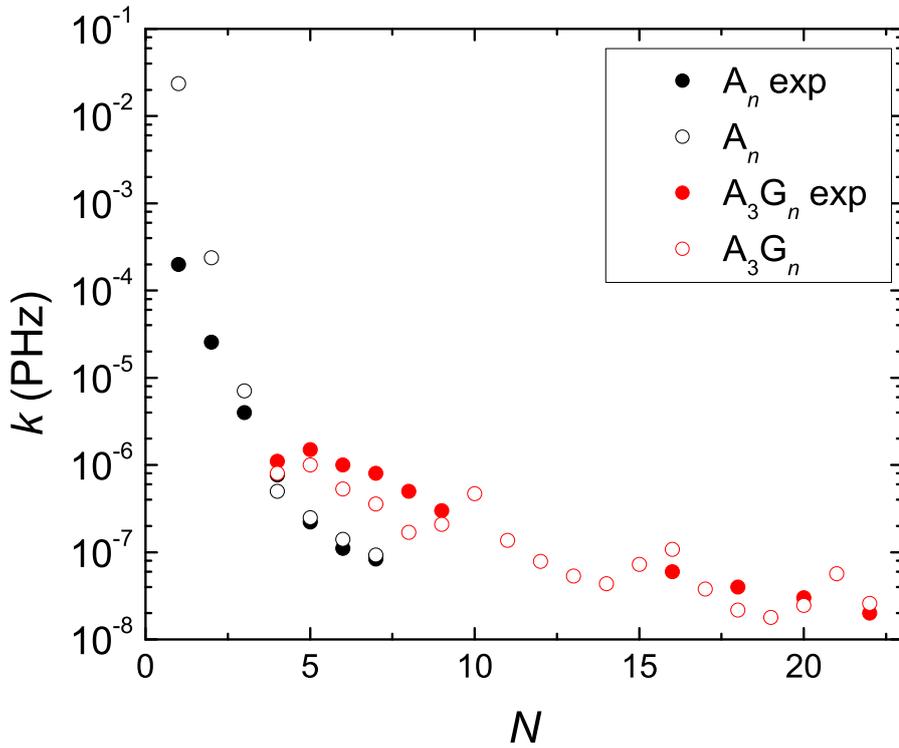
Figure 3.27: Comparison of experimental hole transfer rates K for A_n and A_3G_n segments [169] (full circles) with our theoretical coherent pure mean transfer rates k (empty circles), as a function of the number of monomers in the polymer N. The TB parametrization is described in the main text.

4 | STRUCTURAL VARIABILITY[1]

In canonical B-DNA, the most common DNA form, successive pairs of the four nitrogenous bases, Guanine (G) always paired with Cytosine (C) and Adenine (A) always paired with Thymine (T), are separated and twisted approximately by 3.4 Å and 36°, respectively, relative to the (right-handed) double helix growth axis. But this is an ideal situation. In fact, the shape of the duplex can be analyzed through a set of parameters that have been conventionally defined to this purpose [171].

Local complementary base-pair parameters are employed in order to define the base pair structure and its variability. The parameters describing the relative translations in all axes, involving two bases of a Watson-Crick pair, are *shear* (Sx), *stretch* (Sy) and *stagger* (Sz), while the corresponding rotations around x, y and z axes are *buckle* (κ), *propeller twist* (π) and *opening* (σ) [172]. Figure 4.1 depicts the definitions of these translation and rotation parameters involving two bases of a Watson-Crick pair.
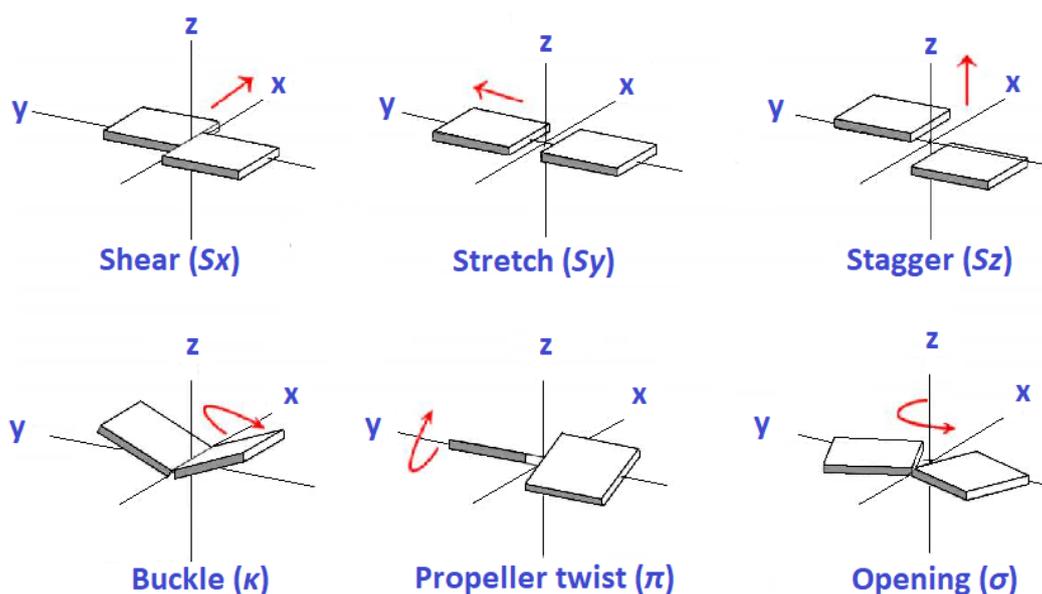
Figure 4.1: Definitions of translation parameters (**top row**) and rotation parameters (**bottom row**) involving two bases of a base pair.

Fig. 4.2 sketches some of the helix shape parameters for nucleic acids (step

---

parameters); distances *shift* (Δx), *slide* (Δy), *rise* (Δz), and angles *tilt* (τ), *roll* (ρ), *twist* (ω). These distances and angles define the relative arrangement of the two monomers in a stacked base pair step.
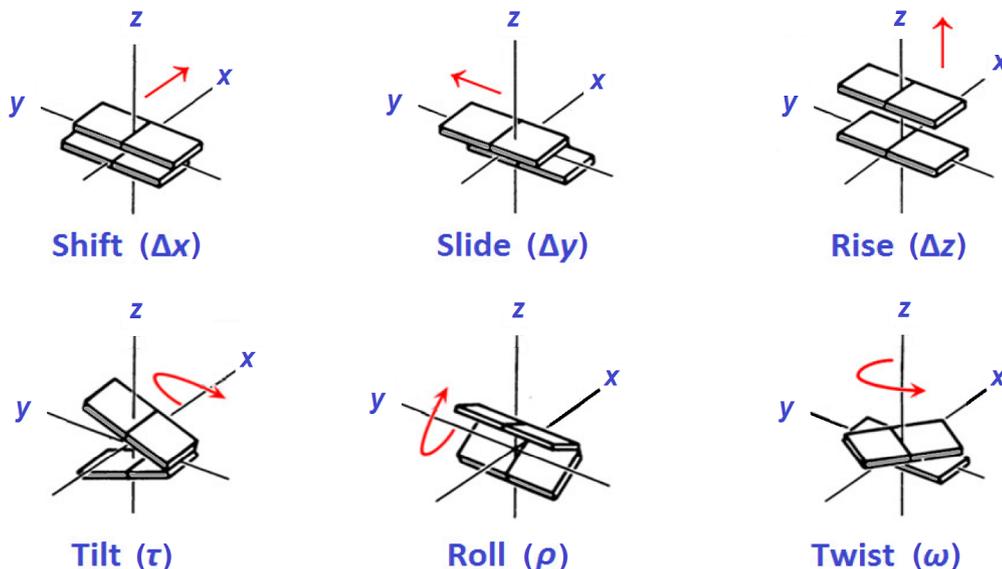
Figure 4.2: Definitions of inter-base-pair helix shape parameters shift (Δx), slide (Δy), rise (Δz), tilt (τ), roll (ρ), twist (ω).

In order to address the effects of structural variability in the electronic structure of B-DNA within the LCAO approach we use deformed base pairs pruned from a 500 ns Molecular Dynamics (MD) trajectory of a 20mer [18, 27].

The rest of this Chapter is organized as follows. In Sec. 4.1 we present the MD method and calculations performed to obtain different geometries of stacked base pairs (dimers) comprised of the same bases on the same strand (AA and GG). In Sec. 4.2 we analyze the effects of structural variability on the electronic structure and charge transfer properties of B-DNA, using the fragments derived from MD. Initially we present the results obtained by the TB method using all valence orbitals [27], then discuss relevant RT-TDDFT results for the different dimer conformations (Subsec. 4.2.1), and finally compare the results from different methods (DFT, TB) for some hole transfer properties (Subsec. 4.2.2).

4.1 Molecular Dynamics simulation of a 20-mer

MD is currently retained as the method of choice to simulate the dynamics of biological molecules on the time scale from tens of nanoseconds to few microseconds [173--176]. Its application to proteins and nucleic acids, as well as lipid membranes, has greatly contributed to the development of the field of structural biology [177--179]. In the context of nucleic acids, we remark the work done by the Ascona B-DNA consortium [180--183], as well as discovery of DNA shape effects in protein-DNA interactions [178, 184, 185].

We simulated a double-stranded DNA 20-mer that contains multiple instances of the AA and GG dimers on which we focus our interest for charge transfer properties [18]. The sequence of the simulated duplex is $5'-$CGAAAAGGGG-AAAAGGGGAT$-3'$, see Table 4.1. The starting configuration for the 20-mer

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
$5'$-	C	G	A	A	A	A	G	G	G	G	A	A	A	A	G	G	G	G	A	T	-$3'$
$3'$-	G	C	T	T	T	T	C	C	C	C	T	T	T	T	C	C	C	C	T	A	-$5'$

Table 4.1: Sequence of the simulated duplex B-DNA oligomer

was constructed with the web interface 3DNA, as regular B-DNA. The B-DNA molecule was located at the center of a periodically repeated supercell and the supercell was filled with TIP3P [186] water molecules. The depth of water separating neighboring replicas was at least 3 nm in each direction. After a standard minimization-equilibration protocol [187] to bring the system at 300 K and 1 bar, the trajectory of the nucleic acid was collected for 500 ns. The simulation was carried out with the open-source software GROMACS [188] 5.0.4, using the amber99sb-parmbsc1 force field for the nucleic acid. [189].

Using GROMACS tools and the GROMOS algorithm [190], we clusterized the instantaneous structures that form the trajectory. We obtained seven clusters that cumulatively represent 71% of the entire trajectory. The centroid structures of the two most populated clusters are labeled as cl1 and cl2 in the following; they represent 35% and 12% ot the trajectory, respectively. The three-dimensional structures of cl1 and cl2 are shown in Fig. 4.3, in black and red, respectively. Although there are clear differences in the shape of the duplex between cl1 and cl2, the B-DNA motif is preserved in the dynamics.

From cl1 and cl2 we extracted all the possible GG and AA base pair steps (or dimers in the following), excluding the A3A4 and G17G18 edge dimers. Furthermore, we removed the backbone and saturated the dangling bonds with H atoms. It has been shown that aspects of charge transport such as coherent transmission can be altered by the inclusion of the backbone in 15 base-pair oligomers [191], and recent experiments suggest that long-distance conduction is mediated by the backbone [192]; however the DNA segments (dimers) studied here are small and the effect of the backbone is not expected to be that significant. AA and GG dimers were chosen as subjects of study since, in principle, are

the most efficient in terms of charge transfer due to the structural resemblance of their constituent base pairs, so that the effect of structural variability can be more clearly demonstrated.

The extracted dimers are labeled as: A4A5_cl1, A4A5_cl2, A5A6_cl1, A5A6_cl2, G7G8_cl1, G7G8_cl2, G8G9_cl1, G8G9_cl2, G9G10_cl1, G9G10_cl2, A11A12_cl1, A11A12_cl2, A12A13_cl1, A12A13_cl2, A13A14_cl1, A13A14_cl2, G15G16_cl1, G15G16_cl2, G16G17_cl1, G16G17_cl2.

The distances Δx, Δy, Δz and the angles τ, ρ, ω of cl1 and cl2, are shown in Fig. 4.4. The obtained values are consistent with finite-temperature variations of the B-DNA form. We denote the corresponding monomers as: A4_cl1, A4_cl2, A5_cl1, A5_cl2, A6_cl1, A6_cl2, G7_cl1, G7_cl2, G8_cl1, G8_cl2, G9_cl1, G9_cl2, G10_cl1, G10_cl2, A11_cl1, A11_cl2, A12_cl1, A12_cl2, A13_cl1, A13_cl2, A14_cl1, A14_cl2, G15_cl1, G15_cl2, G16_cl1, G16_cl2, G17_cl1, G17_cl2. Figure 4.5 sketches the translation and rotation parameters for each one of the studied monomers. The parameters were computed using the web interface 3DNA. Dashed lines denote the mean value of each parameter, that is: 0.03 Å (shear), -0.03 Å (stretch), 0.04 Å (stagger), $6.53°$ (buckle), $-10.40°$ (propeller twist), $1.06°$ (opening) for A-T monomers and -0.09 Å (shear), -0.04 Å (stretch), 0.01 Å (stagger), $0.55°$ (buckle), $-1.13°$ (propeller twist), $-0.66°$ (opening) for G-C monomers. These values together with values found in the literature are listed in Table 4.2.

Figure 4.3: Visualization of the two most representative structures from the MD trajectory; from these structures, duplex AA and GG fragments for the TDDFT calculations were extracted.
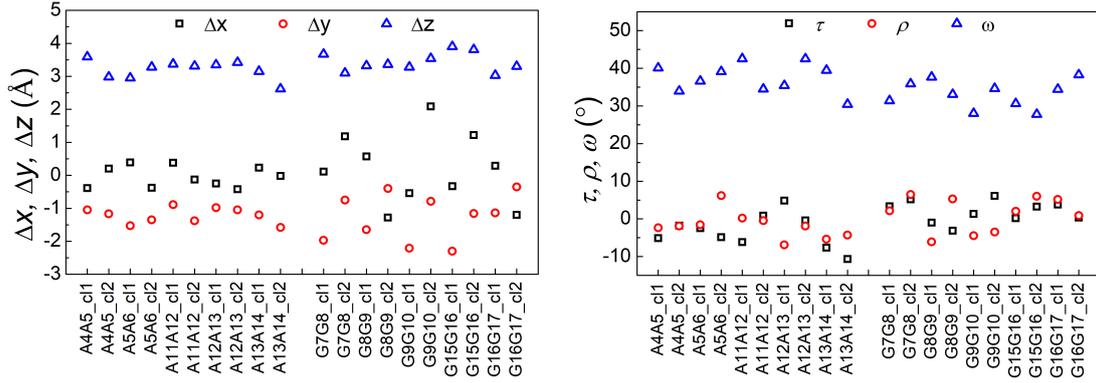
Figure 4.4: Shift (Δx), slide (Δy), rise (Δz), tilt (τ), roll (ρ), twist (ω) of the studied dimers.
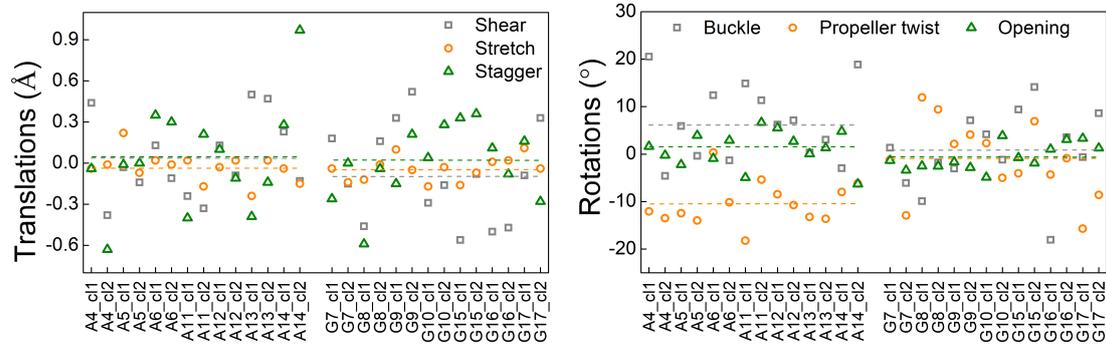
Figure 4.5: Translation [shear (Sx), stretch (Sy), stagger (Sz)] and rotation [buckle (κ), propeller twist (π), opening (σ)] parameters for all studied monomers. Dashed lines denote the mean value of each parameter.

Table 4.2: The second and third column contain the mean values of the translation and rotation parameters for the monomers A-T and G-C, as studied here [27]. The other columns list values found in bibliography.

Parameter	A-T	G-C	[193]	[194]	[195]	[196]
shear (Å)	0.03	−0.09	0.00	−0.04		
stretch (Å)	−0.03	−0.04	−0.15	−0.17		
stagger (Å)	0.04	0.01	0.09	0.21		
buckle (°)	6.53	0.55	0.5	0.3	$(-7.5, 7.5)$	
propeller twist (°)	−10.40	−1.13	−11.4	−13.7	11.5	-12.60 ± 3.2
opening (°)	1.06	−0.66	0.6	1.0	$(-2, 2)$	

4.2 Effects of Structural Variability

In this subsection we analyze the effects of structural variability on the electronic structure and charge transfer properties of B-DNA, using the fragments derived from MD, as detailed in Section 4.1. In Figure 4.6 we present the absolute values of the parameters Δ (difference between the HOMO eigenenergies of the two base pairs of each studied dimer) and t (transfer integral between the two base pairs' HOMOs of each studied dimer), as well as the maximum transfer percentages p as calculated via Equation (3.12) with the parameterization employing all valence orbitals [27]. The values of $|t|$ and p can also be found in Ref. [18] in comparison with results obtained by Density Functional Theory (DFT) techniques. From Equation (3.12) it is expected that ideal dimers (made
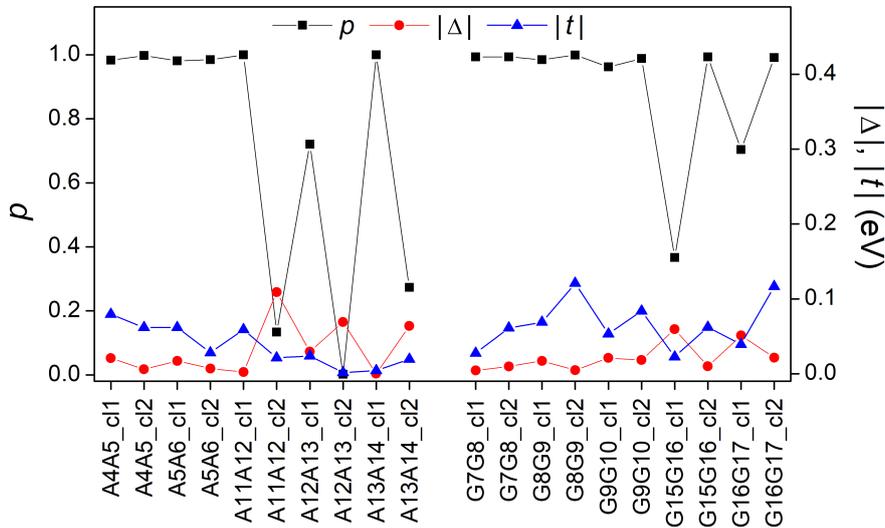
Figure 4.6: The parameters $|\Delta|$ and $|t|$, as well as the maximum transfer percentage p for all the dimers of the MD oligomer.

up of ideal monomers) should have a maximum transfer percentage equal to 1. However, by observing Figure 4.6 one can notice that not all AA and GG dimers have $p = 1$. Specifically, dimers with a p considerably different than unit (and a Δ different than zero) are: A11A12_cl2, A12A13_cl1, A121A13_cl2, A13A14_cl2, G15G16_cl1, G16G17_cl1. This is well expected because the studied monomers are not ideal, that means their consisting bases have relative translations and rotations as depicted in Figure 4.1. More specifically, a small p value is related to a large Δ value, in accordance with Equation (3.12). Thus, it is expected that the structural base-pair parameters (shear, stretch, stagger, buckle, propeller twist, opening) have a reasonable effect to the HOMO (and LUMO) base-pair energy values and consequently to the values of Δ and p.

In order to have a better view of the contribution of transfer integrals to the above discussion, in Fig. 4.7, we present the transfer integral t, as a function of the rise, for different twist values, for ideal (with undistorted geometries) AA and GG dimers, obtained by LCAO. Rise (Δz) and twist (ω) are the two most

important conformational parameters in B-DNA. As it is evident from Fig. 4.7, the absolute value of t decreases with increasing rise. We also observe that, e.g., for AA dimers, for twist close to $30°$, $t \approx 0$. Starting from $36°$ and increasing twist towards $48°$, $|t|$ is increased, while, decreasing twist towards $30°$, $|t|$ is decreased and tends to zero. Further, decreasing twist towards $24°$, increases $|t|$ again. Hence, we expect that for twist around $30°$, the hopping parameter will be very small, hence, p should be very small, unless Δ is very small, too (cf. Eq (3.12)).
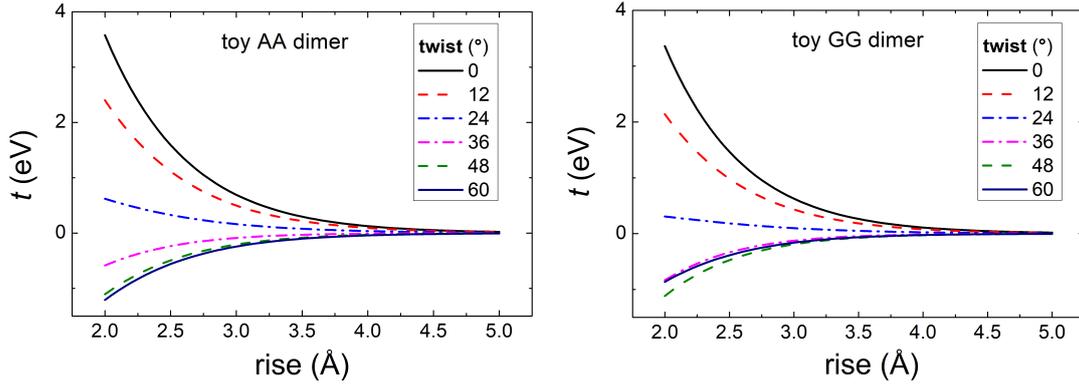
Figure 4.7: Influence of rise (Δz), for various twist (ω) angles, on the hopping parameter, t for an AA (left) and a GG (right) dimer with undistorted geometries, according to LCAO with only $2p_z$ atomic orbitals. For twist $\approx 30°$, $t \approx 0$.

In Fig. 4.8, we depict the values of p, rise (Δz) and twist (ω) for all the cases of the distorted geometries. By observing the fluctuations of these geometry parameters, in a correlation with the p values, a safe deduction cannot be done. The above remark is a definite reminder that charge transfer depends on a large
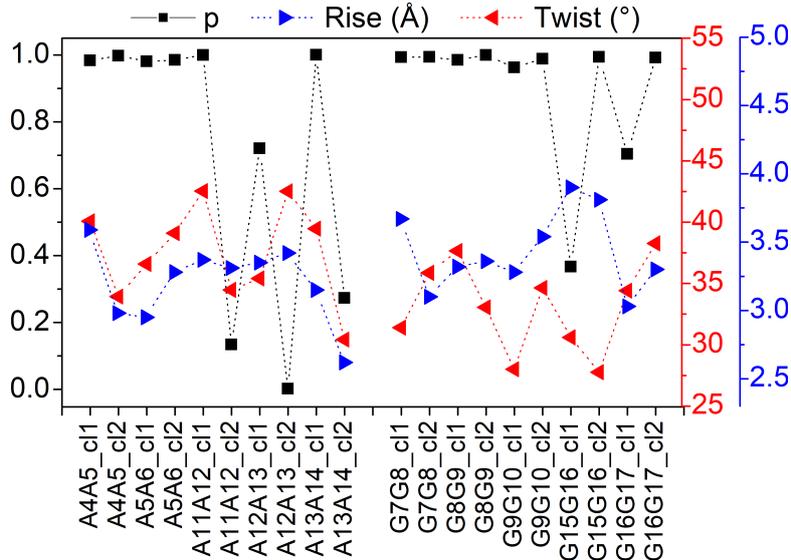
Figure 4.8: p, rise, twist with LCAO using all valence orbitals.

number of structural parameters.

4.2.1 Results obtained by Real-Time Time-Dependent Density Functional Theory

The content of this Subsection can be found in ref. [18]. The content of this Subsection was not produced by the author, so the presentation of the associated results will be limited, in order to have a better view of the effects of structural variability and make a comparative study in Subsection 4.2.2.

Density Functional Theory (DFT) [197--199] is a well established method for treating ground state properties of many electron systems, e.g. molecules or solids. Two decades after its establishment, it was extended to time-dependent DFT (TD-DFT), to treat time-dependent excited state processes. [94, 200--202] The Time-Dependent Kohn-Sham (TDKS) equations with an effective potential energy $v_{\mathrm{KS}}(\mathbf{r}, t)$, uniquely described by the time-dependent charge density, $\rho(\mathbf{r}, t)$, are, in atomic units,

$$i\tfrac{\partial}{\partial t}\Psi_j(\mathbf{r}, t) = \left[-\tfrac{1}{2}\nabla^2 + v_{\mathrm{KS}}(\mathbf{r}, t) \right]\Psi_j(\mathbf{r}, t) = \tag{4.1}$$

$$\left[-\tfrac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\mathbf{r}, t) + v_{\mathrm{H}}(\mathbf{r}, t) + v_{\mathrm{xc}}[\rho](\mathbf{r}, t) \right]\Psi_j(\mathbf{r}, t).$$

$v_{\mathrm{ext}}(\mathbf{r}, t)$ includes external fields and nuclear potentials, $v_{\mathrm{H}}(\mathbf{r}, t)$ is the Hartree potential energy. Exchange and correlation effects are included in $v_{\mathrm{xc}}[\rho](\mathbf{r}, t)$. The charge density is the sum over all occupied orbitals $j = 1, 2, \ldots N_{occ}$, i.e.,

$$\rho(\mathbf{r}, t) = \sum_{j=1}^{N_{occ}} |\Psi_j(\mathbf{r}, t)|^2. \tag{4.2}$$

Real-Time TDDFT (RT-TDDFT) [203, 204] is based on a direct numerical integration of Eq. (4.1). This differs from the traditional linear-response approach, which is not actually a time-resolved method but instead solves Eq. (4.1) in the frequency domain for the excitation energies of a system subject to a small perturbation. Within RT-TDDFT, we solve the TDKS equations and obtain the electron density at each time step. The electron density is then used for the calculation of the Hamiltonian in the next cycle of the self-consistent process. For the DFT and RT-TDDFT calculations, the NWChem open-source computational package [205] was used, as well as a typical time step of 0.5 a.u..

The range-separated functional CAM-B3LYP [206], appropriate for the correct estimation of exchange energy, both at short and long ranges, was used for most DFT and RT-TDDFT results reported in this work. The calculations were performed using the 3-21++G [207, 208],6-31++G** [208--211], aug-cc-pVDZ [212, 213], basis sets, which include diffuse functions, for all dimers. RT-TDDFT with a given basis set and functional started from the DFT density with the same basis set and functional.

In a Gaussian basis set, it is most natural to use the single particle reduced density matrix, whose time evolution is governed by the von Neumann equation. The Magnus propagator is used in NWChem's RT-TDDFT implementation, which is both stable and conserves the density matrix idempotency [204].

In B-DNA, adjacent stacked base pairs are separated by a relatively large distance, so that the stacking interaction is not strong. Thus, the initial state of a stacked dimer, to be evolved by RT-TDDFT, can be determined by DFT calculations of individual base pairs. Here, the procedure adopted for each dimer is the following: First, a DFT ground state calculation was performed for each of the neutral and charged base pair (monomer) appearing in our chosen dimers. The $+1$ charged state simulates the presence of a hole. Then, the resulting eigenstates were combined and subsequently orthogonalized via a Gram-Schmidt process, to obtain the dimer's ground state. RT-TDDFT was finally employed to propagate the electronic density in time from this initial state. At the end of each time step in the excited-state evolution, the fragments' charge was calculated with an appropriate population analysis method, along with the dipole moment. The main frequencies of charge oscillations were extracted from the results via Fourier analysis.

The Löwdin population analysis [214] was used in the RT-TDDFT simulation. The Löwdin population analysis was integrated into the RT-TDDFT module of NWChem for the calculation of each fragment's charge at each time step. It is much less basis-set dependent than the Mulliken analysis (which is the default scheme in NWChem's RT-TDDFT). Furthermore, Löwdin's scheme does not suffer from ultra-fast charge oscillations, which are instead artificially introduced with Mulliken's scheme in RT-TDDFT charge simulations. As a result, Löwdin population analysis gives a more clear picture of charge transfer.

The DFT derived on-site energies of base pairs and transfer integrals between base pairs were calculated following the procedure below: (1) a ground state DFT simulation is performed for each of the (isolated) neutral monomers, as well as for the (neutral) dimer, yielding the corresponding eigenstates and eigenenergies; (2) the eigenstates of monomers are combined and then orthogonalized via the Löwdin's symmetric orthogonalization scheme, to form an orthonormal basis set; (3) the ground state Hamiltonian matrix of the dimer is transformed from the occupied molecular vector subspace to the previously formed basis of monomers's orthogonalized molecular vectors. In this representation, the diagonal Hamiltonian matrix elements that correspond to monomers' HOMOs are the on-site energies, while the relevant non-diagonal element is the hopping integral. In other words, from the Hamiltonian matrix of a dimer we obtain the HOMO on-site energies for each of the two monomers and the transfer integral t between the two monomers. The absolute values of Δ and t, obtained by DFT, are shown in Fig. 4.9.

From the TB parameters obtained by DFT, we can estimate p by Eq. (3.12) and T by Eq. (4.3)

$$f = \frac{1}{T} = \frac{\sqrt{(2t)^2 + \Delta^2}}{h}. \tag{4.3}$$

In Fig. 4.10 (left), we present a comparison between the maximum transfer percentage, for all dimers, as estimated by TB, using Eq. (3.12) with the TB parameters obtained by DFT, and the values obtained graphically by RT-TDDFT. In Fig. 4.10 (right), we present a comparison between the (main) period, for all dimers, as estimated by TB, using Eq. (4.3) with the TB parameters obtained
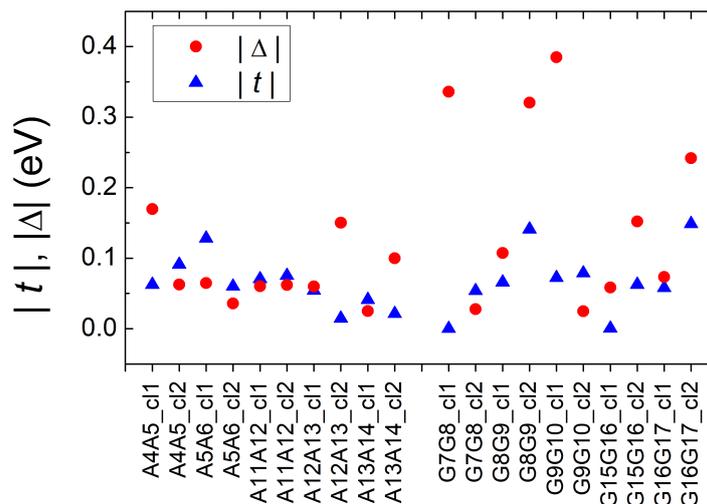
Figure 4.9: The absolute values of the TB parameters Δ (difference between the two monomers' HOMO on-site energies) and t (transfer integral between monomers' HOMOs), obtained by DFT, at the CAM-B3LYP/aug-cc-pVDZ level of theory.

by DFT, and the values obtained graphically by RT-TDDFT.

In Fig. 4.10, we observe that p and T with TB parameters obtained by DFT give the general trends, but the RT-TDDFT results are much richer. We should realize that in the simple TB model, only a constant energy HOMO at each monomer is used and the hole is created at one of the two monomers. In this case, the initial placement of the hole at the one or at the other monomer does not change p or T. In other words, for a specific dimer, p and T are identical irrespective of the initial placement of the hole at monomer 1 (base pair 1, bp1) or at monomer 2 (base pair 2, bp2) [115, 215].
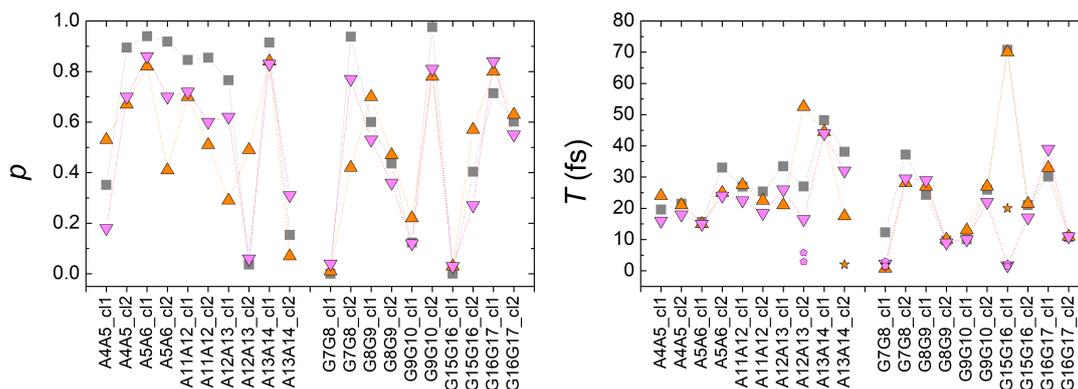
Figure 4.10: Maximum transfer percentage, p (left) and period, T (right) for all dimers. Values estimated by TB, using Eq. (3.12) with TB parameters obtained by DFT for the neutral dimer, are shown with gray squares. Values obtained by RT-TDDFT are shown with orange up and pink down triangles, for initial hole creation at base pair 1 (bp1) and base pair 2 (bp2), respectively. Lines are guides for the eyes. The little stars (case bp1) and pentagons (case bp2) represent some weaker components.

In Fig. 4.11 we present the rise (Δz) and the twist (ω) of our distorted dimers, together with the maximum tranfer percentage, p, as obtained by RT-TDDFT.

In accordance with the above discussion, larger rise leads to smaller p, and vice versa. However, twist is also important, in the sense explained in Fig. 4.7. Twist angles close to 30° correspond to small overlap, hence small transfer integral t, and eventually small p.
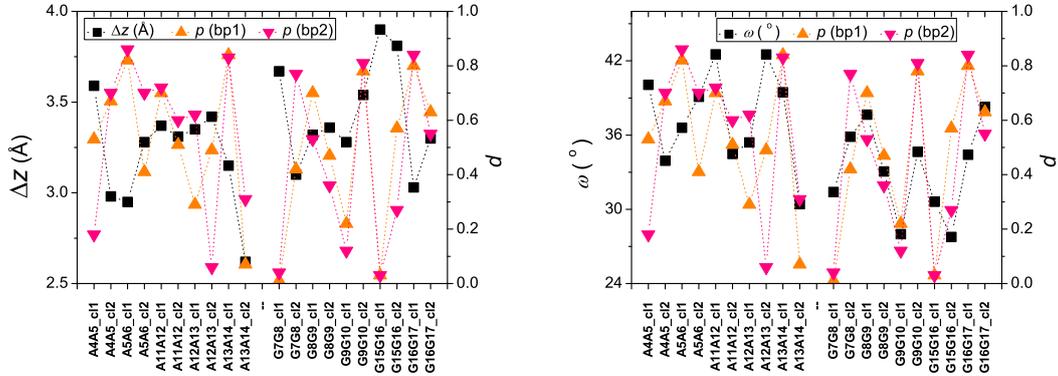
Figure 4.11: The rise (Δz) and twist (ω) of the dimers studied in this work, together with the maximum transfer percentage obtained by RT-TDDFT.

4.2.2   Comparative Study of (RT-TD)DFT and TB results

Even though simple treatments like the TB approach do not prove sufficient for quantitative agreement with RT-TDDFT, they can be useful for a qualitative assessment of our results. The employed TB treatment using all valence orbitals, i.e., $2s$, $2p_x$, $2p_y$, $2p_z$ orbitals for C, N and O atoms and $1s$ orbital for H atoms, has been described in Subsection 2.1.2 and Section 3.1. In Fig. 4.12 we compare the transfer integrals obtained by DFT (cf. Fig. 4.9) with the transfer integrals obtained by LCAO using all valence orbitals, as well as LCAO using only $2p_z$ orbitals, for the MD distorted dimers. Generally, the DFT method and the LCAO method using all valence orbitals give similar trends, while the results obtained by the LCAO method using only $2p_z$ orbitals have slightly larger deviations.

As can be found in Refs. [115, 215], for ideal B-DNA geometries, the period is of the order of 10 to 100 fs. All oscillations between identical monomers, e.g. in dimers like AA and GG, and also AT, TA, GC, and CG are of maximum transfer percentage 1 (100%) and it does not matter whether the carrier is initially placed at the first or at the second monomer, because the two monomers (hence, their on-site energies) are identical (i.e., $\Delta = 0$). On the contrary, all oscillations between different monomers, e.g. in dimers like AC $\equiv$ GT, CA $\equiv$ TG, GA $\equiv$ TC, AG $\equiv$ CT have maximum transfer percentage much smaller than 1, because the two sites are not identical. For DFT and RT-TDDFT applied to structures obtained by MD, even for dimers like AA and GG, i.e., made of "identical" monomers, the sites (monomers) are never exactly identical, primarily due to structural variability. In some cases, the oscillations predicted
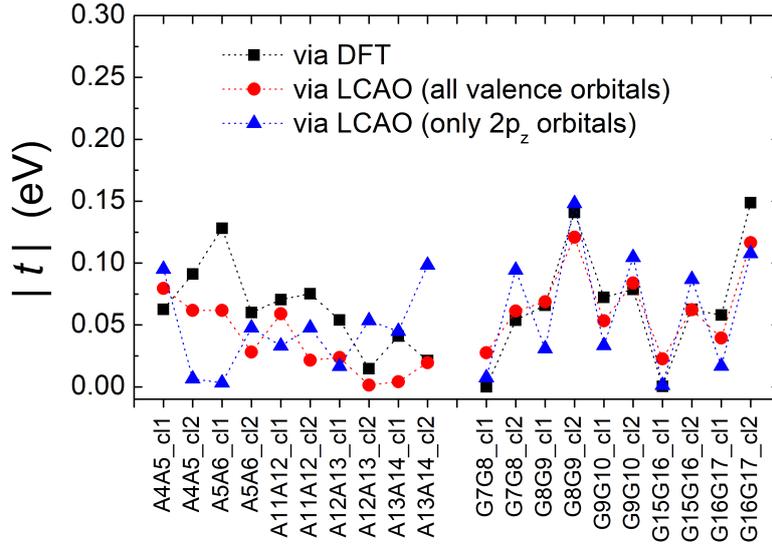
Figure 4.12: Comparison of the transfer integrals obtained by DFT with the ones obtained by LCAO using all valence orbitals.

by MD and RT-TDDFT for placing the hole either at the 1st or at the 2nd monomer are large and almost identical, in accordance with the expected TB picture for oscillations between "identical" monomers. However, in other cases, the oscillations predicted by MD and RT-TDDFT are smaller and/or different for placing the hole initially at the 1st or at the 2nd monomer, seemingly in contrast to the TB picture for oscillations between "identical" monomers. A detailed discussion on this controversy can be found in Ref. [18].

5 | CHARGE TRANSPORT[1]

As already mentioned, the DNA structure allows electron density overlapping of adjacent bases, creating a nearly one-dimensional π-pathway that favours charge transfer and transport, besides stabilising the double helix. The term *transport* implies application of external gradient. Here we use electrodes and apply voltage. The most experimentally relevant quantities for charge transport are the current (I) - voltage (V) curves. Charge transfer and transport through the aromatic base-pair stacking depends on the way stacked bases couple with each other. Therefore, any disturbance in that stacking, through base modifications, insertions, or protein binding, can be observed electrically. In biomedicine, charge transport can be used to detect pathogenic mutations at early stage, e.g., pairing between non-complementary bases leads to point mutations, potentially harmful to the development of organisms (carcinogenesis). Each DNA sequence has a unique electronic signature, which may be useful for identifying a mutant DNA molecule [9, 13, 217]. Thus, charge transfer and transport can bring valuable information about sequencing. It is expected that these properties can be further employed to design electronic circuits as diagnostic tools. Also, charge transport could play a significant role for DNA-repairing deficiency yielding carcinogenesis [16].

Considering the above, we study charge transport along DNA molecules, using the TB method, together with the transfer matrix technique, to solve the time-independent Schrödinger equation and finally obtain I-V curves. We study double-stranded DNA molecules, the ends of which are connected to electrodes, focusing on both ideal and natural geometries concerning two categories of mutations: (i) DNA sequences that contain point substitution mutations, specifically, *transitions* (interchange of purines) G $\leftrightarrow$ A and (ii) sequences extracted from segments of human chromosomes, modified by expanding the CAG triplet to mimic the following diseases: (a) Huntington's disease, (b) Kennedy's disease, (c) Spinocerebellar ataxia 6, (d) Spinocerebellar ataxia 7. Physical quantities such as eigenspectra, density of states, transmission coefficients and current-voltage curves are obtained. The parameters used to describe the molecular electronic structure of nucleic acid bases and extract the on-site energies and the interaction integrals used in the recruited TB wire model were obtained from the LCAO method, considering the molecular wave function as a linear combination of all valence orbitals of all atoms, i.e., 2s, $2p_x$, $2p_y$, $2p_z$ orbitals for C, N, and O atoms and 1s orbital for H atoms.

The novel features of this work compared to state of the art include the following: 1) Ideal and natural DNA geometries are compared. 2) Known mu-

---

tations are examined, and mutated sequences of the above mentioned categories are compared with unmutated ones. 3) The potential use of physical quantities related to charge transport as mutation detectors is investigated. 4) The definition of the normalised deviation of the I-V curve from the origin, NDIV, seems to be a useful quantity for that purpose.

The rest of this Chapter is organized as follows: In Sec. 5.1 we briefly lay down the Transfer Matrix Method we use, in Sec. 5.2 we list the studied sequences and genetic disorders, and in Sec. 5.3 we present our results for various physical quantities and discuss them.

5.1 The Transfer-Matrix Method

In this section, we - once again - employ the wire model variation of the TB method. As mentioned in the previous chapters of this Ph.D. Thesis, the parameters required for the wire model description are the on-site energies of the base pairs and the interaction integrals between successive base pairs. In order to produce the required on-site energies, we employed the LCAO method, considering the molecular wave function as a linear combination of all valence orbitals, i.e., $2s$, $2p_x$, $2p_y$, $2p_z$ orbitals for C, N, and O atoms and $1s$ orbital for H atoms. We used the novel parameterization described in Subsection 2.1.2, initially introduced in Ref. [27]. As for the interaction integrals, we employed the Slater-Koster two-centre interaction form [64] using Harrison-type expressions [65, 66] as described in Subsection 3.1.1 [27].

As already described in Subsection 3.1.1, the problem we aim to solve - i.e. the time independent Schrödinger equation - is reduced to a system of coupled algebraic equations of the form of Eq. (3.7). Once again, strand $5'$-$3'$ is used to denote the DNA segments. Therefore, the notation GG implies two base pairs of GG bases in the $5'$-$3'$ strand and their complementary ones, CC, in the $3'$-$5'$ strand. Equation (3.7) can be written in the matrix form

$$\begin{pmatrix} A_{\lambda+1} \\ A_{\lambda} \end{pmatrix} = \begin{pmatrix} \frac{E-E_{\lambda}}{t_{\lambda}} & -\frac{t_{\lambda-1}}{t_{\lambda}} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} A_n \\ A_{n-1} \end{pmatrix} = P_{\lambda}(E) \begin{pmatrix} A_{\lambda} \\ A_{\lambda-1} \end{pmatrix}, \qquad (5.1)$$

and be solved using the transfer matrix method. $P_{\lambda}(E)$ is called the transfer matrix of monomer λ. Hence, the product

$$M_N(E) = \prod_{\lambda=N}^{1} P_{\lambda}(E). \qquad (5.2)$$

is the global transfer matrix (GTM) of the sequence, and contains all the information about its energetics.

Charge transport properties are studied under the assumption that the DNA sequences of interest lie on sites $n = 1, \ldots, N$ and are connected with two semi-infinite homogeneous metallic electrodes (leads), acting as carrier baths, which lie on sites $(-\infty, 0] \cup [N+1, +\infty)$. The situation is depicted in Fig. 5.1. The leads
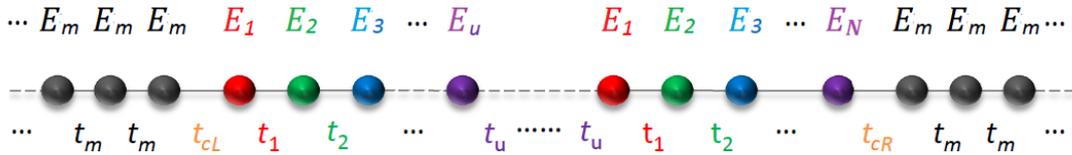
Figure 5.1: Schematic representation of a WM with a unit cell of u sites, sandwiched between two semi-infinite homogeneous metallic leads, and connected to them from the left (right) side with hopping integral $t_{\mathrm{cL}}(t_{\mathrm{cR}})$.

are described by properly chosen on-site energies and interaction parameters.

Here, left and right electrodes are assumed to be identical. The electrodes energy spectrum is given by the *dispersion relation* [218]

$$E = E_\mathrm{m} + 2t_\mathrm{m}\cos(qa), \tag{5.3}$$

where E_m is the on-site energy of the electrodes, t_m is the interaction integral between the electrodes sites, and a is the lattice constant. The electrodes' band lies in the energy interval $[E_\mathrm{m} - 2|t_\mathrm{m}|, E_\mathrm{m} + 2|t_\mathrm{m}|]$. Hence, the energy center and bandwidth of the electrodes are E_m and $4|t_\mathrm{m}|$, respectively. If we imagine the lead as a homogeneous system with one electron per site, then the band is half-filled, the electrodes are metallic, and the Fermi level of the electrodes, E_m^F, is identified with the on-site energy of the electrodes, E_m. The coupling of the sequence with left (L) and right (R) edge electrode sites is described by the interaction integrals $t_\mathrm{cL(R)}$, respectively. The choice of the right parameters is important because it defines the optimum transport profile. For periodic sequences, the *coupling strength factor* is already defined [219] as

$$\omega = \frac{t_\mathrm{m} t_N}{t_\mathrm{cR} t_\mathrm{cL}}, \tag{5.4}$$

in means of the deviation of the real coupling of the system to the leads from the ideal coupling [in which the product of the coupling hopping integrals (t_cL, t_cR) becomes equal to the product of the hopping integrals of the isolated lead ($t_m = t_\mathrm{cL}$) and the isolated system ($t_N = t_\mathrm{cR}$) as if each was cyclically bounded to itself]. The coupling asymmetry factor $\chi = \frac{t_\mathrm{cL}}{t_\mathrm{cR}}$ expresses the difference in coupling strength between the leads and the left/right end of the system. The ideal coupling condition, which is definable only in periodic cases, is $|\omega| = 1$. The symmetric coupling condition is meant as $|\chi| = 1$. In periodic cases, the ideal and symmetric coupling condition, $\omega = 1 = \chi$, leads to the most enhanced transmission [219].

The ***transmission coefficient*** at zero bias, $T(E)$, is a useful quantity for the description of the charge transport properties and refers to the probability that a carrier transmits through the sequence eigenstates. To compute $T(E)$, a transfer matrix formalism [218--220] is used. After some manipulations, we get that $T(E)$ can be analytically expressed as [218]

$$T(E) =$$

$$\frac{4\sin^2(qa)}{\mathrm{Tr}\left(\tilde{M}_N\right)^2 \sin^2(qa) + \left[\tilde{M}_N^{(12)} - \tilde{M}_N^{(21)} + (\tilde{M}_N^{(11)} - \tilde{M}_N^{(22)})\cos(qa)\right]^2}, \tag{5.5}$$

where

$$\tilde{M}_N = P_R M_N P_L, \ P_R = \begin{pmatrix} 1 & 0 \\ 0 & \frac{t_\mathrm{cR}}{t_m} \end{pmatrix}, P_L = \begin{pmatrix} \frac{t_m}{t_\mathrm{cL}} & 0 \\ 0 & 1 \end{pmatrix}, \tag{5.6}$$

Tr denotes the matrix trace and $M_N^{(ij)}$ are the elements of the GTM.

The *I* - *V* curves of the DNA sequences have been calculated using the Landauer - Büttiker formalism [221--223]. A constant bias voltage was applied, V_b,

between the leads, which induces a linear potential energy drop, $U_b = -eV_b$, from one to the other edge of the DNA sequence. Thus, the transmission coefficient becomes bias-dependent. The leads' chemical potential takes the form

$$\mu_{L(R)} = E_m \pm \frac{U_b}{2}. \tag{5.7}$$

The energy regime between them defines the conductance channel. At zero temperature, the Fermi-Dirac distributions are Heaviside step functions, so the electrical current can be computed as

$$I(V) = \frac{2e}{h} \int_{\mu_R}^{\mu_L} dE\, T(E, U_b). \tag{5.8}$$

In Fig. 5.2, the absolute value of the current in logarithmic scale, $\log_{10}|I|$, is demonstrated as a function of both the leads on-site energy, E_m, and the applied voltage between the leads, V, for ideal (left) and natural (right) G_{14} polymers. It is evident that the electrode's on-site energy plays a crucial role in the shape and magnitude of the current–voltage curves. A general trend for homopolymers is that larger currents occur when E_m is closer to the monomer's on-site energy.
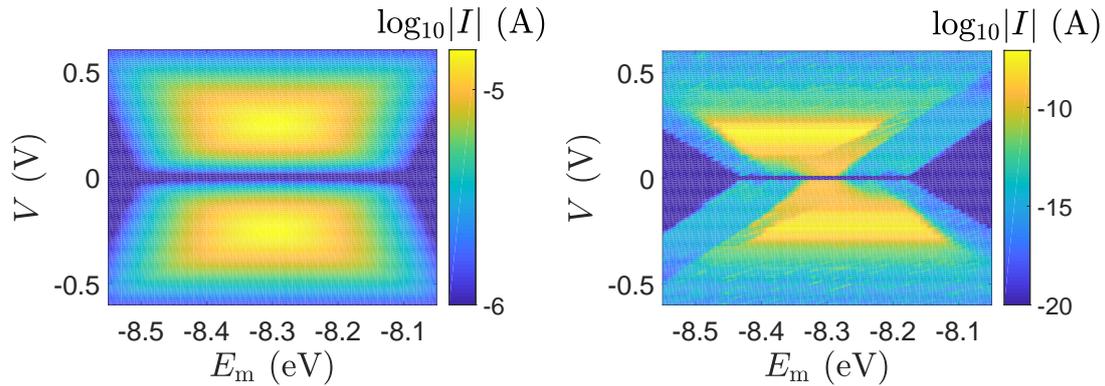
Figure 5.2: Colour maps showing $\log_{10}|I|$, i.e., the absolute value of the current in logarithmic (colour) scale, as function of both the leads on-site energy, E_m, and the applied voltage between the leads, V, for both ideal (left) and natural (right) G_{14} polymers.

5.2 Studied Sequences and Genetic Disorders

Genetic disorders occur when a mutation affects an organism's genes or when the organism has the wrong amount of genetic material. Carrying the mutation does not necessarily mean that the organism will end up with a disease. There are many types of disorder, including single-gene, multifactorial and chromosomal ones. In the present Ph.D. Thesis, we focus on single-gene genetic disorders, i.e., changes or mutations that occur in the DNA sequence of a single gene, otherwise known as monogenetic disorders.

Point substitutional mutations are common. The G-T mismatch mutation alone occurs about once in every 10^4 - 10^5 base pairs. Maintaining a low mutation rate is crucial for cell viability and health. The high fidelity of DNA replication is established and secured by an enzyme, the replicative polymerase, though several mechanisms: (1) sensing proper geometry of the correct base pair, (2) slowing down catalysis in case of a mismatch, and (3) partitioning the mismatched primer to exonuclease active site [224]. However, the performance of polymerases is not error-free. It is estimated [224--226] that even after proofreading, the overall fidelity of DNA synthesis lays in the range of one wrong nucleotide incorporated per 10^3 - 10^5. Besides, DNA replication is constantly challenged by internal and external factors, non-canonical DNA structures, and complex DNA sequences [224].

Another category of DNA mutations related to several diseases is the ***short tandem repeat (STR) expansions***, also known as microsatellites [227, 228]. These are small sections of DNA, usually 2-6 nucleotides long, repeated consecutively at a defined region. At least 6.77% of the human genome is comprised of these repetitive DNA sequences [228]. Large STR expansions are potentially pathogenic, setting the ground for several neurological diseases. In fact, 37 of the already known STR genes that can cause disease when expanded, exhibit primary neurological presentations [228]. In neurological STR diseases, 'CAG' repeat expansions code for the amino acid glutamine. When expanded, they create polyglutamine tract expansions, which are thought to alter and expand the transcribed protein, creating insoluble protein aggregates within neuronal cells. This can cause perturbations in intracellular homeostasis and cell death [229].

Here, we examine two categories of DNA polymers: (i) sequences that contain point substitution mutations, specifically, *transitions* involving G $\leftrightarrow$ A exchange, of both ideal and natural geometries, replacing the out-of-the-rings atoms that are different between A and G, while ensuring that the number of hydrogen bonds is correct, and (ii) sequences of ideal geometry extracted from segments of human chromosomes, subsequently modified by expansion of the CAG triplet [$(CAG)_n$ repeats] to mimic four selected STR diseases, namely, (a) Huntington's disease, (b) Kennedy's disease, (c) Spinocerebellar ataxia 6, (d) Spinocerebellar ataxia 7. The number of pathogenic repeats, i.e., CAG triplets, in Huntington's disease, is n_p = 36-250, located in exon 1 of HTT gene, chromosome 4 [228, 230, 231]. In spinal and bulbar muscular atrophy of Kennedy (Kennedy's disease),

n_p = 38-68, located in exon 1 of AR gene, chromosome X [228, 232--234]. In spinocerebellar ataxia 6, n_p = 19-33, located in exon 47 of CACNA1A gene, chromosome 19 [228, 235, 236]. In spinocerebellar ataxia 7, n_p = 34-460, located in exon 1 of ATXN7 gene, chromosome 3 [228, 237, 238].

5.3 Results and Discussion

We devote Subsection 5.3.1 to eigenenergies and densities of states, Subsection 5.3.2 to transmissions and Subsection 5.3.3 to current (I) - voltage (V) curves, where we introduce the normalised deviation of the I-V curve from the origin (NDIV). In this work, we focus on charge transport through HOMOs.

For DNA segments of ideal geometry, the base pairs are not distorted and base pairs are separated and twisted by 3.4 Å and 36°, respectively, relative to the double helix growth axis. The geometries of the *natural* sequences G_{14} and A_{15} have been extracted from Bioinformatics (RCSB) Protein Data Bank (www.rcsb.org) [accession numbers 4WZW and 6VAA, respectively], from the original Refs. [239] and [240], respectively.

The on-site energies and interaction integrals for all sequences were calculated using all valence orbitals of all atoms, according to the procedure described in Ref. [27]. For ideal sequences, the on-site energies are $E_{\text{A-T}} = -8.49$ eV for the A-T base pair and $E_{\text{G-C}} = -8.30$ eV for the G-C base pair [27], as listed in column 2, Table 2.4. We also calculated the on-site energy of the mismatched A-C base pair. Its HOMO value is -8.43 eV, i.e., very close to that of the A-T base pair, and very close to that of the A base, as expected, since A has higher HOMO than that of C. Its LUMO value is -4.43 eV (σ^*), -4.23 eV, very close to that of the C base, as expected, since C has a lower LUMO than that of A. The HOMO (LUMO) interaction integrals between successive base pairs of ideal geometry, without mismatches, calculated with the method described in [27] using all valence orbitals of all atoms, can be found in Table 3.3, column 2(5). Mutations and distortions change the values of interaction integrals; this effect is included in our work, via the same method [27]. Table 5.1 contains the absolute values of interaction parameters between HOMOs (LUMOs), obtained from LCAO, using all valence orbitals, for close-to-ideal geometrical conformations [27], for dimers containing A-C mismatched monomers (denoted by Am) and G-C monomers, for all possible combinations of successive base pairs. Studying charge transport

Table 5.1: Absolute values of interaction parameters between HOMOs (LUMOs), $|t_{\text{LCAO}}|$, obtained from LCAO using all valence orbitals, for close-to-ideal geometrical conformations [27], for dimers containing A-C mismatched monomers (denoted by Am) and G-C monomers, for all possible combinations of successive base pairs. All values are given in meV. XY denotes the sequence in the 5′-3′ direction.

XY	$|t_{\text{H}}|$ [27]	$|t_{\text{L}}|$ [27]
GAm	130	$89(\sigma^*)$ 8
AmG	31	$90(\sigma^*)$ 20
AmAm	36	$90(\sigma^*)$ 25

through HOMOs within the TB wire model (as we do in this work), it is practically easier to examine mutations concerning purine substitution by another

purine; given that purines have higher HOMO than pyrimidines, this substitution will have substantial effect on the base pair on-site energy. This substitution will generate important diagonal disorder within the TB wire model Hamiltonian matrix, in addition to the always-present off-diagonal disorder caused by the modification of interaction parameters.

For completeness, in Appendix D we display the pure mean transfer rates k (see Subsection 3.1.3, Eq. (3.39)) of the studied G_{14} sequences: a) with one A - C mismatch mutation of varying position (sites 1, 2, 3) in the sequence for both ideal and natural polymers (Fig. 6), b) with zero, one, and two randomly positioned A - C mismatch mutations for both ideal and natural polymers (Fig. 7). The general trend for ideal polymers is that any inserted mutation decreases k drastically. In the case of natural polymers, not a safe deduction can be done, because the off-diagonal disorder inserted by the interaction parameters is random.

5.3.1   Eigenspectra and Density Of States

The eigenspectra of the studied sequences were calculated by numerical diagonalization of the Hamiltonian matrices, which are real, tridiagonal and symmetric matrices, within the TB wire model [106, 116]. In order to describe the electronic structure of the systems under examination we also calculate DOS, which shows the number (N_E) of states that can be occupied by electrons per energy (E) interval, or $\frac{\mathrm{d}N_E}{\mathrm{d}E}$ to be more precise, as defined by Eq. (3.21).

The *integrated density of states (IDOS)* refers to the number of states that have energy smaller than E, and is defined as

$$\mathrm{IDOS}(E) = \int_{-\infty}^{E} g(E')\mathrm{d}(E'). \qquad (5.9)$$

A. Unmutated Sequences
As examples, we show in Fig. 5.3 the eigenspectra and corresponding DOS for a few ideal and natural DNA homopolymers (*or* homo-oligomers), G... and A..., without mutations. It is clear that the eigenstates of ideal G... and A... homopolymers are symmetrically positioned around the on-site energy of the ideal G-C or A-T base pair, respectively. Regarding natural sequences, although the eigenspectra are still close to the corresponding ideal G-C or A-T base pair on-site energy, they are no longer symmetrically positioned, due to the presence of diagonal and off-diagonal disorder. The corresponding normalised IDOS can be found in Fig. 5.4.
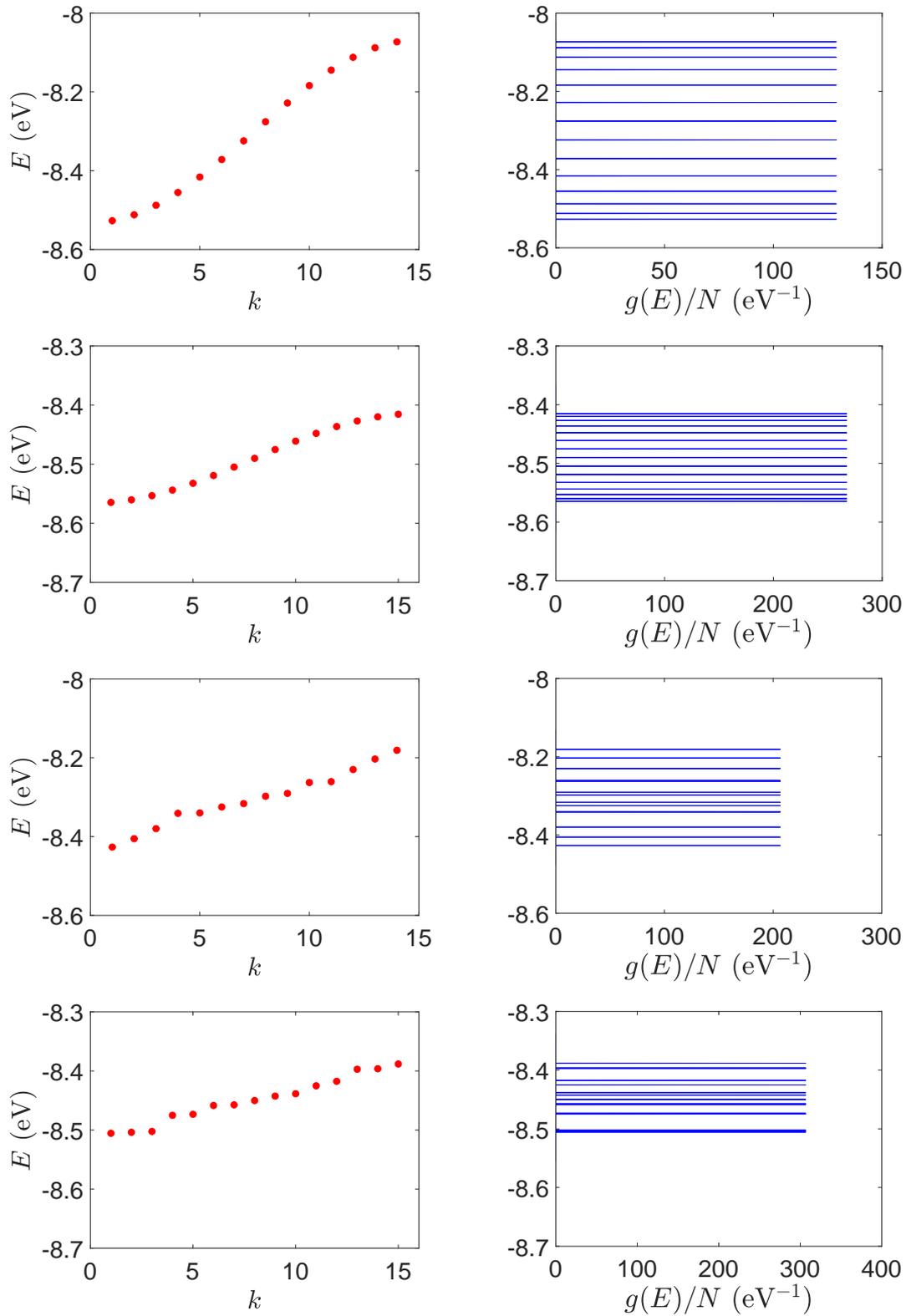
Figure 5.3: Eigenspectra (left) and DOS(right) of unmutated DNA homopolymers. First row: ideal G_{14}, second row: ideal A_{15}, third row: natural G_{14}, fourth row: natural A_{15}. The geometries of the *natural* sequences G_{14} and A_{15} have been extracted from Bioinformatics (RCSB) Protein Data Bank (www.rcsb.org) [accession numbers 4WZW and 6VAA, respectively] from the original Refs. [239] and [240], respectively. k is the eigenenergy index.
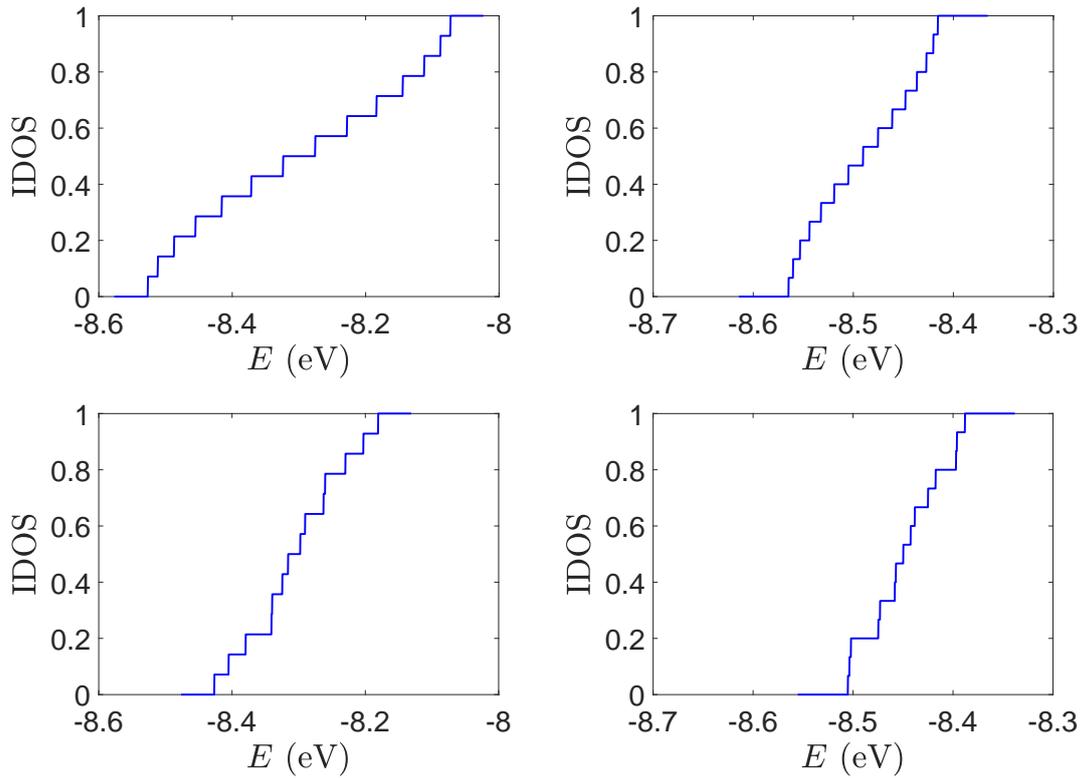
Figure 5.4: Normalized IDOS of the unmutated DNA sequences depicted in Fig. 5.3. Upper panels: ideal, lower panels: natural, left panels: G_{14}, right panels: A_{15}.

B. Mutated sequences
i) Point substitution mutations.

The transition $G \leftrightarrow A$ occurs by introducing A instead of G in G... polymers: the pyrimidine strand still contains only cytosines, but in the purine strand we replace guanines with adenines. Hence, the replaced base pairs are A-C instead of G-C. When such mismatches are introduced, the respective interaction integrals are modified, according the procedure described above, i.e., using the input geometry and LCAO with all valence orbitals of all atoms [27]. The eigenspectra and the corresponding DOS for the studied ideal and natural G_{14} sequences with 7 A-C mismatch mutations, randomly positioned in the sequence, are presented in Fig. 5.5. The corresponding normalised IDOS can be found in Fig. 5.6. In other words, the purine strand contains 7 G and 7 A randomly distributed, while, the pyrimidine strand still contains 14 C. Comparing Fig. 5.3 with Fig. 5.5, we observe that apart from the increased irregularity of the mutated sequences, there is roughly a movement of the mean value of eigenenergies from around $E_{G\text{-}C}$ towards around $E_{A\text{-}C}$ (cf. also Table 2.4). It can be observed that the number of levels close to $E_{A\text{-}C}$ is increased. In particular, the natural mutated sequence displays a high density of levels closer to both $E_{G\text{-}C}$ and $E_{A\text{-}C}$. This can also be understood by inspecting the IDOS, i.e., by comparing Fig. 5.4 with Fig. 5.6.
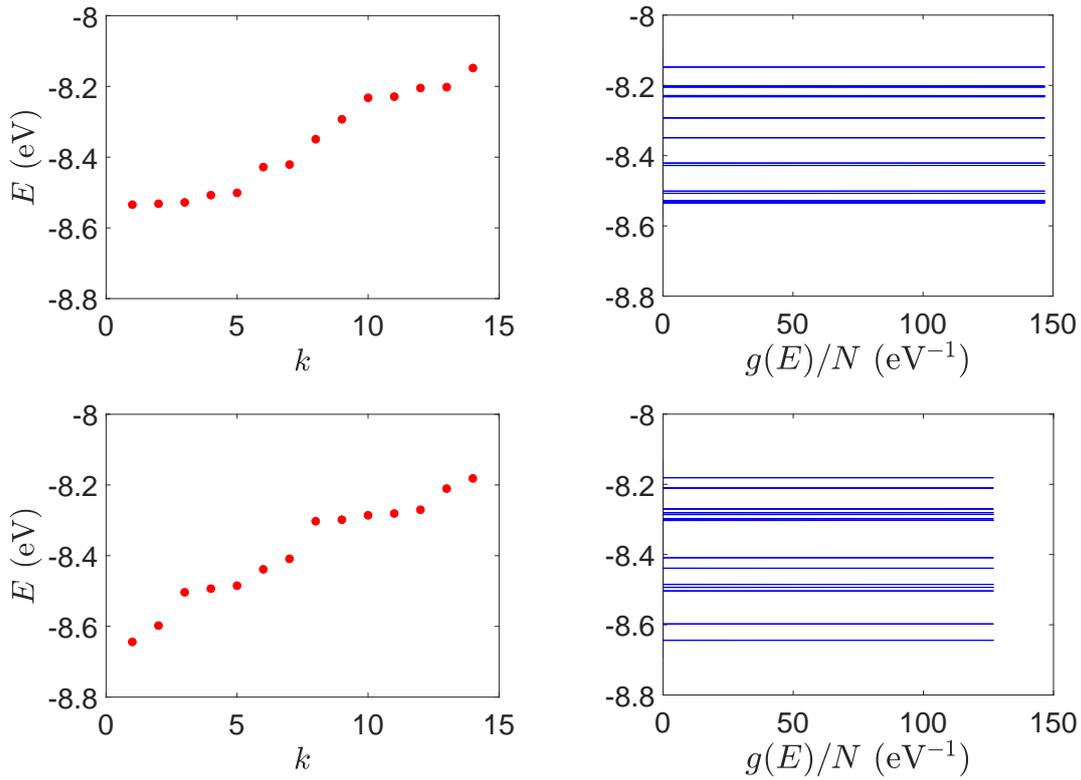
Figure 5.5: Eigenspectra and DOS of initially G_{14} sequences, but with 7 A-C mismatch mutations, randomly positioned in the sequence. The purine strand contains 7 G and 7 A randomly distributed, while, the pyrimidine strand still contains 14 C. Upper: ideal polymers, lower: natural polymers. This figure should be compared with Fig. 5.3.
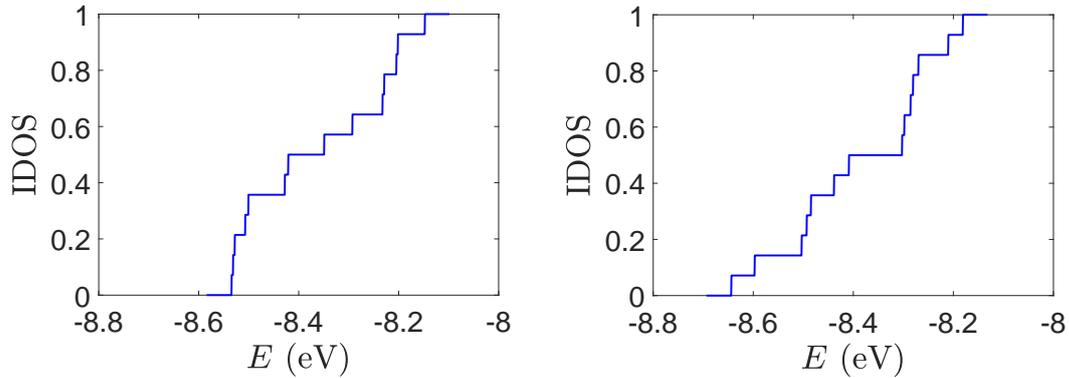
Figure 5.6: Normalized IDOS of initially G_{14} sequences, but with 7 A-C mismatch mutations, randomly positioned in the sequence. The purine strand contains 7 G and 7 A randomly distributed, while, the pyrimidine strand still contains 14 C. Left: ideal polymers, right: natural polymers. This figure should be compared with the left part of Fig. 5.4.

ii) Short tandem repeat (STR) expansions.

We examine four important cases of STR expansions: (a) Huntington's disease, (b) Kennedy's disease, (c) Spinocerebellar ataxia 6, (d) Spinocerebel-

lar ataxia 7. The complete sequences, including primers [241], used in these examples are (a) Huntington's disease: AAGTCCTTC(CAG)$_{100}$CAACA-GCCG, (b) Kennedy's disease: CTGCTGCTG(CAG)$_{45}$CAAGAGACT, (c) Spinocerebellar ataxia 6: GGGCCCCCG(CAG)$_{30}$GCGGTGGCC, (d) Spinocerebellar ataxia 7: GCCGCCCGG(CAG)$_{100}$CCGCCGCCT. The eigenspectra and corresponding DOS for the studied sequences with STR expansion mutations are presented in Fig. 5.8. These show two subbands around the on-site energies $E_{\text{G-C}} = -8.30$ eV and $E_{\text{A-T}} = -8.49$ eV, plus scattered features due to the presence of primers. The corresponding normalised IDOS can be found in Fig. 5.7.
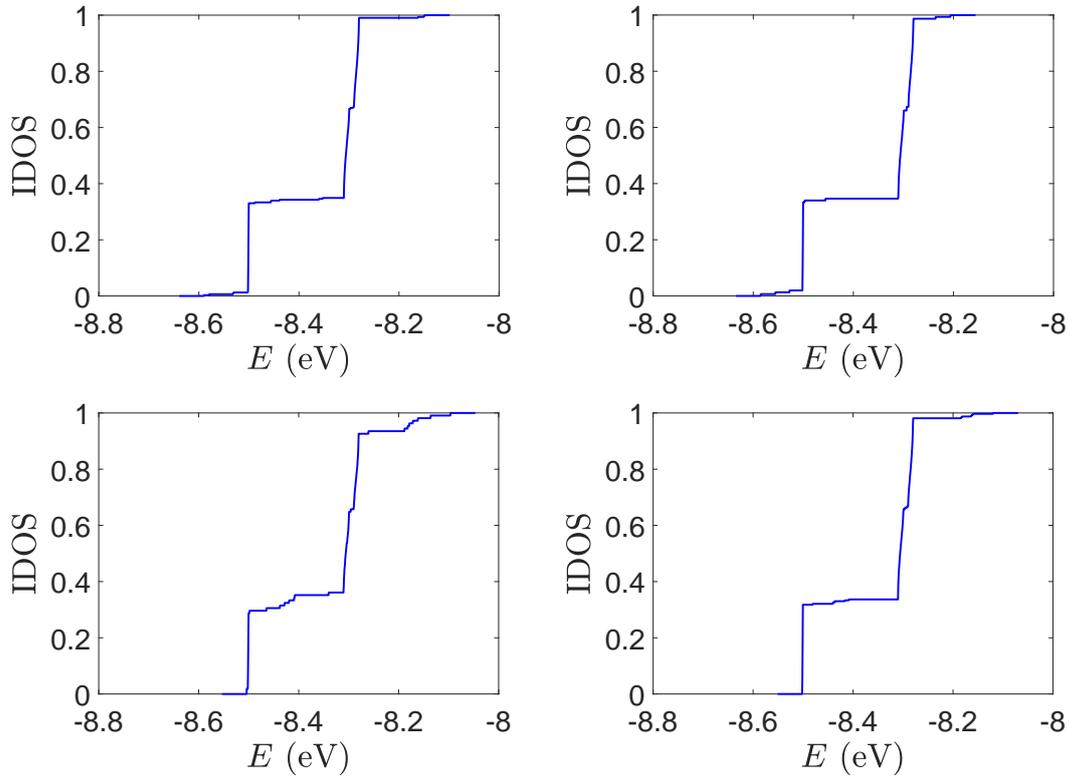
Figure 5.7: IDOS of the studied DNA sequences with STR expansion mutations. (a) Upper left: Huntington's disease with 100 STR expansions, (b) upper right: Kennedy's disease with 45 STR expansions, (c) left lower: Spinocerebellar ataxia 6 with 30 STR expansions, (d) lower right: Spinocerebellar ataxia 7 with 100 STR expansions.
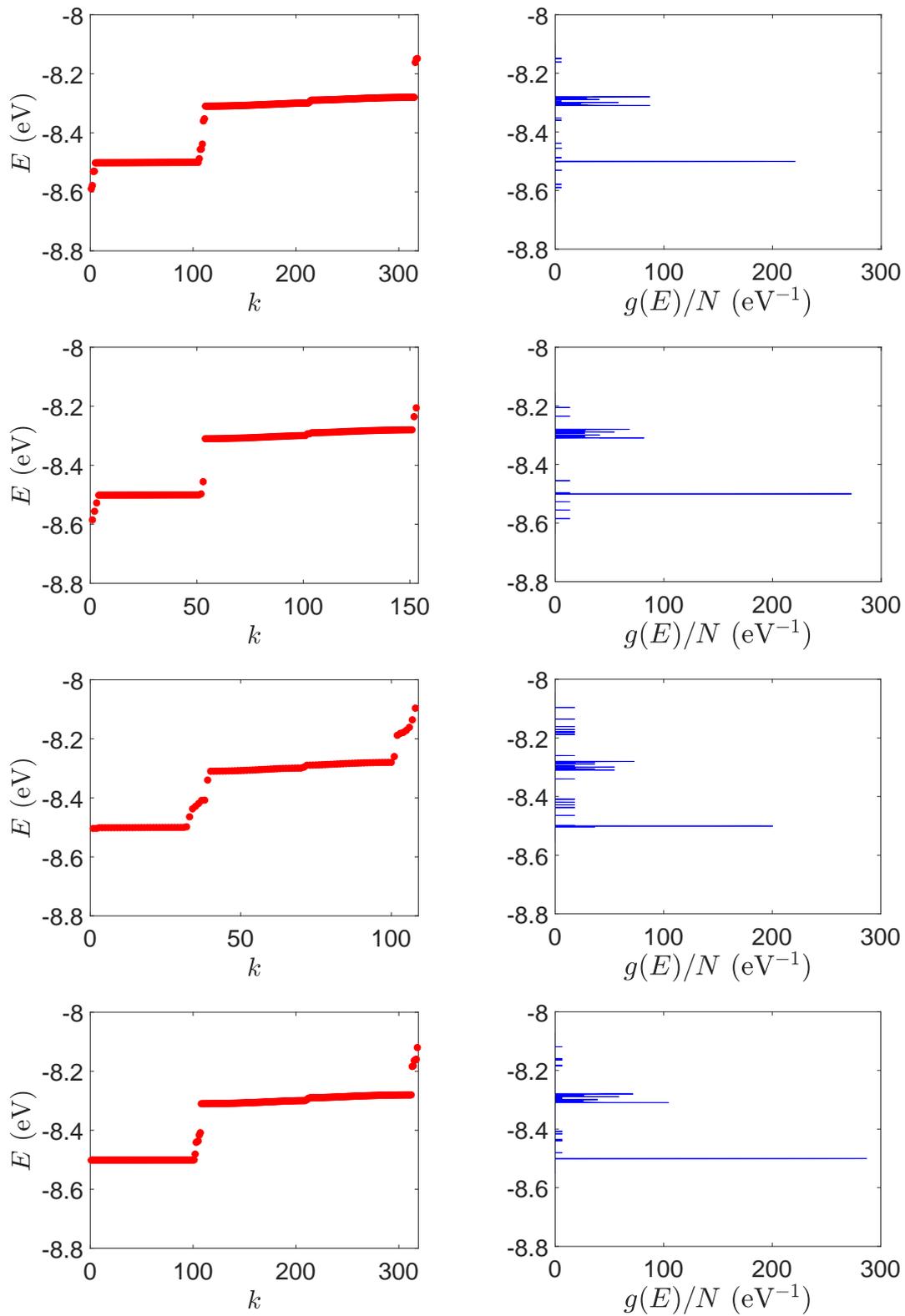
Figure 5.8: Eigenspectra and DOS of the studied DNA sequences with STR expansion mutations. (a) first line: Huntington's disease with 100 STR expansions, (b) second line: Kennedy's disease with 45 STR expansions, (c) third line: Spinocerebellar ataxia 6 with 30 STR expansions, (d) fourth line: Spinocerebellar ataxia 7 with 100 STR expansions.

5.3.2   Transmission coefficient

For E_m we assume that it is either equal to E_{G-C} or E_{A-T}: for ideal or natural G... we take $E_m = E_{G-C}$; for ideal or natural A... we take $E_m = E_{A-T}$; for G... sequences containing A-C mutations we use $E_m = E_{G-C}$; for diseases we still use $E_m = E_{G-C}$. Here, we take $|t_m| = 0.5$ eV, so that the leads' bandwidth contains all the eigenstates of our sequences, cf. Figs. 5.3, 5.5, 5.8.

We choose t_{cL} and t_{cR} from the ideal and symmetric coupling conditions of periodic cases of ideal homopolymers G... and A..., i.e., when dealing with G... or A... we take t_N equal either to $t_{GG} = 0.116$ eV or to $t_{AA} = 0.038$ eV, according to our TB parametrization [27], employing Eq. (5.4). This procedure results in $t_{cL} = t_{cR} = 0.24$ eV for G... and 0.14 eV for A.... In natural homopolymers G..., we still use $t_{cL} = t_{cR} = 0.24$ eV. In natural homopolymers A..., we still use $t_{cL} = t_{cR} = 0.14$ eV. For A-C mismatches in G... as well as for diseases, we still use $t_{cL} = t_{cR} = 0.24$ eV.

$T(E)$ for the studied ideal and natural DNA sequences, G_{14} and A_{15}, without mutations, are presented in Fig. 5.9. In ideal periodic segments, it is expected
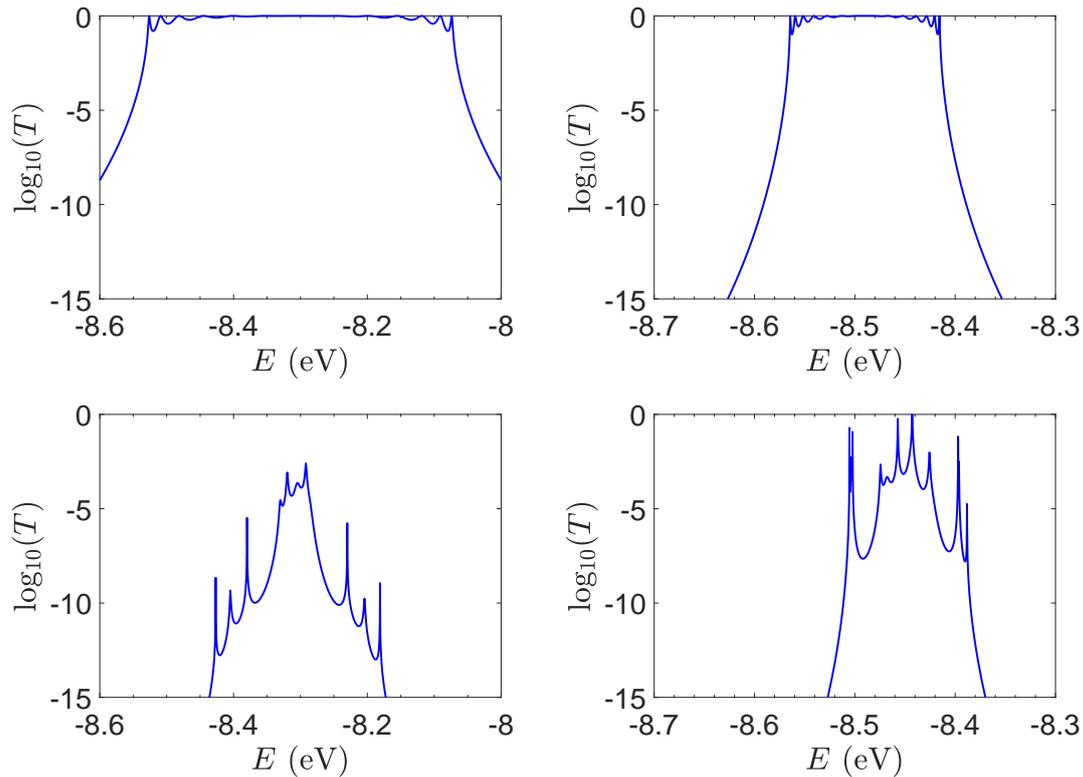
Figure 5.9: The $\log_{10}(T(E))$ of the studied DNA sequences, without mutations. Upper panels: ideal polymers. There are $N-1$ peaks with full transmission (left) and N peaks with full transmission (right). Theory[219] guarantees *at least* $N-1$ peaks. Lower panels: natural polymers. Left: G_{14}, right: A_{15}.

from theory[219] that full transmission $(T(E) = 1)$ occurs at specific energies, at least $N-1$ in number, which is actually the case in the upper panels of Fig. 5.9 (not all peaks are seen clearly at this scale). The natural sequences have a signif-

icantly less symmetric profile, and significantly reduced overall transmission, as expected, because neither the on-site energies nor the interaction integrals are any more equal, i.e., in natural homopolymers both diagonal and off-diagonal disorder are present.

In Fig. 5.10, we depict the on-site energies (left) and absolute values of the interaction integrals (right), together with their mean values, μ, and standard deviations, σ, of the natural G_{14} and A_{15} sequences whose transmission is shown in the lower panels of Fig. 5.9. The corresponding values of ideal sequences are also shown, for reference. The mean values and standard deviations (μ, σ) of the on-site energies, which account for diagonal disorder, are $\approx (-8.304$ eV, 0.005 eV) for G_{14} and $(-8.449$ eV, 0.004 eV) for A_{15}, while, those of the magnitude of the interaction integrals, which account for off-diagonal disorder, are $(0.040$ eV, 0.034 eV) for G_{14} $(0.024$ eV, 0.014 eV) for A_{15}. In terms of coefficients of variation $CV = \frac{|\mu|}{\sigma}$, diagonal disorder is small and of comparable magnitude between G_{14} and A_{15}, i.e., $\approx 0.06\%$ and 0.05%, respectively. On the other hand, off-diagonal disorder is much larger, i.e., $\approx 85.00\%$ and 58.33%, respectively. Clearly, off-diagonal disorder is more pronounced in G_{14}. This explains qualitatively the smaller transmission peaks G_{14} displays compared to A_{15} (cf., bottom panels of Fig. 5.9). Notice that $|t_n|$ was used to assess the off-diagonal disorder, since the spectrum of tridiagonal, irreducible, real, symmetric matrices (as all our matrices are, within the wire model) does not depend on the signs of their off-diagonal entries [96].
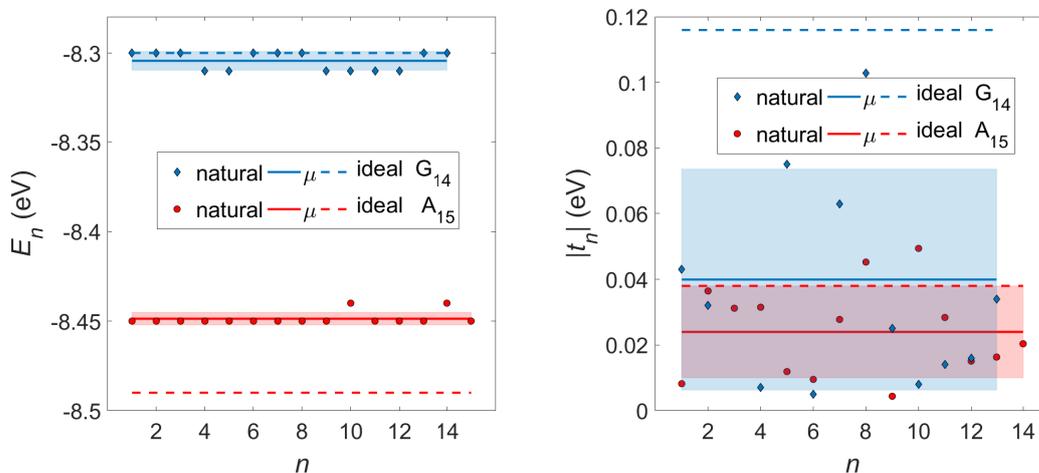
Figure 5.10: TB parameters for the natural G_{14} and A_{15} polymers whose transmission is shown in the lower panels of Fig. 5.9. Left: On-site energies, E_n. Right: Absolute values of interaction parameters, $|t_n|$. Blue (G_{14}) diamonds and red (A_{15}) circles represent the values of the parameters at each site, continuous lines their mean values, μ, and shaded areas include the region $\mu \pm \sigma$, where σ is the standard deviation. The values of parameters for ideal polymers are shown in dashed lines, for reference.

In Figure 5.11 the effect of including zero, one, and two A–C mismatches, randomly distributed in the sequence, is shown, for ideal and natural G_{14} segments. Transmission coefficients, $T(E)$, are also displayed in log-scale (lower

panels). The randomly positioned mismatches are at the same sites for ideal and natural sequences. $\log_{10}(T(E))$ for zero A-C mismatch mutations is also displayed in the left panels of Fig. 5.9. The values of $\int_{-\infty}^{+\infty} dE\, T(E)$ (which act as a measure of the overall transmission) for the three ideal cases are: 0.3856 eV, 0.0430 eV, and 0.0237 eV for 0, 1, and 2 (A-C) mismatches, respectively. Hence, in the ideal cases inclusion of more mismatches decreases transmission, because the sequence homogeneity in terms of on-site energies and interaction integrals is deteriorated. For the three natural cases, the values of the integrals are 1.0651 $\times 10^{-5}$ eV, 3.0247 $\times 10^{-4}$ eV, and 2.2113 $\times 10^{-4}$ eV, respectively. In the natural sequence with zero mismatches there is no homogeneity to be lost by inserting (A-C) mismatches: the sequences are already disordered. Therefore it is difficult to expect to characterize natural sequences based only upon $T(E)$.
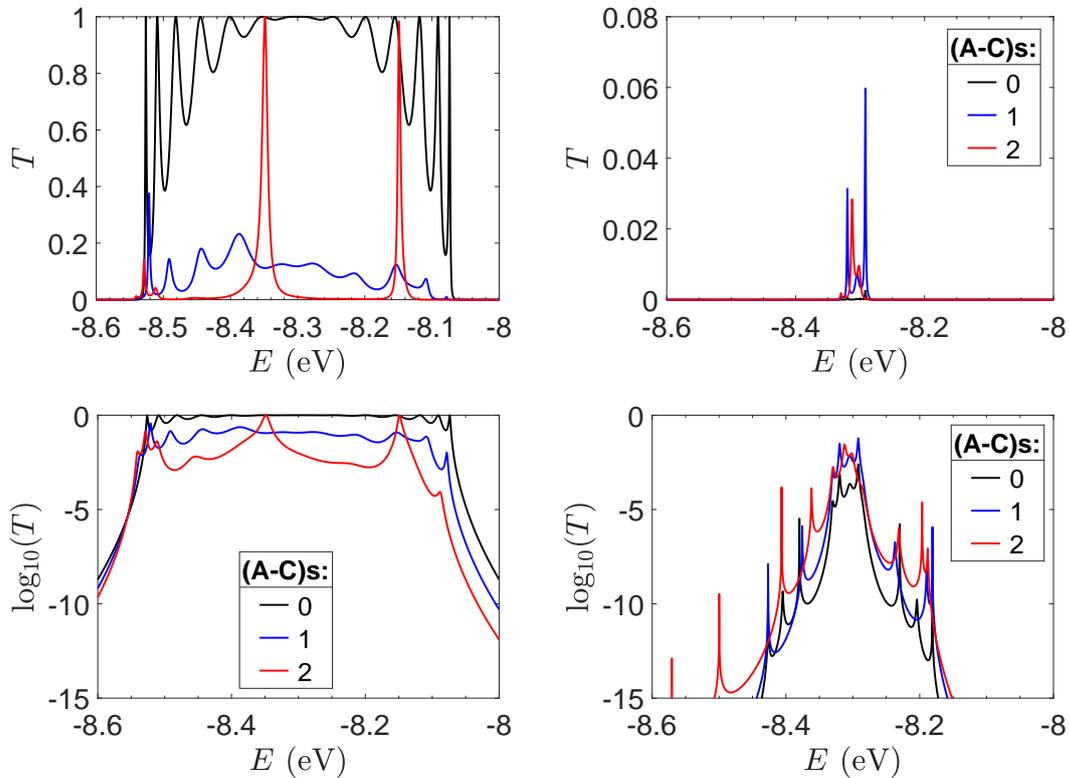
Figure 5.11: Transmission coefficient of the studied, initially G_{14} sequences with zero, one and two randomly positioned A-C mismatch mutations. Mutations are placed at the same sites for both ideal and natural sequences. Upper panels: Normal scale, lower panels: $\log_{10}(T(E))$ scale. Left: ideal polymers, right: natural polymers.

$T(E)$ for ideal and natural G_{14} sequences with seven A-C mismatch mutations, randomly positioned in the sequence (at the same sites for ideal and natural sequences), are presented in Fig. 5.12. Fig. 5.12 should therefore be compared with the left of Fig. 5.9. Including seven mutations, i.e., 50% of the monomers, the polymer becomes a random binary sequence. We observe an influence of the inclusion of A-C monomers; there are some lightly conducting

states closer to $E_{\text{A-C}}$, cf. Table 2.4. However, since E_{m} is positioned at $E_{\text{G−C}}$, this effect is small.
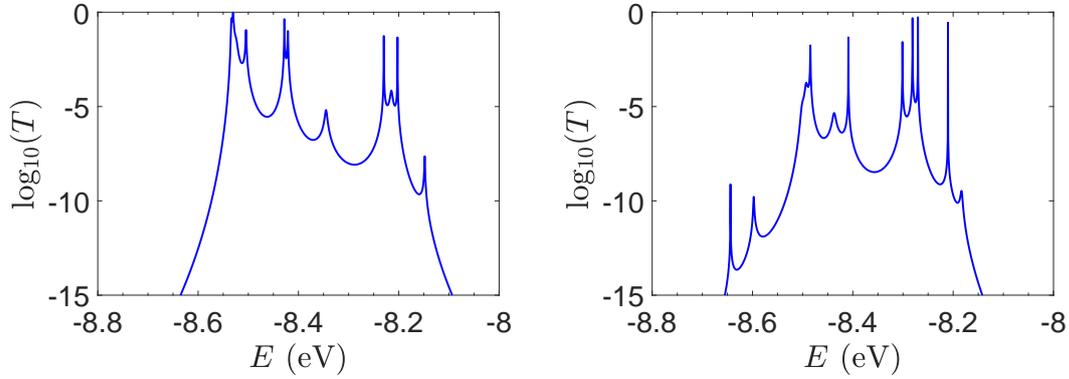
Figure 5.12: The $\log_{10}(T(E))$ of the studied, initially G_{14} sequences, but with seven A-C mismatch mutations, randomly positioned in the sequence (at the same sites for ideal and natural sequences). Left: ideal polymers, right: natural polymers.

$T(E)$ for the studied DNA sequences of ideal geometries with STR expansion mutations are presented in Fig. 5.13. As their DOS suggest (cf. Fig. 5.8), these
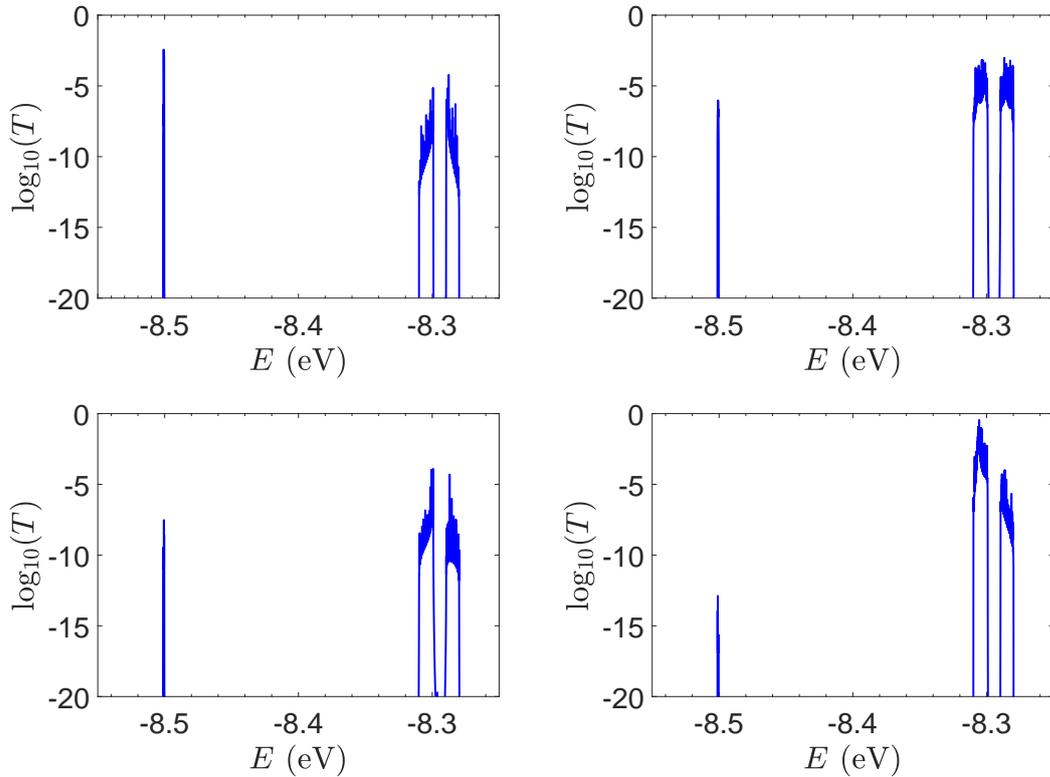
Figure 5.13: The $\log_{10}(T(E))$ of the studied sequences (ideal geometries) with STR expansion mutations. Upper left: Huntington's disease with 100 STR expansions, upper right: Kennedy's disease with 45 STR expansions, left lower: Spinocerebellar ataxia 6 with 30 STR expansions, lower right: Spinocerebellar ataxia 7 with 100 STR expansions.

sequences display narrow regions close to $E_{\text{G-C}}$ and $E_{\text{A-T}}$ within which transmission is allowed. The relative contribution of each region as well as the overall transmission profile is different for each sequence, allowing for distinct current-voltage curves, as it will be seen below.

5.3.3   Current - Voltage curves

The I-V curves of the studied ideal and natural DNA sequences without mutations are shown in Fig. 5.14, assuming $E_{\text{m}} = E_{\text{G-C}}$ for G_{14} and $E_{\text{m}} = E_{\text{A-T}}$ for A_{15}. The left panels of Fig. 5.14 are a subset of Fig. 5.2, for $E_{\text{m}} = E_{\text{G-C}} = -8.3$ eV.
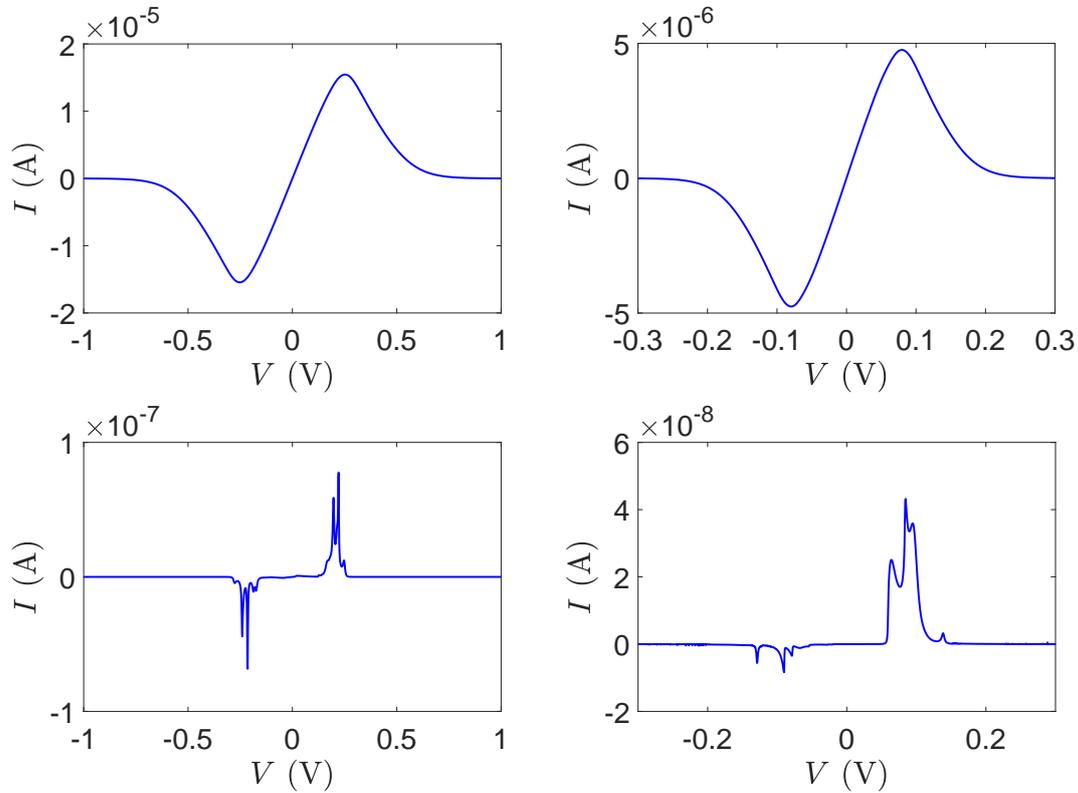
Figure 5.14: The I-V curves of the studied DNA sequences without mutations. Upper: ideal polymers, lower: natural polymers, left: G_{14} (with $E_{\text{m}} = E_{\text{G-C}}$), right: A_{15} (with $E_{\text{m}} = E_{\text{A-T}}$).

The order of magnitude of the I-V curves and their shape varies dramatically when many mutations are included. Hence, we have tried to devise another physical magnitude that may be used to characterise the I-V curves. This is the normalised deviation of the I-V from the origin, defined as

$$\text{NDIV}^+ = \frac{\int_0^\infty dV\ I(V)\ V}{\int_0^\infty dV\ I(V)}, \tag{5.10}$$

for the positive V regime and as

$$\text{NDIV}^- = \frac{\int_{-\infty}^0 dV\; I(V)\; V}{\int_{-\infty}^0 dV\; I(V)},\tag{5.11}$$

for the negative V regime. Then, NDIV is defined as

$$\text{NDIV} = \frac{\left|\text{NDIV}^+\right| + \left|\text{NDIV}^-\right|}{2}.\tag{5.12}$$

Figs. 5.15 and 5.16 display I-V related diagrams of the studied ideal and natural G_{14} DNA sequences, respectively, with one A-C mismatch mutation of varying position (left columns) and with varying number of A-C mismatch mutations randomly distributed in the sequence (right columns). We sketch the I-V curves, the $\log_{10}|I| - V$ curves (i.e., in logarithmic $|I|$ scale) and the newly introduced quantity, i.e., the normalised deviation of the I-V from the origin, NDIV. It can be seen that, for ideal segments, generally, the I-V curves do not vary significantly with the position of one A-C mismatch in the sequence ($\approx$ half an order of magnitude) for ideal segments; in natural segments the position of the mismatch affects the current more significantly (some orders of magnitude). The variation of the I-V curves becomes much more significant with increasing the number of A-C mismatches (many orders of magnitude). As a particular example, we show in Fig. 5.17 the I-V curves of the studied ideal (left) and natural (right) DNA sequences with 7 A-C mismatch mutations, randomly inserted in the sequence.
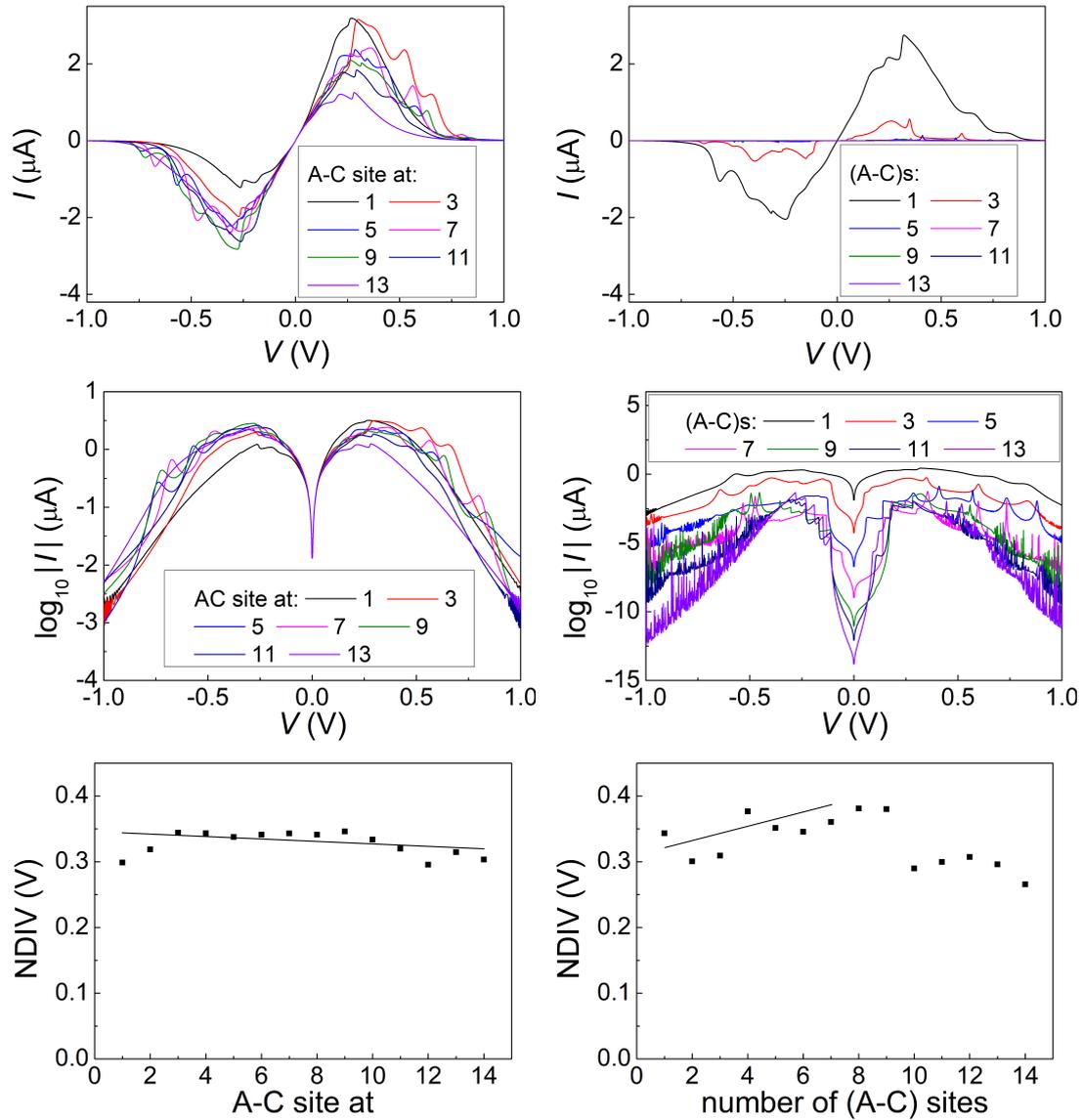
Figure 5.15: The *I-V* related diagrams of the studied G_{14} DNA sequences of ideal geometry with one A-C mismatch mutation of varying position in the sequence (left panels) and varying number of A-C mismatch mutations randomly inserted in the sequence (right panels). First row: *I-V* curves, second row: $\log_{10}|I| - V$ curves, i.e., in logarithmic scale, third row: NDIV, i.e., normalised deviation of the *I-V* curve from the origin.
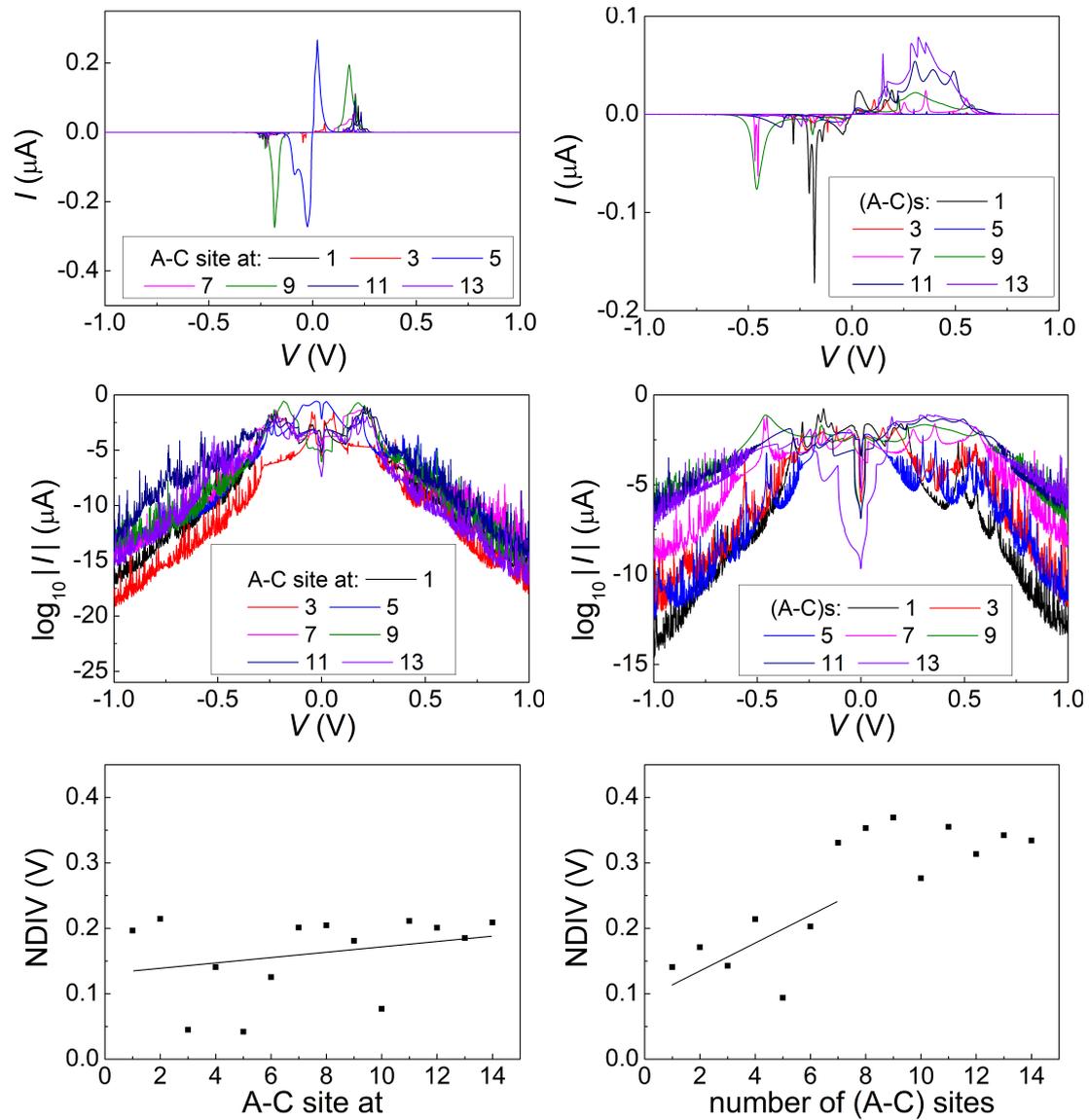
Figure 5.16: The *I-V* related diagrams of the studied G_{14} DNA sequences of natural geometry with one A-C mismatch mutation of varying position in the sequence (left panels) and varying number of A-C mismatch mutations randomly inserted in the sequence (right panels). First row: *I-V* curves, second row: $\log_{10}|I| - V$ curves, i.e., in logarithmic scale, third row: NDIV, i.e., normalised deviation of the *I-V* curve from the origin.
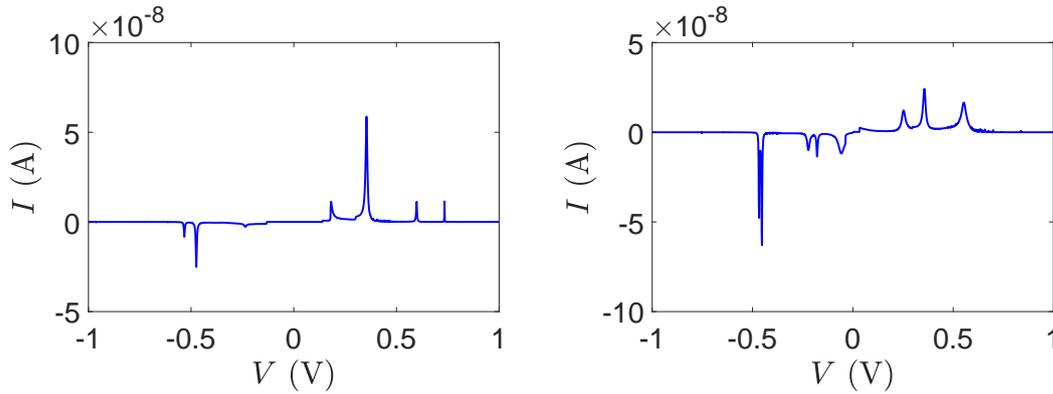
Figure 5.17: The *I-V* curves of the studied DNA sequences, initially G_{14}, but with 7 A-C mismatch mutations, randomly inserted in the sequence. Left: ideal polymers, right: natural polymers.

NDIV in ideal sequences with one A-C mismatch of varying position remains almost constant; the slope of NDIV versus the A-C site position is close to zero. However, NDIV in ideal sequences with increasing number of A-C mismatch mutations does not remain constant; the slope of NDIV versus the number of A-C mismatch mutations is positive until the number of (A-C)s becomes equal to the number of (G-C)s. Of course, after that point, the number of (A-C)s becomes larger than the number of (G-C)s; a further increase of the number of (A-C)s stabilises the situation. NDIV in natural sequences is similar but with pronounced slopes, especially when introducing an increasing number of A-C mismatch mutations. Hence, NDIV is a useful quantity to characterise these sequences.

In Figs. 5.18 and 5.19 we present the *I-V* related diagrams of the studied DNA sequences (ideal geometry) with STR expansion mutations. For all studied cases, changes in the *I-V* curves become more pronounced with increasing the number of STR expansion mutations (i.e., the number of CAG repeats). The respective NDIV display significant but almost monotonous variations, and can, therefore, be used to evaluate the number of (CAG) repeats in the sequence. This behaviour of the NDIV versus the number of CAG repeats suggests that it can be used to characterise the grade of danger for developing the studied diseases.
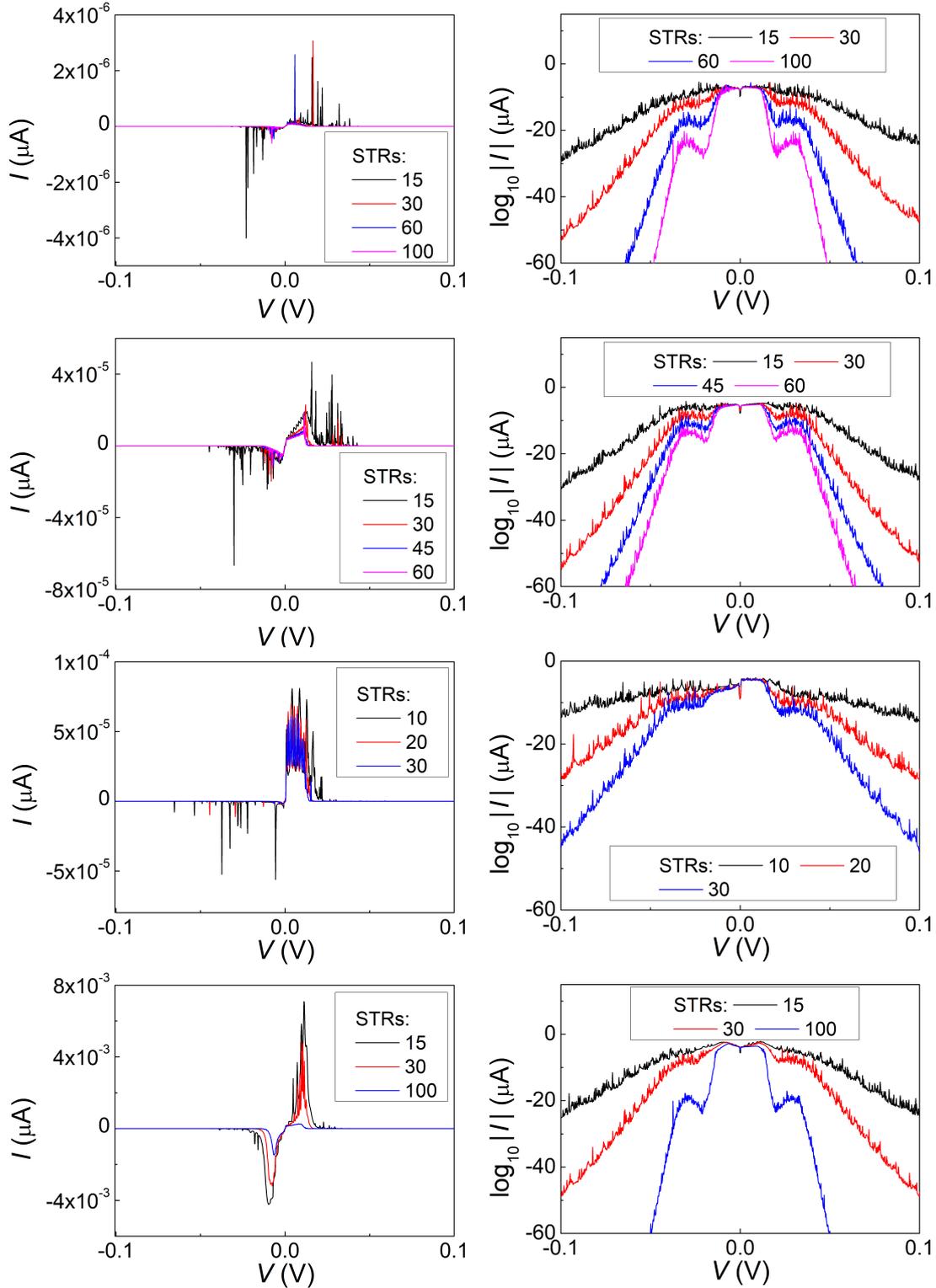
Figure 5.18: The *I-V* related diagrams (*I-V* curves and $\log_{10}|I|(V)$ curves, i.e, in logarithmic scale) of the studied DNA sequences (ideal geometry) with varying STR expansions, i.e., with different number of (CAG) triplets. First row: Huntington's disease, second row: Kennedy's disease, third row: Spinocerebellar ataxia 6, fourth row: Spinocerebellar ataxia 7.
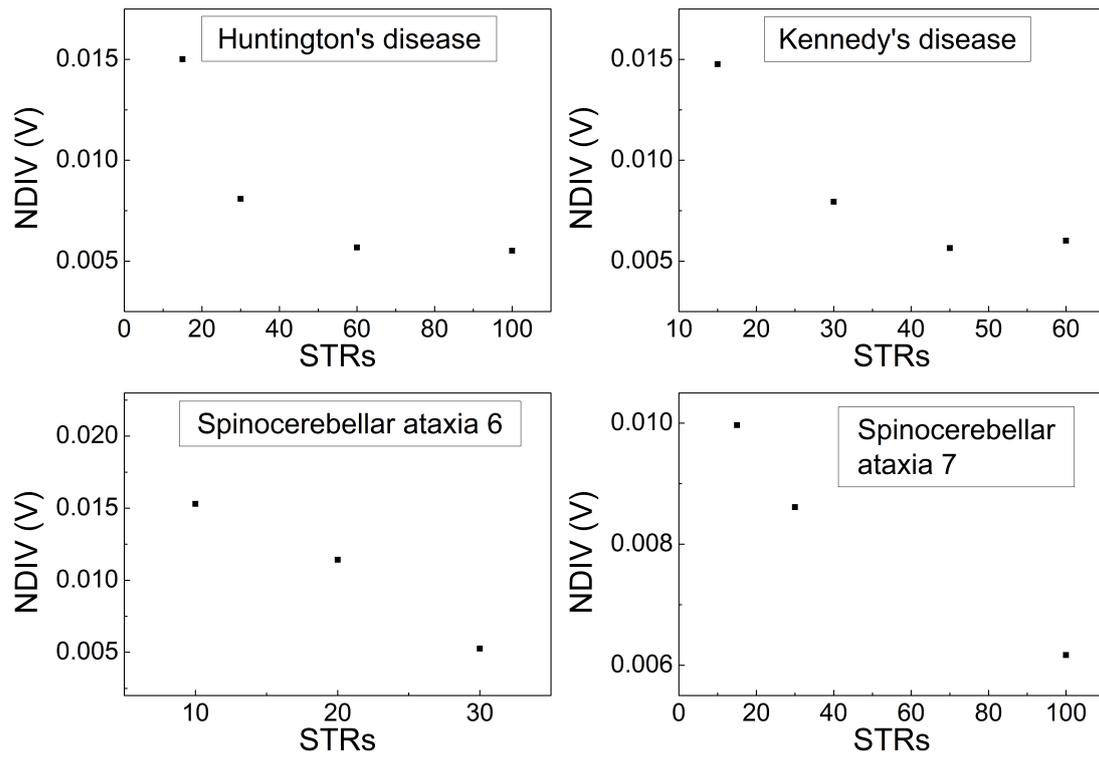
Figure 5.19: Normalised deviation of the *I-V* curve from the origin (NDIV) as a function of the number of (CAG) repeats aka short tandem repeat (STR) expansions. Upper left panel: Huntington's disease, upper right panel: Kennedy's disease, left lower panel: Spinocerebellar ataxia 6, lower right panel: Spinocerebellar ataxia 7.

6 | CONCLUSION AND PERSPECTIVES

The present Ph.D. Thesis aimed to a systematic study of the electronic structure of DNA molecular wires, as well as of the charge transfer and transport properties along them, using the LCAO method and a TB wire model. The computational part for the electronic structure employing all valence orbitals was implemented entirely by the author. The computational part regarding charge transfer was a result of a series of additions and modifications made by the author to already existing programs, implemented within the group "Physics of Nanosctructures and Biomaterials[1]". As for the code used for the charge transport computations, it was developed by Dr. Konstantinos Lambropoulos, a member of the aforementioned group, and was optimized by the author in order to implement computations for mutant sequences. The MD computations have been delivered by Prof. Rosa di Felice[2]. Other supportive computational codes were implemented by the author.

In Chapter 2, we calculated the lowest ionization and excitation energies of various biologically important molecules, such as adenine and isomers, guanine and isomers, purine and isomers, thymine, cytosine, pyrimidine and isomers, uracil and isomers, and other related planar heterocyclic molecules, which are π-conjugated systems. First, we employed the LCAO method in order to calculate the lowest ionization and excitation energies of the aforementioned "ideal" (frozen) heterocyclic organic molecules with a biological function, including the DNA and RNA bases and isomers. We briefly described the MMTS parameterization [55] within a simple semi-empirical LCAO approach that takes into account only the $2p_z$ atomic orbitals, and compared its results with the former HKS parameterization [62, 63] and with results from ab-initio (post Hartree-Fock) techniques. In most cases, the MMTS parameterization is more successful than the HKS parameterization with respect to both the experimental and the CC results. Then, we introduced a novel parameterization [27] that accounts for all valence orbitals, i.e., 2s, 2px, 2py, 2pz orbitals for C, N and O atoms and 1s orbital for H atoms. This LCAO approach is more suitable than the standard LCAO parameterization to investigate non-planar geometries. We predict ionization and excitation energies with RMSPE 3.65% and 6.49%, respectively, compared to the experimental values. Based on these errors, we infer that the proposed computational strategy is an adequate tool for a quick and relatively

---

[1]National and Kapodistrian University of Athens, Department of Physics, Section of Condensed Matter Physics;

[2]Department of Physics and Astronomy and Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, United States; CNR-NANO Modena, 41125 Modena, Italy;

accurate estimation of the electronic structure for a variety of organic molecules. We also computed the transition oscillator strengths f in a simplistic approximation, considering point contribution of the corresponding orbitals. Using the computed energies of the HOMO and LUMO within the proposed LCAO method, we then evaluated the energy levels of DNA base pairs (A-T, G-C). Our results are in good agreement with reference data.

In Chapter 3, we have systematically studied all the TB parameters which are necessary for the description of charge transfer along DNA. We evaluated the interaction integrals between successive base pairs using all valence orbitals, for both electrons and holes. We compared our results with interaction integrals evaluated by using the TB model employing only $2p_z$ valence orbitals, with HKS [62, 63] and MMTS [55] parameterization, and with Ref. [115] where interaction intergrals from different methods where taken into account. Our predictions are in good agreement with reference data. The obtained interaction integrals can be used in further studies of charge transfer/transport in DNA oligomers and polymers.

Having obtained the complete set of TB parameters for charge transfer, we moved on to employ it for the study of periodic and aperiodic DNA sequences. First, we comparatively studied the energy structure and the transfer of an extra carrier, electron or hole, along N monomer periodic polymers, made of the same monomer, i.e., I1, I2, I4, I6, I8, I10, I20, as well as made of different monomers, i.e., D2, D4, D6, D8, D10, D20 [97]. The TB wire model used here, employs only $2p_z$ valence orbitals. We determined various physical quantities: the HOMO and LUMO eigenspectra and density of states, the HOMO-LUMO gap, the mean over time probability to find the carrier at each monomer, the frequency content of carrier transfer, and the pure mean transfer rate k. To express clearly the frequency content, using the Fourier spectra, we defined two new physical quantities: the WMF of each monomer and the TWMF of the whole polymer. We found that, for I polymers, as the repetition unit P increases, all studied quantities related to charge transfer have the homopolymers as a limit regarding their efficiency. In the case of D polymers, the limit is the union of the two related homopolymers. Increasing P, k (from the first to the last monomer) falls drastically. As for the frequency content of carrier transfer (in terms of the TWMF), it lies within the THz regime, for both I and D polymers. Although $k(N)$ is a decreasing function, it can be increased, for the same N, by many orders of magnitude with appropriate sequence choice. Finally, the homopolymers display higher k, hence they are more efficient in terms of electron and hole transfer.

We found that as the repetition unit increases, in the former case, all studied quantities related to charge transfer (eigenspectra, DOS, energy gaps, mean over time probabilities to find the extra carrier at each base pair, TWMFs, pure mean transfer rates) have the homopolymers as a limit regarding their efficiency, while, in the latter case, the limit is the union of the two possible homopolymers.

Subsequently, we systematically studied the electronic structure and the coherent transfer of an extra carrier, along various categories of binary quasi-periodic (Fibonacci, Thue-Morse, Double-Period, Rudin-Shapiro) and fractal

(Cantor Set, Asymmetric Cantor Set) polymers consisting of either the same monomer (I polymers) or different monomers (D polymers) [96]. We used the same simple TB wire model with parameters extracted by using only $2p_z$ atomic orbitals (parameterization introduced in [62, 63]). Regarding the energy structure of the polymers, the eigenenergies lie around the monomers' on-site energies. We demonstrated that for I polymers, the eigenenergies are always symmetric relative to the (constant) monomer on-site energy. For both I and D polymers, in quasi-periodic cases the DOS has rather acute subbands, while in fractal cases it is fragmented and spiky. D polymers pose smaller HOMO-LUMO gaps than I polymers and their band limits lie within the energy regions defined by the respective limits of I polymers. As for the mean over time probabilities, they are significant only rather close to the first monomer, although in some cases we observe non-negligible probabilities at more distant monomers. For D polymers, the mean over time probabilities are generally negligible further than the first monomer. This situation in aperiodic polymers, where the mean over time probabilities generally decline a little away from the site where the carrier was initially placed, is in contrast to the situation in periodic polymers [97, 106, 116], where, generally, non-negligible probabilities exist at distant sites.

We, as well, investigated the frequency content of coherent extra carrier transfer via the TWMF of the polymer, using the WMF of the Fourier spectra that correspond to the probabilities to find the carrier at each monomer. In all cases, the TWMF lies in the THz regime, $\approx 10^{-2} - 10^2$ THz, and generally stabilizes after a few generations. This is different from various cases of periodic I and D polymers [97], where the TWMFs were found in the region $\approx 10^0 - 10^2$ THz. The study of the pure mean transfer rates, $k(N)$, shows that I polymers, which are simpler cases in terms of energy intricacy, are more efficient than D polymers regarding coherent hole and electron transfer. Furthermore, a random shuffle of a quasi-periodic or fractal monomer sequence destroys the deterministic character of its construction rules, thus leading to vanishing transfer rates. Comparison with the experiment, revealed that large variations of the TB parameters are expected in real situations, hence modifications are necessary. Using a modified parameterization, we were able to find hole pure mean transfer rates k of similar magnitude with experimental transfer rates K obtained by time-resolved spectroscopy.

Comparing periodic [97] and aperiodic polymers reveals that although generally periodic polymers are more efficient in terms of charge transfer, specific aperiodic polymers can be better than periodic ones [96]. However, the structurally simplest periodic polymers, i.e., the homopolymers [97], represent an unreachable limit for all aperiodic polymers.

Subsequently, we were interested to address the impact of structural flexibility (dynamics) on the electronic structure and charge transfer ability of B-DNA [18, 27]. To this end, in Chapter 4, we applied our LCAO method to 20 AA and GG dimers, extracted from representative structures in a classical MD trajectory of a 20mer. For all these systems, we calculated the difference between the on-site energies Δ and the interaction parameters t, as well as the maximum transfer percentage between the two monomers of a dimer p. We found that the

values of Δ and p are significantly affected by geometrical changes. Nevertheless, in the majority of the studied dimers, the maximum transfer percentage is very close to unity. RT-TDDFT results show that structural changes affect the DNA charge transfer properties [18], such as the time evolution of the probability to find the carrier at each monomer, its mean value and the maximum transfer percentage. Our study examines and quantifies the impact structural changes have on the time evolution of the hole population, which is one possible physical quantity to characterize charge transfer. In brief, charge transfer in AA and GG dimers, although affected by structural variability, remains significant, on the average. The TB results for the periods T and the maximum transfer percentages p were compared with those obtained by RT-TDDFT. Using the TB method employing only $2p_z$ orbitals, quantitative agreement could not be reached but a qualitative assessment of the RT-TDDFT results is possible. Nevertheless, the TB method employing all valence orbitals gives results with similar trends for the transfer integrals as with DFT method, which is a promising result for accurate and less computationally costly parameter calculations.

We suggest that the proposed methodology can be used in a high-throughput manner to characterize dynamical effects on charge transfer in organic polymers constituted of heterocyclic building blocks. Our cost-effective simple method is suitable for very fast computations of electronic structure and transfer integrals. It can greatly facilitate charge transfer calculations, illuminating its fundamental mechanisms, in sequences of arbitrary geometry taken, e.g., by MD simulations, as far as purines, pyrimidines, and similar molecules are the constituents. Although we took only valence orbitals for carbon, nitrogen, oxygen, and hydrogen into account, this approach could be generalized to include other atomic species and orbitals.

Chapter 5 of the present Thesis was devoted to the investigation of charge transport properties of DNA, and more specifically to the effect of two types of mutations: point substitution transitions and STR expansions. We focused on the following physical quantities: eigenspectra, density of states, transmission coefficients and current-voltage curves. Both ideal (textbook geometry) and natural (naturally distorted, geometry from databases) sequences were considered. Once again, the TB wire model utilizing all valence orbitals was recruited in order to include the structural variability effects, in conjunction with a transfer matrix technique. The results point out interesting features in all the aforementioned physical quantities that discriminate mutated from unmutated sequences.

The most experimentally relevant quantities are the I-V curves. Their characteristics, when introducing mutations, change, rather strongly. Since both the order of magnitude and the shape of the I-V curves varies when introducing mutations, the Normalized Deviation of the I-V curve from the origin (NDIV) was defined to make things clearer. In ideal sequences with one A-C mismatch of varying position, the NDIV remains almost constant: its slope versus the mismatch position is close to zero. However, in ideal sequences with increasing number of A-C mismatch mutations, the NDIV varies. Its slope versus the number of A-C mismatch mutations is positive, until the number of (A-C)s becomes equal to the number of (G-C)s. After that point, since the number of

(A-C)s becomes larger than the number of (G-C)s, a further increase of the number of (A-C)s stabilises the situation. NDIV in natural sequences is similar but with pronounced slopes, especially when introducing an increasing number of A-C mismatch mutations. Hence, NDIV is a useful quantity to characterise these sequences. As far as the STR expansions are concerned, although dramatic changes in the I-V curves occur for all studied cases as the number of CAG repeats increases, NDIV shows significant but almost monotonous variation. Therefore, it can be used to evaluate the number of (CAG) repeats in the sequence, and so the NDIV can be used to characterise the grade of danger for developing the studied diseases. Overall, the NDIV is generally insensitive to the position of a point mutation, but rather sensitive to the number of point mutations and STR expansion mutations.

Transitions of C $\leftrightarrow$ T exchange have not been included because their effects could not be properly grasped within the wire model. These mutations could be properly studied within the extended ladder model [106], i.e., a TB description at the single-base level; this will be hopefully done in the future. Transitions are more likely than transversions (interchange purine $\leftrightarrow$ pyrimidine), because it is easier to substitute a single ring by another single ring than a double ring for a single ring or vice versa. To study transversions, a careful geometry optimization is necessary.

Another category of mutations is that of germline mutations [242], i.e., a gene change in a reproductive cell that becomes incorporated into the DNA of every cell in the body of the offspring. This way the mutation can be passed from parent to offspring, and is, therefore, hereditary. This could be the subject of a future study, as well.

All diseases studied in this Ph.D. Thesis, have the same triplet motif, i.e. $(CAG)_n$, between 9 base pairs at the start and other 9 base pairs at the other end (primers). Of course, the primers are different for each disease, but they only contain 18 base pairs altogether, which is not a large number when dealing with a sequence of 180 or 300 base pairs. Under these conditions, it is not safe to draw direct conclusions regarding the identity of the disease. This issue could be possibly tackled by including larger primers, which would produce much more distinctive DOS or IDOS features and allow for sequence recognition. Another perspective would be to consider, e.g., a 300 base-pair sequence and change both n and the length of primers, while keeping the same number of total base pairs (e.g., 300). These different perspectives could hopefully be included in future work.

Overall, the interplay between periodicity and aperiodicity in biology [243] is a vast area of extreme interest to us and novel methods must be devised to explore it. In the present Ph.D. Thesis, DNA was considered as a prototype system, as far as it provides the opportunity to construct easily different polymers by combining a given number of monomers. As we already demonstrated, charge transport and transfer properties of DNA are strongly sequence-dependent. Additionally, considering its fairly large persistent length, one can assume the reason why DNA is a promising candidate as an element of nano-circuits and relevant applications.

Appendices

A Schrödinger Equation and Hamiltonian Matrix Elements

The general formula of time-dependent Schrödinger equation in one dimension is:

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle \Rightarrow \tag{1}$$

$$i\hbar\frac{\partial}{\partial t}\langle x|\psi(t)\rangle = \langle x|\hat{H}|\psi(t)\rangle \Rightarrow \tag{2}$$

$$i\hbar\frac{\partial}{\partial t}\langle x|\psi(t)\rangle = \int dx'\langle x|\hat{H}|x'\rangle\langle x'|\psi(t)\rangle \Rightarrow \tag{3}$$

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \int dx'\hat{H}(\frac{\hbar}{i}\frac{\partial}{\partial x},x)\delta(x-x')\psi(x',t) \Rightarrow \tag{4}$$

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \hat{H}(\frac{\hbar}{i}\frac{\partial}{\partial x},x)\psi(x,t). \tag{5}$$

The above equation stands for the position representation of the time-dependant Schrödinger equation.

Thus:

$$H_{j\mu i\nu} = \langle\phi_{j\mu}|\hat{H}|\phi_{i\nu}\rangle = \tag{6}$$

$$\int d^3\vec{r'}\int d^3\vec{r}\langle\phi_{j\mu}|\vec{r'}\rangle\langle\vec{r'}|\hat{H}|\vec{r}\rangle\langle\vec{r}|\phi_{i\nu}\rangle =$$

$$\int d^3\vec{r'}\int d^3\vec{r}\phi_{j\mu}(\vec{r'})^*\hat{H}\delta(\vec{r'}-\vec{r})\phi_{i\nu}(\vec{r}) =$$

$$\int d^3\vec{r}\phi_{j\mu}(\vec{r})^*\hat{H}\phi_{i\nu}(\vec{r}).$$

and

$$S_{j\mu i\nu} = \langle\phi_{j\mu}|\phi_{i\nu}\rangle = \tag{7}$$

$$\int d^3\vec{r'}\int d^3\vec{r}\langle\phi_{j\mu}|\vec{r'}\rangle\langle\vec{r'}|\vec{r}\rangle\langle\vec{r}|\phi_{i\nu}\rangle =$$

$$\int d^3\vec{r'}\int d^3\vec{r}\phi_{j\mu}(\vec{r'})^*\delta(\vec{r'}-\vec{r})\phi_{i\nu}(\vec{r}) =$$

$$\int d^3\vec{r}\phi_{j\mu}(\vec{r})^*\phi_{i\nu}(\vec{r}).$$

B Hamiltonian Diagonalization

Example 1: Diagonalization of the Hamiltonian matrix of the cyclobutadiene molecule (chemical formula C_4H_4), using the LCAO method, employing only the $2p_z$ valence orbitals.

In Fig. 1 we depict the structure of the cyclobutadiene molecule, consisted of 4 carbon atoms and 4 hydrogen atoms. The geometry in this particular example is taken from the National Institute of Standards and Technology (NIST) webpage: https://www.nist.gov.
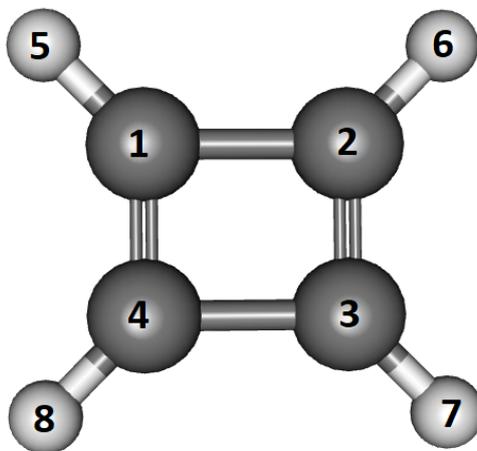
Figure 1: The cyclobutadiene molecule with chemical formula C_4H_4, consisted of 4 carbon atoms and 4 hydrogen atoms.

In this simplistic LCAO approximation we neglect the hydrogen 1s orbitals and use only the carbon $2p_z$ orbitals to build the molecular Hamiltonian. Using the procedure of Subsection 2.1.1, the Hamiltonian matrix reads:

$$\hat{H}^b = \begin{bmatrix} E_{C(1)} & V_{1,2} & 0 & V_{1,4} \\ V_{2,1} & E_{C(2)} & V_{2,3} & 0 \\ 0 & V_{3,2} & E_{C(3)} & V_{3,4} \\ V_{4,1} & 0 & V_{4,3} & E_{C(4)} \end{bmatrix}. \tag{8}$$

Employing the MMTS parameterization [55], we get:

$$\hat{H}^b = \begin{bmatrix} -6.56 & -3.29 & 0 & -2.35 \\ -3.29 & -6.56 & -2.35 & 0 \\ 0 & -2.35 & -6.56 & -3.29 \\ -2.35 & 0 & -3.29 & -6.56 \end{bmatrix}. \tag{9}$$

At this point we diagonalize the above matrix using the **eig** command in MATLAB programming language, using LAPACK functions. This way we obtain the

energy eigenvalues

$$E_k^b = \begin{bmatrix} -12.21 \\ -7.50 \\ -5.62 \\ -0.91 \end{bmatrix}, \tag{10}$$

of which $E_{HOMO}^b = -7.50$ and $E_{LUMO}^b = -5.61$. We also obtain the eigenvector matrix

$$C_k^b = \begin{bmatrix} -0.50 & 0.50 & -0.50 & 0.50 \\ -0.50 & 0.50 & 0.50 & -0.50 \\ -0.50 & -0.50 & 0.50 & 0.50 \\ -0.50 & -0.50 & -0.50 & -0.50 \end{bmatrix}, \tag{11}$$

of which

$$C_{HOMO}^b = \begin{bmatrix} 0.50 \\ 0.50 \\ -0.50 \\ -0.50 \end{bmatrix}, \tag{12}$$

and

$$C_{LUMO}^b = \begin{bmatrix} -0.50 \\ 0.50 \\ 0.50 \\ -0.50 \end{bmatrix}. \tag{13}$$

Example 2: Diagonalization of the Hamiltonian matrix of the cyclobutadiene molecule (chemical formula C_4H_4), using the LCAO method, employing all valence orbitals.

In this LCAO approximation we use all valence orbitals, that means, the hydrogen 1s orbitals and the carbon 2s, $2p_x$, $2p_y$, $2p_z$ orbitals to build the molecular Hamiltonian. Using the procedure of Subsection 2.1.2, the Hamiltonian matrix reads:

$$
\hat{H}^b =
\begin{bmatrix}
E_{C(2s)(1)} & 0 & 0 & 0 & V_{ss(1,2)} & V_{sx(1,2)} & V_{sy(1,2)} & V_{sz(1,2)} & 0 & 0 \\
0 & E_{C(2p)(1)} & 0 & 0 & V_{xs(1,2)} & V_{xx(1,2)} & V_{xy(1,2)} & V_{xz(1,2)} & 0 & 0 \\
0 & 0 & E_{C(2p)(1)} & 0 & V_{ys(1,2)} & V_{yx(1,2)} & V_{yy(1,2)} & V_{yz(1,2)} & 0 & 0 \\
0 & 0 & 0 & E_{C(2p)(1)} & V_{zs(1,2)} & V_{zx(1,2)} & V_{zy(1,2)} & V_{zz(1,2)} & 0 & 0 \\
V_{ss(2,1)} & V_{sx(2,1)} & V_{sy(2,1)} & V_{sz(2,1)} & E_{C(2s)(2)} & 0 & 0 & 0 & V_{ss(2,3)} & V_{sx(2,3)} \\
V_{xs(2,1)} & V_{xx(2,1)} & V_{xy(2,1)} & V_{xz(2,1)} & 0 & E_{C(2p)(2)} & 0 & 0 & V_{xs(2,3)} & V_{xx(2,3)} \\
V_{ys(2,1)} & V_{yx(2,1)} & V_{yy(2,1)} & V_{yz(2,1)} & 0 & 0 & E_{C(2p)(2)} & 0 & V_{ys(2,3)} & V_{yx(2,3)} \\
V_{zs(2,1)} & V_{zx(2,1)} & V_{zy(2,1)} & V_{zz(2,1)} & 0 & 0 & 0 & E_{C(2p)(2)} & V_{zs(2,3)} & V_{zx(2,3)} \\
0 & 0 & 0 & 0 & V_{ss(3,2)} & V_{sx(3,2)} & V_{sy(3,2)} & V_{sz(3,2)} & E_{C(2s)(3)} & 0 \\
0 & 0 & 0 & 0 & V_{xs(3,2)} & V_{xx(3,2)} & V_{xy(3,2)} & V_{xz(3,2)} & 0 & E_{C(2p)(3)} \\
0 & 0 & 0 & 0 & V_{ys(3,2)} & V_{yx(3,2)} & V_{yy(3,2)} & V_{yz(3,2)} & 0 & 0 \\
0 & 0 & 0 & 0 & V_{zs(3,2)} & V_{zx(3,2)} & V_{zy(3,2)} & V_{zz(3,2)} & 0 & 0 \\
V_{ss(4,1)} & V_{sx(4,1)} & V_{sy(4,1)} & V_{sz(4,1)} & 0 & 0 & 0 & 0 & V_{ss(4,3)} & V_{sx(4,3)} \\
V_{xs(4,1)} & V_{xx(4,1)} & V_{xy(4,1)} & V_{xz(4,1)} & 0 & 0 & 0 & 0 & V_{xs(4,3)} & V_{xx(4,3)} \\
V_{ys(4,1)} & V_{yx(4,1)} & V_{yy(4,1)} & V_{yz(4,1)} & 0 & 0 & 0 & 0 & V_{ys(4,3)} & V_{yx(4,3)} \\
V_{zs(4,1)} & V_{zx(4,1)} & V_{zy(4,1)} & V_{zz(4,1)} & 0 & 0 & 0 & 0 & V_{zs(4,3)} & V_{zx(4,3)} \\
V_{ss(5,1)} & V_{sx(5,1)} & V_{sy(5,1)} & V_{sz(5,1)} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & V_{ss(6,1)} & V_{sx(6,1)} & V_{sy(6,1)} & V_{sz(6,1)} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & V_{ss(7,1)} & V_{sx(7,1)} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$$
\begin{bmatrix}
0 & 0 & V_{ss(1,4)} & V_{sx(1,4)} & V_{sy(1,4)} & V_{sz(1,4)} & V_{ss(1,5)} & 0 & 0 & 0 \\
0 & 0 & V_{xs(1,4)} & V_{xx(1,4)} & V_{xy(1,4)} & V_{xz(1,4)} & V_{xs(1,5)} & 0 & 0 & 0 \\
0 & 0 & V_{ys(1,4)} & V_{yx(1,4)} & V_{yy(1,4)} & V_{yz(1,4)} & V_{ys(1,5)} & 0 & 0 & 0 \\
0 & 0 & V_{zs(1,4)} & V_{zx(1,4)}6 & V_{zy(1,4)} & V_{zz(1,4)} & V_{zs(1,5)} & 0 & 0 & 0 \\
V_{sy(2,3)} & V_{sz(2,3)} & 0 & 0 & 0 & 0 & 0 & V_{ss(2,6)} & 0 & 0 \\
V_{xy(2,3)} & V_{xz(2,3)} & 0 & 0 & 0 & 0 & 0 & V_{xs(2,6)} & 0 & 0 \\
V_{yy(2,3)} & V_{yz(2,3)} & 0 & 0 & 0 & 0 & 0 & V_{ys(2,6)} & 0 & 0 \\
V_{zy(2,3)} & V_{zz(2,3)} & 0 & 0 & 0 & 0 & 0 & V_{zs(2,6)} & 0 & 0 \\
0 & 0 & V_{ss(3,4)} & V_{sx(3,4)} & V_{sy(3,4)} & V_{sz(3,4)} & 0 & 0 & V_{ss(3,7)} & 0 \\
0 & 0 & V_{xs(3,2)} & V_{xx(3,4)} & V_{xy(3,4)} & V_{xz(3,4)} & 0 & 0 & V_{xs(3,7)} & 0 \\
E_{C(2p)(3)} & 0 & V_{ys(3,4)} & V_{yx(3,4)} & V_{yy(3,4)} & V_{yz(3,4)} & 0 & 0 & V_{ys(3,7)} & 0 \\
0 & E_{C(2p)(3)} & V_{zs(3,4)} & V_{zx(3,4)} & V_{zy(3,4)} & V_{zz(3,4)} & 0 & 0 & V_{zs(3,7)} & 0 \\
V_{sy(4,3)} & V_{sz(4,3)} & E_{C(2s)(4)} & 0 & 0 & 0 & 0 & 0 & 0 & V_{ss(4,8)} \\
V_{xy(4,3)} & V_{xz(4,3)} & 0 & E_{C(2p)(4)} & 0 & 0 & 0 & 0 & 0 & V_{xs(4,8)} \\
V_{yy(4,3)} & V_{yz(4,3)} & 0 & 0 & E_{C(2p)(4)} & 0 & 0 & 0 & 0 & V_{ys(4,8)} \\
V_{zy(4,3)} & V_{zz(4,3)} & 0 & 0 & 0 & E_{C(2p)(4)} & 0 & 0 & 0 & V_{zs(4,8)} \\
0 & 0 & 0 & 0 & 0 & 0 & E_{H(1s)(5)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & E_{H(1s)(6)} & 0 & 0 \\
V_{sy(7,1)} & V_{sz(7,1)} & 0 & 0 & 0 & 0 & 0 & 0 & E_{H(1s)(7)} & 0 \\
0 & 0 & V_{ss(8,1)} & V_{sx(8,1)} & V_{sy(8,1)} & V_{sz(8,1)} & 0 & 0 & 0 & E_{H(1s)(8)}
\end{bmatrix}
$$

Employing the MSF parameterization [27], we get:

$$\hat{H}^b = \begin{bmatrix}
-13.18 & 0 & 0 & 0 & -5.65 & 4.25 & 4.35 & 0 & 0 & 0 & 0 & 0 & -4.04 & 3.10 & -3.03 & -0.28 & -5.99 & 0 & 0 & 0 \\
0 & -6.70 & 0 & 0 & -4.25 & 3.04 & 6.31 & 0 & 0 & 0 & 0 & 0 & -3.10 & 2.36 & -4.49 & -0.42 & 6.44 & 0 & 0 & 0 \\
0 & 0 & -6.70 & 0 & -4.35 & 6.31 & 3.33 & 0 & 0 & 0 & 0 & 0 & 3.03 & -4.49 & 2.15 & 0.41 & 0.13 & 0 & 0 & 0 \\
0 & 0 & 0 & -6.70 & 0 & 0 & 0 & -3.12 & 0 & 0 & 0 & 0 & 0.28 & -0.42 & 0.41 & -2.19 & -0.29 & 0 & 0 & 0 \\
-5.65 & -4.25 & -4.35 & 0 & -13.18 & 0 & 0 & 0 & -4.04 & 3.10 & -3.03 & -0.28 & 0 & 0 & 0 & 0 & 0 & -5.99 & 0 & 0 \\
4.25 & 3.04 & 6.31 & 0 & 0 & -6.70 & 0 & 0 & -3.10 & 2.36 & -4.49 & -0.42 & 0 & 0 & 0 & 0 & 0 & 0.02 & 0 & 0 \\
4.35 & 6.31 & 3.33 & 0 & 0 & 0 & -6.70 & 0 & 3.03 & -4.49 & 2.15 & 0.41 & 0 & 0 & 0 & 0 & 0 & -6.44 & 0 & 0 \\
0 & 0 & 0 & -3.12 & 0 & 0 & 0 & -6.70 & 0.28 & -0.42 & 0.41 & -2.19 & 0 & 0 & 0 & 0 & 0 & -0.29 & 0 & 0 \\
0 & 0 & 0 & 0 & -4.04 & -3.10 & 3.03 & 0.28 & -13.18 & 0 & 0 & 0 & -5.65 & -4.25 & -4.35 & 0 & 0 & 0 & -5.99 & 0 \\
0 & 0 & 0 & 0 & 3.10 & 2.36 & -4.49 & -0.42 & 0 & -6.70 & 0 & 0 & 4.25 & 3.04 & 6.31 & 0 & 0 & 0 & -6.44 & 0 \\
0 & 0 & 0 & 0 & -3.03 & -4.49 & 2.15 & 0.41 & 0 & 0 & -6.70 & 0 & 4.35 & 6.31 & 3.34 & 0 & 0 & 0 & -0.13 & 0 \\
0 & 0 & 0 & 0 & -0.28 & -0.42 & 0.41 & -2.19 & 0 & 0 & 0 & -6.70 & 0 & 0 & 0 & -3.12 & 0 & 0 & 0.29 & 0 \\
-4.04 & -3.10 & 3.03 & 0.28 & 0 & 0 & 0 & 0 & -5.65 & 4.25 & 4.35 & 0 & -13.18 & 0 & 0 & 0 & 0 & 0 & 0 & -5.99 \\
3.10 & 2.36 & -4.49 & -0.42 & 0 & 0 & 0 & 0 & -4.25 & 3.04 & 6.31 & 0 & 0 & -6.70 & 0 & 0 & 0 & 0 & 0 & -0.02 \\
-3.03 & -4.49 & 2.15 & 0.41 & 0 & 0 & 0 & 0 & -4.35 & 6.31 & 3.34 & 0 & 0 & 0 & -6.70 & 0 & 0 & 0 & 0 & 6.44 \\
-0.28 & -0.42 & 0.41 & -2.19 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3.12 & 0 & 0 & 0 & -6.70 & 0 & 0 & 0 & 0.29 \\
-5.99 & 6.44 & 0.13 & -0.29 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -13.64 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -5.99 & 0.02 & -6.44 & -0.29 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -13.64 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.99 & -6.44 & -0.13 & 0.29 & 0 & 0 & 0 & 0 & 0 & 0 & -13.64 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.99 & -0.02 & 6.44 & 0.29 & 0 & 0 & 0 & -13.64
\end{bmatrix}$$

At this point we diagonalize the above matrix using the **eig** command in MAT-LAB programming language, using LAPACK functions. This way we obtain the

energy eigenvalues

$$
E_k^b = \begin{bmatrix}
-28.39 \\
-22.61 \\
-21.63 \\
-21.25 \\
-19.88 \\
-18.82 \\
-17.30 \\
-15.03 \\
-12.06 \\
-7.59 \\
-5.81 \\
-4.23 \\
-3.03 \\
-2.13 \\
-2.12 \\
-1.34 \\
-0.98 \\
3.42 \\
3.82 \\
9.28
\end{bmatrix}, \tag{14}
$$

of which $E_{HOMO}^b = -7.59$ eV, $E_{LUMO}^b = -5.81$ eV. We also obtain the eigenvector matrix

$$
C_k^b = \left[
\begin{array}{cccccccccccccccccccc}
0.42 & -0.38 & 0.15 & -0.31 & -0.23 & 0.09 & 0.05 & -0.23 & 0 & 0 & 0 & 0.22 & -0.30 & 0.27 & 0.37 & 0 & -0.10 & 0.18 & 0.02 & -0.23 \\
0.25 & 0.01 & -0.31 & 0.13 & 0.18 & 0.32 & 0.06 & -0.26 & -0.02 & 0.02 & -0.02 & -0.29 & 0.19 & -0.28 & -0.14 & 0.02 & -0.34 & 0.30 & 0.14 & -0.42 \\
0.07 & -0.19 & -0.03 & 0.10 & 0.01 & 0.24 & -0.49 & 0.30 & 0.02 & -0.02 & 0.02 & -0.01 & 0.21 & 0.14 & -0.06 & -0.02 & 0.33 & 0.36 & -0.48 & -0.13 \\
-0.01 & -0.01 & 0.01 & 0 & -0.01 & 0 & -0.03 & 0.03 & -0.50 & 0.50 & -0.50 & 0.01 & 0 & 0.02 & 0 & 0.50 & 0.03 & 0 & -0.03 & 0.01 \\
0.42 & -0.38 & 0.14 & 0.31 & 0.23 & 0.09 & 0.05 & 0.23 & 0 & 0 & 0 & 0.22 & -0.30 & -0.27 & -0.37 & 0 & -0.10 & -0.18 & -0.02 & 0.23 \\
-0.07 & 0.19 & 0.03 & 0.10 & 0 & -0.23 & 0.49 & 0.31 & -0.02 & 0.02 & 0.02 & 0 & -0.21 & 0.15 & -0.05 & -0.02 & -0.34 & 0.35 & -0.48 & -0.12 \\
-0.26 & -0.01 & 0.31 & 0.13 & 0.18 & -0.32 & -0.05 & -0.25 & 0.02 & -0.02 & -0.02 & 0.29 & -0.19 & -0.28 & -0.14 & 0.02 & 0.33 & 0.30 & 0.13 & -0.42 \\
-0.01 & -0.01 & 0.01 & 0 & 0.01 & 0 & -0.03 & -0.03 & -0.50 & 0.50 & 0.50 & 0.01 & 0 & -0.02 & 0 & -0.50 & 0.03 & 0 & 0.03 & -0.01 \\
0.42 & 0.38 & 0.14 & 0.31 & -0.23 & -0.09 & 0.05 & 0.23 & 0 & 0 & 0 & 0.22 & 0.30 & -0.25 & 0.39 & 0 & 0.10 & -0.18 & 0.02 & -0.23 \\
-0.25 & 0.01 & 0.31 & 0.13 & -0.18 & 0.32 & -0.06 & -0.26 & -0.02 & -0.02 & 0.02 & 0.29 & 0.19 & -0.27 & 0.16 & 0.02 & -0.34 & 0.30 & -0.14 & 0.42 \\
-0.07 & -0.19 & 0.03 & 0.10 & -0.01 & 0.24 & 0.49 & 0.30 & 0.02 & 0.02 & -0.02 & 0.01 & 0.21 & 0.14 & 0.05 & -0.02 & 0.33 & 0.36 & 0.48 & 0.13 \\
0.01 & -0.01 & -0.01 & 0 & 0.01 & 0 & 0.03 & 0.03 & -0.50 & -0.50 & 0.50 & -0.01 & 0 & 0.02 & 0 & 0.50 & 0.03 & 0 & 0.03 & -0.01 \\
0.42 & 0.38 & 0.14 & -0.31 & 0.23 & -0.09 & 0.05 & -0.23 & 0 & 0 & 0 & 0.22 & 0.30 & 0.25 & -0.39 & 0 & 0.10 & 0.18 & -0.02 & 0.23 \\
0.07 & 0.19 & -0.03 & 0.10 & 0 & -0.23 & -0.49 & 0.31 & -0.02 & -0.02 & -0.02 & 0 & -0.21 & 0.15 & 0.05 & -0.02 & -0.34 & 0.35 & 0.48 & 0.12 \\
0.26 & -0.01 & -0.31 & 0.13 & -0.18 & -0.32 & 0.05 & -0.25 & 0.02 & 0.02 & 0.02 & -0.29 & -0.19 & -0.27 & 0.16 & 0.02 & 0.33 & 0.30 & -0.13 & 0.42 \\
0.01 & -0.01 & -0.01 & 0 & -0.01 & 0 & 0.03 & -0.03 & -0.50 & -0.50 & -0.50 & -0.01 & 0 & -0.02 & 0 & -0.50 & 0.03 & 0 & -0.03 & 0.01 \\
0.06 & -0.26 & 0.36 & -0.36 & -0.41 & -0.29 & -0.01 & 0.19 & 0 & 0 & 0 & -0.34 & 0.28 & -0.30 & -0.28 & 0 & -0.12 & 0.05 & 0.04 & -0.06 \\
0.06 & -0.26 & 0.36 & 0.36 & 0.41 & -0.29 & -0.01 & -0.19 & 0 & 0 & 0 & -0.34 & 0.28 & 0.30 & 0.28 & 0 & -0.12 & -0.05 & -0.04 & 0.06 \\
0.06 & 0.26 & 0.36 & 0.36 & -0.41 & 0.29 & -0.01 & -0.19 & 0 & 0 & 0 & -0.34 & -0.28 & 0.28 & -0.29 & 0 & 0.12 & -0.05 & 0.04 & -0.06 \\
0.06 & 0.26 & 0.36 & -0.36 & 0.41 & 0.29 & -0.01 & 0.19 & 0 & 0 & 0 & -0.34 & -0.28 & -0.28 & 0.29 & 0 & 0.12 & 0.05 & -0.04 & 0.06
\end{array}
\right]
$$

of which

$$
C^b_{HOMO} = \begin{bmatrix} 0 \\ 0.02 \\ -0.02 \\ 0.50 \\ 0 \\ 0.02 \\ -0.02 \\ 0.50 \\ 0 \\ -0.02 \\ 0.02 \\ -0.50 \\ 0 \\ -0.02 \\ 0.02 \\ -0.50 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} , \qquad (15) \qquad\qquad C^b_{LUMO} = \begin{bmatrix} 0 \\ -0.02 \\ 0.02 \\ -0.50 \\ 0 \\ 0.02 \\ -0.02 \\ 0.50 \\ 0 \\ 0.02 \\ -0.02 \\ 0.50 \\ 0 \\ -0.02 \\ 0.02 \\ -0.50 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} . \qquad (16)
$$

Table 1 resumes our results for the two methods, along with the corresponding experimental data, where available.

Table 1: HOMO and LUMO eigenenergies of the cyclobutadiene molecule, obtained using LCAO with only $2p_z$ valence orbitals and the MMTS parameterization [55] (columns 2 and 3), along with the corresponding values obtained using LCAO with all valence orbitals and the MSF parameterization [27] (columns 4 and 5). Column 6 contains the available experimental HOMO value [244].

$E_{\mathrm{MMTS,H}}$ [55]	$E_{\mathrm{MSF,H}}$ [27]	$E_{\mathrm{exp,H}}$ [244]	$E_{\mathrm{MMTS,L}}$ [55]	$E_{\mathrm{MSF,L}}$ [27]	$E_{\mathrm{exp,L}}$
-7.50	-7.59	-8.24	-5.61	-5.81	-

C MEAN OVER TIME PROBABILITIES FOR CONSECUTIVE GENERATIONS OF APERIODIC POLYMERS
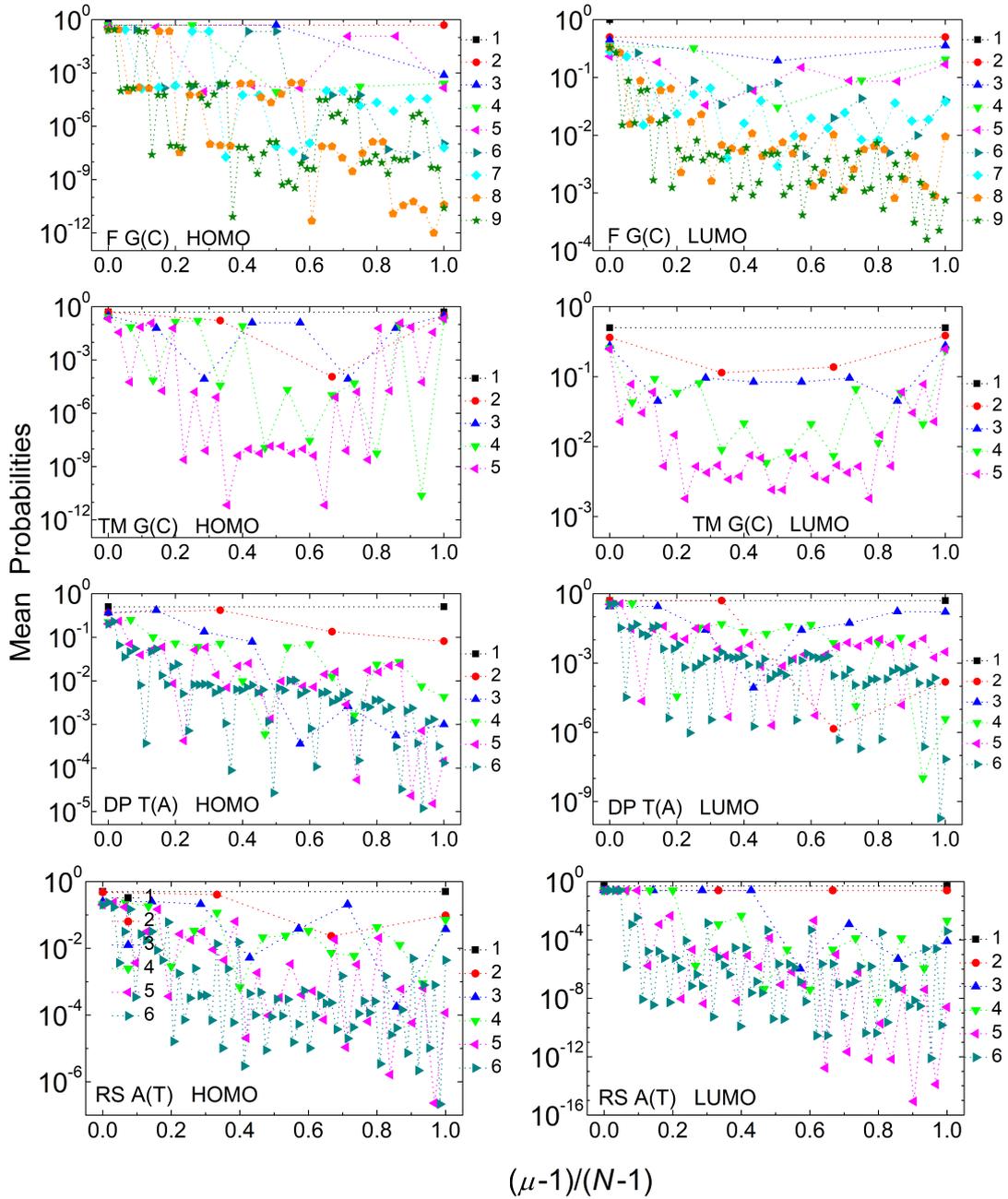
Figure 2: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for some consecutive generations, for F G(C), TM G(C), DP T(A), RS A(T) polymers (quasi-periodic), for HOMO (left) and LUMO (right).
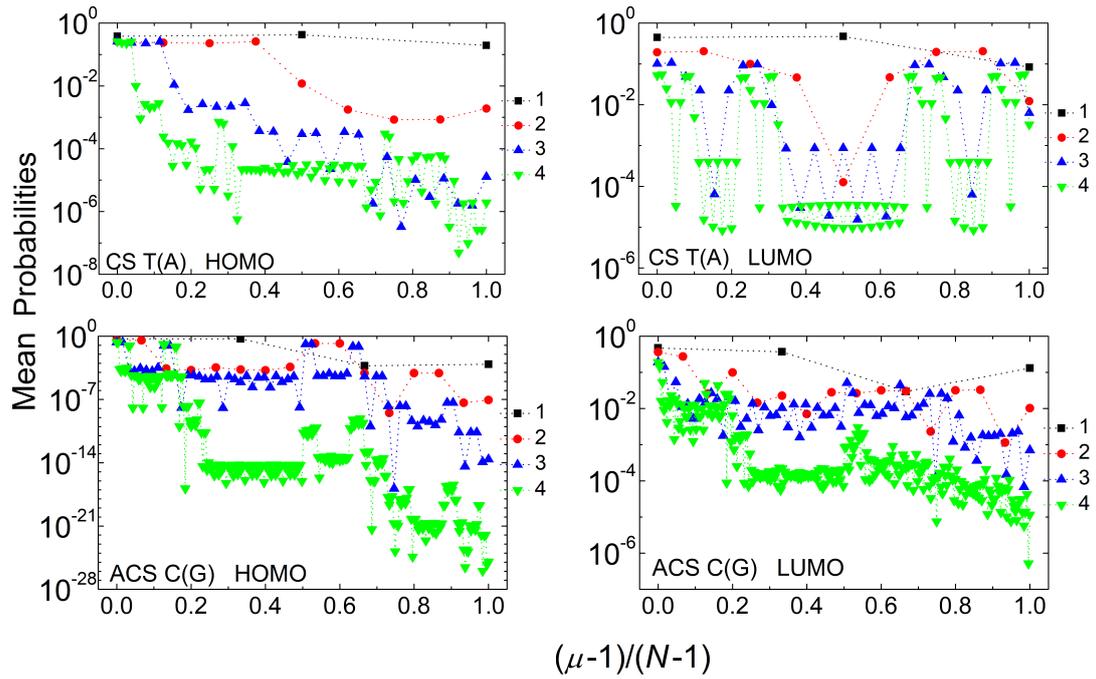
Figure 3: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for some consecutive generations, for CS T(A), ACS C(G) polymers (fractal), for HOMO (left) and LUMO (right).
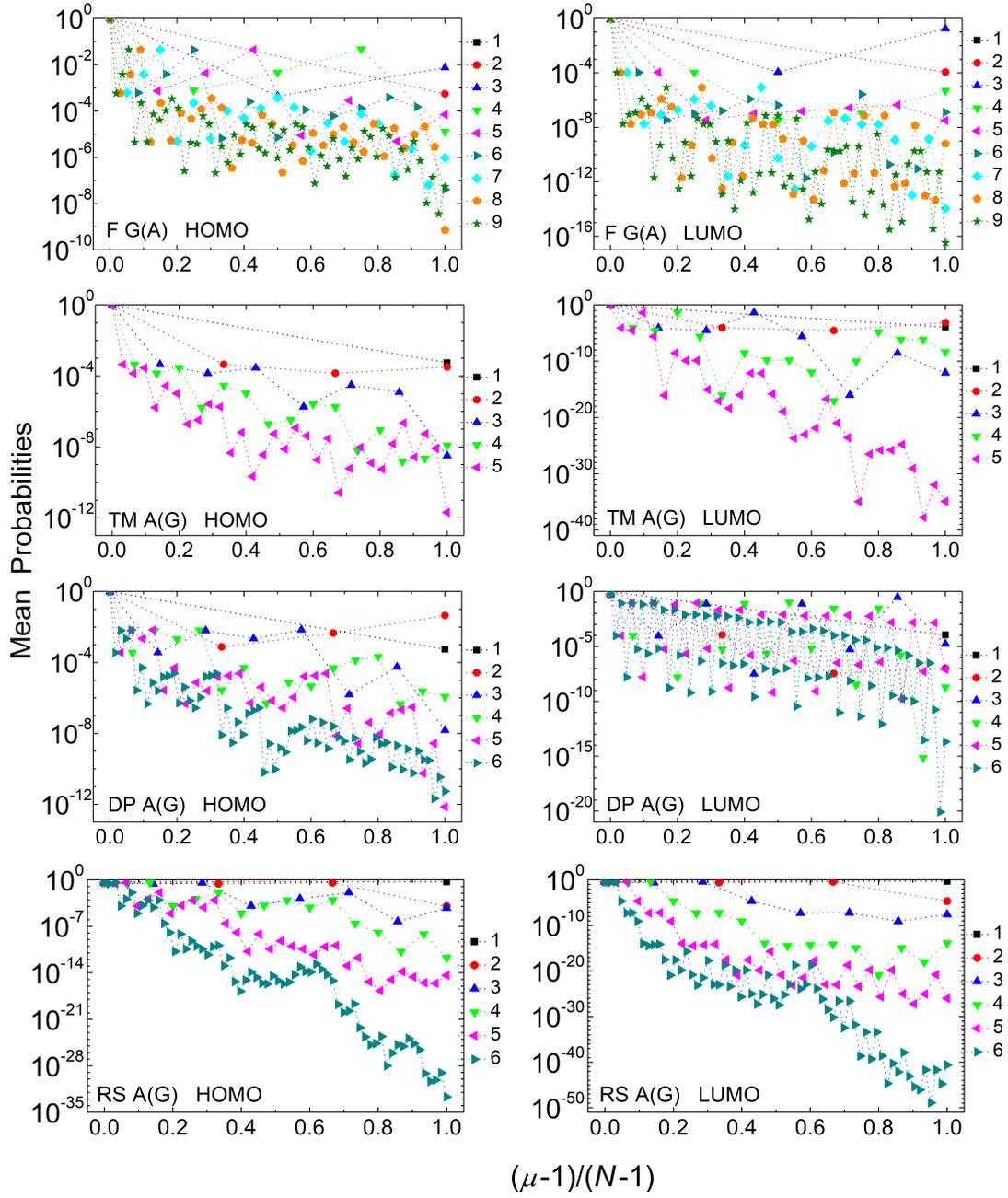
Figure 4: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for some consecutive generations, for F G(A), TM A(G), DP A(G), RS A(G), CS A(G), ACS A(G) polymers (quasi-periodic), for HOMO (left) and LUMO (right).
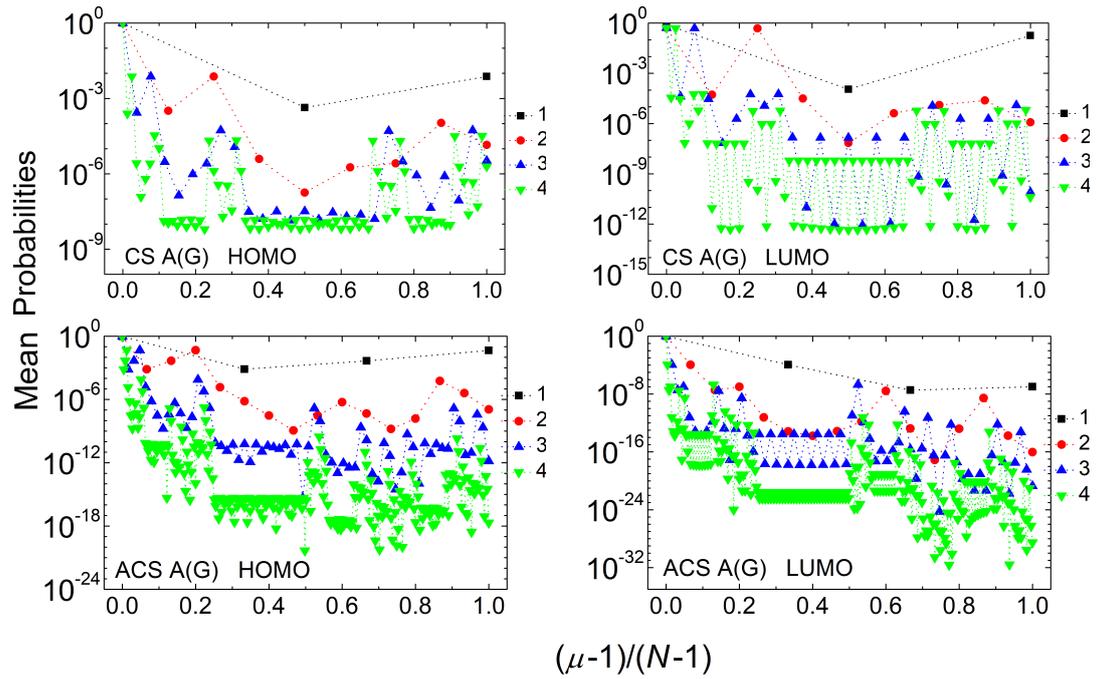
Figure 5: Mean over time probabilities to find the extra carrier at each monomer $\mu = 1, \ldots, N$, having placed it initially at the first monomer, for some consecutive generations, for CS A(G), ACS A(G) polymers (fractal), for HOMO (left) and LUMO (right).

D Mean Transfer Rates for Mutated Sequences
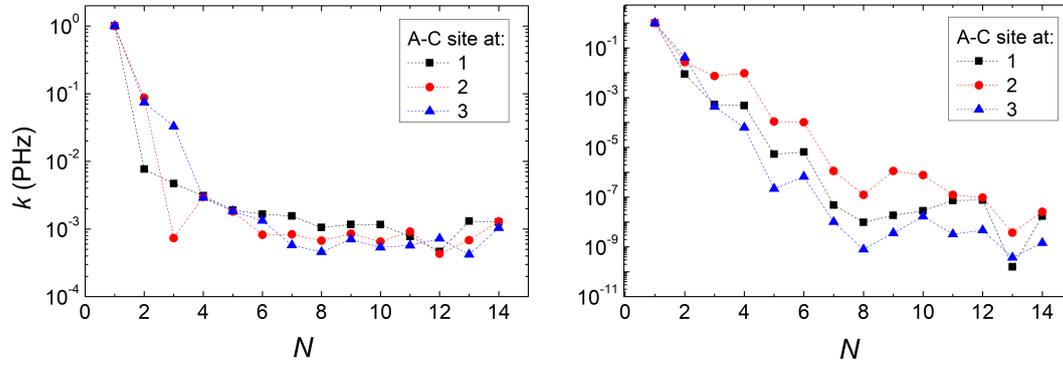
Figure 6: *Pure* mean transfer rates k of the studied G14 sequences with one A - C mismatch mutation of varying position (sites 1, 2, 3) in the sequence. Left: Ideal polymers, right: natural polymers.
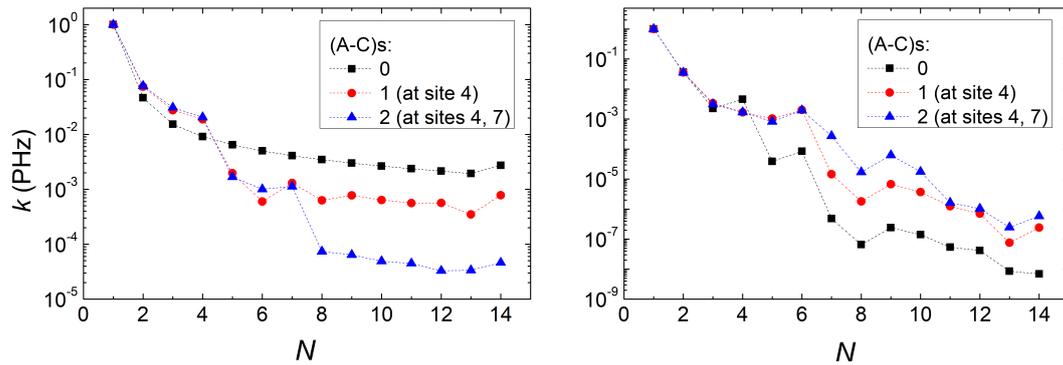
Figure 7: *Pure* mean transfer rates k of the studied G14 sequences with zero, one, and two randomly positioned A - C mismatch mutations. Left: Ideal polymers, right: natural polymers. Mutations are placed at the same sites for both ideal and natural sequences.

Bibliography

[1] Watson, J.D.; Crick, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737--738. https://doi.org/10.1038/171737a0. 1

[2] Franklin, R.E.; Gosling, R.G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171*, 740--741. https://doi.org/10.1038/171740a0. 1

[3] Roberts, R.B. Microsomal particles and protein synthesis. *Papers presented at the First Symposium of the Biophysical Society, Massachusetts Institute of Technology, Cambridge* **1958**. https://doi.org/https://doi.org/10.5962/bhl.title.6261. 1

[4] Brenner, S.; Jacob, F.; Meselson, M. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **1961**, *190*, 576--581. https://doi.org/10.1038/190576a0.

[5] Watson, J.D. Involvement of RNA in the Synthesis of Proteins. *Science* **1963**, *140*, 17--26, [https://www.science.org/doi/pdf/10.1126/science.140.3562.17]. https://doi.org/10.1126/science.140.3562.17. 1

[6] Manning, G.S. The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophysical Journal* **2006**, *91*, 3607–3616. https://doi.org/10.1529/biophysj.106.089029. 4, 47

[7] Kawai, K.; Majima, T. Hole Transfer Kinetics of DNA. *Accounts of Chemical Research* **2013**, *46*, 2616--2625. https://doi.org/10.1021/ar400079s. 4, 63, 64

[8] Dandliker, P.J.; Holmlin, R.E.; Barton, J.K. Oxidative Thymine Dimer Repair in the DNA Helix. *Science* **1997**, *275*, 1465--1468. https://doi.org/10.1126/science.275.5305.1465. 4, 25

[9] Rajski, S.R.; Jackson, B.A.; Barton, J.K. DNA repair: models for damage and mismatch recognition. *Mutation Research* **2000**, *447*, 49--72. https://doi.org/10.1016/s0027-5107(99)00195-5. 78

[10] Giese, B. Electron transfer through DNA and peptides. *Bioorganic and Medicinal Chemistry* **2006**, *14*, 6139--6143. https://doi.org/10.1016/j.bmc.2006.05.067. 4, 25

[11] Burrows, C.J.; Muller, J.G. Oxidative Nucleobase Modifications Leading to Strand Scission. *Chemical Reviews* **1998**, *98*, 1109--1152. https://doi.org/10.1021/cr960421s. 4

[12] Cadet, J. *DNA damage caused by oxidation, deamination, ultraviolet radiation and photoexcited psoralens.* Hemminki, K.; Dipple, A.; Shuker, D.E.G.; Kadlubar, F.F.; Segerback, D.; Bartsch, H. (eds.); International Agency for Research on Cancer, Scientific Publication No 125, 1994; pp. 245--276. 4

[13] Shih, C.T.; Cheng, Y.Y.; Wells, S.A.; Hsu, C.L.; Römer, R.A. Charge transport in cancer-related genes and early carcinogenesis. *Computer Physics Communications* **2011**, *182*, 36--38. https://doi.org/10.1016/j.cpc.2010.06.029. 4, 78

[14] Páez, C.J.; Schulz, P.A.; Wilson, N.R.; Römer, R.A. Robust signatures in the current–voltage characteristics of DNA molecules oriented between two graphene nanoribbon electrodes. *New Journal of Physics* **2012**, *14*, 093049. https://doi.org/10.1088/1367-2630/14/9/093049. 4, 26

[15] Oliveira, J.I.N.; Albuquerque, E.L.; Fulco, U.L.; Mauriz, P.W.; Sarmento, R.G.; Caetano, E.W.S.; Freire, V.N. Conductance of single microRNAs chains related to the autism spectrum disorder. *Europhysics Letters* **2014**, *107*, 68006. https://doi.org/10.1209/0295-5075/107/68006. 4

[16] Shih, C.T.; Roche, S.; Römer, R.A. Point-Mutation Effects on Charge-Transport Properties of the Tumor-Suppressor Gene p53. *Physical Review Letters* **2008**, *100*, 018105. https://doi.org/10.1103/PhysRevLett.100.018105. 4, 26, 78

[17] Capobianco, A.; Velardo, A.; Peluso, A. Single-Stranded DNA Oligonucleotides Retain Rise Coordinates Characteristic of Double Helices. *The Journal of Physical Chemistry B* **2018**, *122*, 7978--7989. https://doi.org/10.1021/acs.jpcb.8b04542. 4

[18] Mantela, M.; Morphis, A.; Lambropoulos, K.; Simserides, C.; Di Felice, R. Effects of Structural Dynamics on Charge Carrier Transfer in B-DNA: A Combined MD and RT-TDDFT Study. *The Journal of Physical Chemistry B* **2021**, *125*, 3986--4003. https://doi.org/10.1021/acs.jpcb.0c11489. 4, 7, 66, 67, 68, 71, 73, 77, 105, 106

[19] Zwang, T.J.; Tse, E.C.M.; Barton, J.K. Sensing DNA through DNA Charge Transport. *ACS Chemical Biology* **2018**, *13*, 1799--1809. https://doi.org/10.1021/acschembio.8b00347. 5

[20] Wohlgamuth, C.H.; McWilliams, M.A.; Slinker, J.D. DNA as a Molecular Wire: Distance and Sequence Dependence. *Analytical Chemistry* **2013**, *85*, 8634--8640. https://doi.org/10.1021/ac401229q. 5

[21] Lewis, F.D.; Wasielewski, M.R. Dynamics and efficiency of photoinduced charge transport in DNA: Toward the elusive molecular wire. *Pure and Applied Chemistry* **2013**, *85*, 1379--1387. https://doi.org/10.1351/PAC-CON-13-01-09. 5

[22] Wang, K. DNA-Based Single-Molecule Electronics: From Concept to Function. *Journal of Functional Biomaterials* **2018**, *9*, 8. https://doi.org/10.3390/jfb9010008. 5

[23] Seeman, N.C. DNA in a material world. *Nature Nanotechnology* **2003**, *421*, 427--431. https://doi.org/10.1038/nature01406. 5, 25

[24] Shapir, E.; Cohen, H.; Calzolari, A.; Cavazzoni, C.; Ryndik, D.; Cuniberti, G.; Kotlyar, A.; Di Felice, R.; Porath, D. Electronic structure of single DNA molecules resolved by transverse scanning tunnelling spectroscopy. *Nature Materials* **2008**, *7*, 68--74. https://doi.org/10.1038/nmat2060. 6, 7

[25] Genereux, J.C.; Barton, J.K. Mechanisms for DNA Charge Transport. *Chemical Reviews* **2010**, *110*, 1642--1662. https://doi.org/10.1021/cr900228f. 5

[26] Livshits, G.; Stern, A.; Rotem, D.; Borovok, N.; Eidelshtein, G.; Migliore, A.; Penzo, E.; Wind, S.; Di Felice, R.; Skourtis, S.; et al. Long-range charge transport in single G-quadruplex DNA molecules. *Nature Nanotechnology* **2014**, *9*, 1040--1046. https://doi.org/10.1038/NNANO.2014.246. 5

[27] Mantela, M.; Simserides, C.; Di Felice, R. LCAO Electronic Structure of Nucleic Acid Bases and Other Heterocycles and Transfer Integrals in B-DNA, Including Structural Variability. *Materials* **2021**, *14*. https://doi.org/10.3390/ma14174930. 6, 7, 11, 16, 17, 23, 25, 27, 40, 41, 66, 67, 70, 71, 80, 85, 88, 92, 103, 105, 111, 115

[28] Cadet, J.; Vigny, P. *Bioorganic Photochemistry, vol. 1*. Morrison, H. (ed.); Wiley & Sons: New York, 1990; pp. 1--272. 6

[29] Ruzsicska, B.P.; Lemaire, D.G.E. *CRC Handbook of Organic Photochemistry and Photobiology*. Horspool, W.M.; Francesco, L. (eds.); CRC Press: Boca Raton, FL, 1995; pp. 1289--1317. 6

[30] Abrahamson, E.; Ostroy, S.E. The photochemical and macromolecular aspects of vision. *Progress in Biophysics and Molecular Biology* **1967**, *17*, 179--215. https://doi.org/10.1016/0079-6107(67)90007-7. 6

[31] Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Essential cell biology (2nd ed.)*; Garland Science Publishing: New York, NY, 2004. 6

[32] Colson, A.O.; Besler, B.; Close, D.M.; Sevilla, M.D. Ab initio molecular orbital calculations of DNA bases and their radical ions in various protonation states: evidence for proton transfer in GC base pair radical anions. *The Journal of Physical Chemistry* **1992**, *96*, 661--668. https://doi.org/10.1021/j100181a028. 6

[33] Sevilla, M.D.; Besler, B.; Colson, A.O. Ab Initio Molecular Orbital Calculations of DNA Radical Ions. 5. Scaling of Calculated Electron Affinities and Ionization Potentials to Experimental Values. *The Journal of Physical Chemistry* **1995**, *99*, 1060--1063. https://doi.org/10.1021/j100003a032. 6

[34] Wetmore, S.D.; Boyd, R.J.; Eriksson, L.A. Electron affinities and ionization potentials of nucleotide bases. *Chemical Physics Letters* **2000**, *322*, 129--135. https://doi.org/10.1016/S0009-2614(00)00391-2. 6

[35] Walch, S.P. Model calculations of the electron affinities and ionization potentials of DNA. *Chemical Physics Letters* **2003**, *374*, 496--500. https://doi.org/10.1016/S0009-2614(03)00735-8.

[36] Mishra, D.; Pal, S. Ionization potential and structure relaxation of adenine, thymine, guanine and cytosine bases and their base pairs: A quantification of reactive sites. *Journal of Molecular Structure: THEOCHEM* **2009**, *902*, 96--102. https://doi.org/10.1016/j.theochem.2009.02.018.

[37] Sun, L.; Bu, Y. Oxidative damage to DNA: Theoretical determination of ionization potential of deoxyriboguanosine (dG)--deoxyribocytidine (dC) and proton transfer in its cation. *Journal of Molecular Structure: THEOCHEM* **2009**, *909*, 25--32. https://doi.org/10.1016/j.theochem.2009.05.022. 6

[38] Chen, E.S.; Chen, E.C. A proposed model for electron conduction in DNA based upon pairwise anion π stacking: electron affinities and ionization potentials of the hydrogen bonded base pairs. *Bioelectrochemistry and Bioenergetics* **1998**, *46*, 15--19. https://doi.org/10.1016/S0302-4598(98)00115-9. 6

[39] Gomzi, V.; Herak, J.N. Ionization potentials of nucleobase analogs using partial third-order electron propagator method. *Journal of Molecular Structure: THEOCHEM* **2004**, *683*, 155--157. https://doi.org/10.1016/j.theochem.2004.07.005. 6

[40] Schwell, M.; Hochlaf, M. *Photoinduced Phenomena in Nucleic Acids I: Nucleobases in the Gas Phase and in Solvents.* Barbatti, M.; Borin, A.C.; Ullrich, S. (eds.); Springer International Publishing: Cham, 2015; pp. 155--208. https://doi.org/10.1007/128_2014_550. 6

[41] Bravaya, K.B.; Kostko, O.; Dolgikh, S.; Landau, A.; Ahmed, M.; Krylov, A.I. Electronic Structure and Spectroscopy of Nucleic Acid Bases: Ionization Energies, Ionization-Induced Structural Changes, and Photoelectron Spectra. *The Journal of Physical Chemistry A* **2010**, *114*, 12305--12317. PMID: 21038927, https://doi.org/10.1021/jp1063726. 6

[42] Trofimov, A.B.; Schirmer, J.; Kobychev, V.B.; Potts, A.W.; Holland, D.M.P.; Karlsson, L. Photoelectron spectra of the nucleobases cytosine, thymine and adenine. *Journal of Physics B: Atomic, Molecular and Optical Physics* **2005**, *39*, 305--329. https://doi.org/10.1088/0953-4075/39/2/007. 6

[43] Matos, J.M.O.; Roos, B.O. Ab initio quantum chemical study of the .pi.-electron spectrum of the cytosine molecule. *Journal of the American Chemical Society* **1988**, *110*, 7664--7671. https://doi.org/10.1021/ja00231a014. 6, 7

[44] Fülscher, M.P.; Roos, B.O. Theoretical Study of the Electronic Spectrum of Cytosine. *Journal of the American Chemical Society* **1995**, *117*, 2089--2095. https://doi.org/10.1021/ja00112a024. 6

[45] Lorentzon, J.; Fülscher, M.P.; Roos, B.O. Theoretical Study of the Electronic Spectra of Uracil and Thymine. *Journal of the American Chemical Society* **1995**, *117*, 9265--9273. https://doi.org/10.1021/ja00141a019.

[46] Broo, A.; Holmén, A. Calculations and Characterization of the Electronic Spectra of DNA Bases Based on ab Initio MP2 Geometries of Different Tautomeric Forms. *The Journal of Physical Chemistry A* **1997**, *101*, 3589--3600. https://doi.org/10.1021/jp963928f. 7

[47] Fülscher, M.P.; Serrano-Andrés, L.; Roos, B.O. A Theoretical Study of the Electronic Spectra of Adenine and Guanine. *Journal of the American Chemical Society* **1997**, *119*, 6168--6176. https://doi.org/10.1021/ja964426i. 6

[48] Broo, A. A Theoretical Investigation of the Physical Reason for the Very Different Luminescence Properties of the Two Isomers Adenine and 2-Aminopurine. *The Journal of Physical Chemistry A* **1998**, *102*, 526--531. https://doi.org/10.1021/jp9713625. 7

[49] Jørgen, H.; Jensen, A.; Koch, H.; Jørgensen, P.; Olsen, J. Direct iterative RPA calculations. Applications to ethylene, benzene and cytosine. *Chemical Physics* **1988**, *119*, 297--306. https://doi.org/10.1016/0301-0104(88)87192-1. 7

[50] Mennucci, B.; Toniolo, A.; Tomasi, J. Theoretical Study of Guanine from Gas Phase to Aqueous Solution: Role of Tautomerism and Its Implications in Absorption and Emission Spectra. *The Journal of Physical Chemistry A* **2001**, *105*, 7126--7134. https://doi.org/10.1021/jp0111362. 7

[51] Mennucci, B.; Toniolo, A.; Tomasi, J. Theoretical Study of the Photophysics of Adenine in Solution: Tautomerism, Deactivation Mechanisms, and Comparison with the 2-Aminopurine Fluorescent Isomer. *The Journal of Physical Chemistry A* **2001**, *105*, 4749--4757. https://doi.org/10.1021/jp0045843. 7

[52] Kowalski, K.; Valiev, M. Asymptotic Extrapolation Scheme for Large-Scale Calculations with Hybrid Coupled Cluster and Molecular Dynamics Simulations. *The Journal of Physical Chemistry A* **2006**, *110*, 13106--13111. PMID: 17134172, https://doi.org/10.1021/jp064266p. 7

[53] Valiev, M.; Kowalski, K. Hybrid coupled cluster and molecular dynamics approach: Application to the excitation spectrum of cytosine in the native DNA environment. *The Journal of Chemical Physics* **2006**, *125*, 211101. https://doi.org/10.1063/1.2403847. 7

[54] Kowalski, K.; Valiev, M. Structure of Low-Lying Excited States of Guanine in DNA and Solution: Combined Molecular Mechanics and High-Level Coupled Cluster Studies. *Research Letters in Physical Chemistry* **2007**, *2007*, 085978. https://doi.org/10.1155/2007/85978.

[55] Mantela, M.; Morphis, A.; Tassi, M.; Simserides, C. Lowest ionisation and excitation energies of biologically important heterocyclic planar molecules. *Molecular Physics* **2016**, *114*, 709--718. https://doi.org/10.1080/00268976.2015.1113313. 7, 11, 17, 19, 21, 22, 40, 41, 103, 104, 109, 115

[56] Elhadj, S.; Singh, G.; Saraf, R.F. Optical Properties of an Immobilized DNA Monolayer from 255 to 700 nm. *Langmuir* **2004**, *20*, 5539--5543. PMID: 15986697, https://doi.org/10.1021/la049653+. 7

[57] Barnes, N.R.; Schreiner, A.F.; Finnegan, M.G.; Johnson, M.K. Insight into externally bound 5,10,15,20-tetrakis(2-N-methylpyridyl)porphyrinatopalladium(II), PdP(2), with B-form DNA duplexes poly(G-C)2, poly(A-T)2, and CT DNA by using combined MCD, CD, and optical data. *Biospectroscopy* **1998**, *4*, 341--352. https://doi.org/10.1002/(SICI)1520-6343(1998)4:5<341::AID-BSPY5>3.0.CO;2-E. 7

[58] Hückel, E. Quantentheoretische Beiträge zum Benzolproblem. *Zeitschrift für Physik* **1931**, *70*, 204--286. https://doi.org/10.1007/BF01339530. 7

[59] Hückel, E. Quanstentheoretische Beiträge zum Benzolproblem. *Zeitschrift für Physik* **1931**, *72*, 310--337. https://doi.org/10.1007/BF01341953.

[60] Hückel, E. Quantentheoretische Beiträge zum Problem der aromatischen und ungesättigten Verbindungen. III. *Zeitschrift für Physik* **1932**, *76*, 628--648. https://doi.org/10.1007/BF01341936.

[61] Hückel, E. Die freien Radikale der organischen Chemie. *Zeitschrift für Physik* **1933**, *83*, 632--668. https://doi.org/10.1007/BF01330865. 7

[62] Hawke, L.G.D.; Kalosakas, G.; Simserides, C. Empirical LCAO parameters for π molecular orbitals in planar organic molecules. *Molecular Physics* **2009**, *107*, 1755--1771. https://doi.org/10.1080/00268970903049089. 7, 10, 17, 19, 21, 22, 103, 104, 105

[63] Hawke, L.G.D.; Kalosakas, G.; Simserides, C. Electronic parameters for charge transfer along DNA. *European Physical Journal E* **2010**, *32*, 291--305. ibid: Erratum to: Electronic parameters for charge transfer along DNA, **34**, 118 (2011)., https://doi.org/10.1140/epje/i2010-10650-y. 7, 10, 17, 19, 21, 22, 23, 25, 40, 41, 42, 46, 103, 104, 105

[64] Slater, J.C.; Koster, G.F. Simplified LCAO Method for the Periodic Potential Problem. *Physical Review* **1954**, *94*, 1498--1524. https://doi.org/10.1103/PhysRev.94.1498. 11, 14, 80

[65] Harrison, W.A. *Electronic Structure and the Properties of Solids: the Physics of the Chemical Bond*. Dover. (ed.); Dover Publications Inc.: 2nd ed. New York, 1989. 11, 80

[66] Harrison, W.A. *Elementary Electronic Structure*; World Scientific: River Edge, NJ, 1999. https://doi.org/10.1142/5432. 11, 80

[67] Guerra, F.C.; Bickelhaupt, F.M.; Snijders, J.G.; Baerends, E.J. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *Journal of the American Chemical Society* **2000**, *122*, 4930. https://doi.org/10.1021/ja993262d. 13

[68] Endres, R.G.; Cox, D.; Singh, R.R.P. Colloquium: The quest for high-conductance DNA. *Reviews of Modern Physics* **2004**, *76*, 195--214. https://doi.org/10.1103/RevModPhys.76.195. 14, 25

[69] Menon, M.; Allen, R.E. Simulations of atomic processes at semiconductor surfaces: General method and chemisorption on GaAs(110). *Physical Review B* **1988**, *38*, 6196--6205. https://doi.org/10.1103/PhysRevB.38.6196. 16

[70] Menon, M.; Subbaswamy, K.R. Nonorthogonal tight-binding molecular-dynamics study of silicon clusters. *Physical Review B* **1993**, *47*, 12754--12759. https://doi.org/10.1103/PhysRevB.47.12754.

[71] Menon, M.; Connolly, J.; Lathiotakis, N.; Andriotis, A. Tight-binding molecular-dynamics study of transition-metal clusters. *Physical Review B* **1994**, *50*, 8903--8906. https://doi.org/10.1103/PhysRevB.50.8903. 16

[72] Hilborn, R.C. Einstein coefficients, cross sections, f values, dipole moments, and all that. *American Journal of Physics* **1982**, *50*, 982--986. https://doi.org/10.1119/1.12937. 18

[73] Hush, N.; Cheung, A.S. Ionization potentials and donor properties of nucleic acid bases and related compounds. *Chemical Physics Letters* **1975**, *34*, 11--13. https://doi.org/https://doi.org/10.1016/0009-2614(75)80190-4. 19, 20

[74] Clark, L.B.; Peschel, G.G.; Tinoco, I. Vapor Spectra and Heats of Vaporization of Some Purine and Pyrimidine Bases. *The Journal of Physical Chemistry* **1965**, *69*, 3615--3618. https://doi.org/10.1021/j100894a063. 19

[75] Voet, D.; Gratzer, W.B.; Cox, R.A.; Doty, P. Absorption spectra of nucleotides, polynucleotides, and nucleic acids in the far ultraviolet. *Biopolymers* **1963**, *1*, 193--208. https://doi.org/10.1002/bip.360010302. 19, 20

[76] Santhosh, C.; Mishra, P. Electronic spectra of 2-aminopurine and 2,6-diaminopurine: phototautomerism and fluorescence reabsorption. *Spectrochimica Acta Part A: Molecular Spectroscopy* **1991**, *47*, 1685--1693. https://doi.org/10.1016/0584-8539(91)80006-5. 19

[77] Clark, L.B.; Tinoco, I. Correlations in the Ultraviolet Spectra of the Purine and Pyrimidine Bases1. *Journal of the American Chemical Society* **1965**, *87*, 11--15. https://doi.org/10.1021/ja01079a003. 19, 20

[78] Maier, J.P.; Muller, J.F.; Kubota, T. Ionisation Energies and the Electronic Structures of the N-oxides of diazabenzenes. *Helvetica Chimica Acta* **1975**, *58*, 1634--1640. https://doi.org/10.1002/hlca.19750580618. 20

[79] Kubota, T. Electronic Spectra and Electronic Structures of Some Basic Heterocyclic N-Oxides. *Bulletin of the Chemical Society of Japan* **1962**, *35*, 946--955. https://doi.org/10.1246/bcsj.35.946. 20

[80] Gleiter, R.; Heilbronner, E.; Hornung, V. Photoelectron Spectra of Azabenzenes and Azanaphthalenes: I. Pyridine, diazines, s-triazine and s-tetrazine. *Helvetica Chimica Acta* **1972**, *55*, 255--274. https://doi.org/10.1002/hlca.19720550130. 20

[81] Pisanias, M.N.; Christophorou, L.G.; Carter, J.G.; McCorkle, D.L. Compound-negative-ion resonance states and threshold-electron excitation spectra of N-heterocyclic molecules: Pyridine, pyridazine, pyrimidine, pyrazine, and sym-triazine. *The Journal of Chemical Physics* **1973**, *58*, 2110--2124. https://doi.org/10.1063/1.1679477. 20

[82] Bolovinos, A.; Tsekeris, P.; Philis, J.; Pantos, E.; Andritsopoulos, G. Absolute vacuum ultraviolet absorption spectra of some gaseous azabenzenes. *Journal of Molecular Spectroscopy* **1984**, *103*, 240--256. https://doi.org/10.1016/0022-2852(84)90051-1. 20

[83] Halverson, F.; Hirt, R.C. Near Ultraviolet Solution Spectra of the Diazines. *The Journal of Chemical Physics* **1951**, *19*, 711--718. https://doi.org/10.1063/1.1748338. 20

[84] Ramsey, B.G. Substituent effects on imidazole basicity and photoelectron spectroscopy determined ionization energies. *The Journal of Organic Chemistry* **1979**, *44*, 2093--2097. https://doi.org/10.1021/jo01327a010. 20

[85] Caswell, D.S.; Spiro, T.G. Ultraviolet resonance Raman spectroscopy of imidazole, histidine, and Cu(imidazole)42+: implications for protein studies. *Journal of the American Chemical Society* **1986**, *108*, 6470--6477. https://doi.org/10.1021/ja00281a004. 20

[86] Lichtenberger, D.L.; Copenhaver, A.S. Ionization band profile analysis in valence photoelectron spectroscopy. *Journal of Electron Spectroscopy and Related Phenomena* **1990**, *50*, 335--352. https://doi.org/10.1016/0368-2048(90)87076-Z. 20

[87] Noyce, D.S.; Ryder, E.; Walker, B.H. The ultraviolet absorption spectra of substituted pyrazoles. *The Journal of Organic Chemistry* **1955**, *20*, 1681--1686. https://doi.org/10.1021/jo01364a013. 20

[88] Gordon, R.D.; Yang, R.F. Vapor absorption spectra of benzoxazole, benzimidazole, and benzothiazole near 2850 Å. *Canadian Journal of Chemistry* **1970**, *48*, 1722--1729. https://doi.org/10.1139/v70-283. 20

[89] Kovać, B.; Klasinc, L.; Stanovnik, B.; Tišler, M. Photoelectron spectroscopy of heterocycles. Azaindenes and azaindolizines. *Journal of Heterocyclic Chemistry* **1980**, *17*, 689--694. https://doi.org/10.1002/jhet.5570170413. 20

[90] Cané, E.; Trombetti, A.; Velino, B.; Caminati, W. Assignment of the 290-nm electronic band system of indazole [1,2-benzodiazole] as $\pi^* - \pi$ by rotational band contour analysis. *Journal of Molecular Spectroscopy* **1992**, *155*, 307--314. https://doi.org/10.1016/0022-2852(92)90519-T. 20

[91] Fuke, K.; Yoshiuchi, H.; Kaya, K.; Achiba, Y.; Sato, K.; Kimura, K. Multiphoton ionization photoelectron spectroscopy and two-color multiphoton ionization threshold spectroscopy on the hydrogen bonded phenol and 7-azaindole in a supersonic jet. *Chemical Physics Letters* **1984**, *108*, 179--184. https://doi.org/10.1016/0009-2614(84)85716-4. 20

[92] Ilich, P. 7-Azaindole: the low-temperature near-UV/vis spectra and electronic structure. *Journal of Molecular Structure* **1995**, *354*, 37--47. https://doi.org/10.1016/0022-2860(95)08856-Q. 20

[93] Pluhařová, E.; Slavíček, P.; Jungwirth, P. Modeling Photoionization of Aqueous DNA and Its Components. *Accounts of Chemical Research* **2015**, *48*, 1209--1217. https://doi.org/10.1021/ar500366z. 22

[94] Varsano, D.; Di Felice, R.; Marques, M.; Rubio, A. A TDDFT Study of the Excited States of DNA Bases and Their Assemblies. *Journal of Physical Chemistry B* **2006**, *110*, 7129--7138. https://doi.org/10.1021/jp056120g. 23, 73

[95] Mallajosyula, S.S.; Datta, A.; Pati, S.K. Structure and electronic properties of the Watson–Crick base pairs: Role of hydrogen bonding. *Synthetic Metals* **2005**, *155*, 398--401. Proceedings of the Sixth International Topical Conference on Optical Probes of Conjugated Polymers and Biosystems, Bangalore-INDIA, January 4-8th, 2005, https://doi.org/https://doi.org/10.1016/j.synthmet.2005.09.022. 23

[96] Mantela, M.; Lambropoulos, K.; Theodorakou, M.; Simserides, C. Quasi-Periodic and Fractal Polymers: Energy Structure and Carrier Transfer. *Materials* **2019**, *12*, 2177. https://doi.org/10.3390/ma12132177. 25, 27, 36, 93, 105

[97] Lambropoulos, K.; Vantaraki, C.; Bilia, P.; Mantela, M.; Simserides, C. Periodic polymers with increasing repetition unit: Energy structure and carrier transfer. *Physical Review E* **2018**, *98*, 032412. https://doi.org/10.1103/PhysRevE.98.032412. 25, 26, 27, 35, 42, 45, 46, 55, 62, 104, 105

[98] Dekker, C.; Ratner, M. Electronic properties of DNA. *Physics World* **2001**, *14*, 29. https://doi.org/10.1088/2058-7058/14/8/33. 25

[99] Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U.; Braun, E. Sequence-Specific Molecular Lithography on Single DNA Molecules. *Science* **2002**, *297*, 72--75. https://doi.org/10.1126/science.1071247.

[100] Kannan, A.M.; Renugopalakrishnan, V.; Filipek, S.; Li, P.; Audette, G.F.; Munukutla, L. Bio-Batteries and Bio-Fuel Cells: Leveraging on Electronic Charge Transfer Proteins. *J. Nanosci. Nanotechnol.* **2009**, *9*, 1665--1678. https://doi.org/10.1166/jnn.2009.si03.

[101] Moser, C.C.; Ross Anderson, J.L.; Dutton, P.L. Guidelines for tunneling in enzymes. *Biochim. Biophys. Acta* **2010**, *1797*, 1573--1586. https://doi.org/10.1016/j.bbabio.2010.04.441.

[102] Gray, H.B.; Winkler, J.R. Electron flow through metalloproteins. *Biochim. Biophys. Acta* **2010**, *1797*, 1563--1572. https://doi.org/10.1016/j.bbabio.2010.05.001.

[103] Artés, J.M.; López-Martínez, M.; Díez-Pérez, I.; Sanz, F.; Gorostiza, P. Nanoscale charge transfer in redox proteins and DNA: Towards biomolecular electronics. *Electrochim. Acta* **2014**, *140*, 83--95. https://doi.org/10.1016/j.electacta.2014.05.089. 25

[104] Maciá, E.; Triozon, F.; Roche, S. Contact-dependent effects and tunneling currents in DNA molecules. *Physical Review B* **2005**, *71*, 113106. https://doi.org/10.1103/PhysRevB.71.113106. 25, 26

[105] Rawtani, D.; Kuntmal, B.; Agrawal, Y. Charge transfer in DNA and its diverse modelling approaches. *Frontiers in Life Science* **2016**, *9*, 214--225. https://doi.org/10.1080/21553769.2016.1207570. 25

[106] Lambropoulos, K.; Chatzieleftheriou, M.; Morphis, A.; Kaklamanis, K.; Lopp, R.; Theodorakou, M.; Tassi, M.; Simserides, C. Electronic structure and carrier transfer in B-DNA monomer polymers and dimer polymers: Stationary and time-dependent aspects of a wire model versus an extended ladder model. *Physical Review E* **2016**, *94*, 062403. https://doi.org/10.1103/PhysRevE.94.062403. 25, 26, 31, 86, 105, 107

[107] Arkin, M.R.; Stemp, E.D.A.; Holmlin, R.E.; Barton, J.K.; Hörmann, A.; Olson, E.J.C.; Barbara, P.F. Rates of DNA-Mediated Electron Transfer Between Metallointercalators. *Science* **1996**, *273*, 475--480, [https://www.science.org/doi/pdf/10.1126/science.273.5274.475]. https://doi.org/10.1126/science.273.5274.475. 25

[108] Hall, D.B.; Holmlin, R.E.; Barton, J.K. Oxidative DNA damage through long-range electron transfer. *Nature* **1996**, *382*, 731--735. https://doi.org/10.1038/382731a0. 25

[109] Fink, H.W.; Schönenberger, C. Electrical conduction through DNA molecules. *Nature* **1999**, *398*, 407--410. https://doi.org/10.1038/18855. 25

[110] Porath, D.; Bezryadin, A.; de Vries, S.; Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **2000**, *403*, 635--638. https://doi.org/10.1038/35001029.

[111] Yoo, K.H.; Ha, D.H.; Lee, J.O.; Park, J.W.; Kim, J.; Kim, J.J.; Lee, H.Y.; Kawai, T.; Choi, H.Y. Electrical Conduction through Poly(dA)-Poly(dT) and Poly(dG)-Poly(dC) DNA Molecules. *Physical Review Letters* **2001**, *87*, 198102.

[112] Xu, B.; Zhang, P.; Li, X.; Tao, N. Direct Conductance Measurement of Single DNA Molecules in Aqueous Solution. *Nano Lett.* **2004**, *4*, 1105--1108. https://doi.org/10.1021/nl0494295.

[113] H. Cohen and C. Nogues and R. Naaman and D. Porath. Direct measurement of electrical transport through single DNA molecules of complex sequence. *Proceedings of the National Academy of Sciences* **2005**, *102*, 11589--11593. https://doi.org/10.1073/pnas.0505272102.

[114] Grozema, F.C.; Berlin, Y.A.; Siebbeles, L.D.A. Sequence-dependent charge transfer in donor-DNA-acceptor systems: A theoretical study. *International Journal of Quantum Chemistry* **1999**, *75*, 1009--1016. https://doi.org/10.1002/(sici)1097-461x(1999)75:6<1009::aid-qua5>3.0.co;2-a. 25

[115] Simserides, C. A systematic study of electron or hole transfer along DNA dimers, trimers and polymers. *Chemical Physics* **2014**, *440*, 31 -- 41. https://doi.org/10.1016/j.chemphys.2014.05.024. 26, 34, 40, 41, 42, 63, 64, 75, 76, 104

[116] Lambropoulos, K.; Chatzieleftheriou, M.; Morphis, A.; Kaklamanis, K.; Theodorakou, M.; Simserides, C. Unbiased charge oscillations in B-DNA: Monomer polymers and dimer polymers. *Physical Review E* **2015**, *92*, 032725. https://doi.org/10.1103/PhysRevE.92.032725. 26, 33, 45, 86, 105

[117] Cuniberti, G.; Craco, L.; Porath, D.; Dekker, C. Backbone-induced semiconducting behavior in short DNA wires. *Physical Review B* **2002**, *65*, 241314. https://doi.org/10.1103/PhysRevB.65.241314.

[118] Roche, S.; Bicout, D.; Maciá, E.; Kats, E. Long Range Correlations in DNA: Scaling Properties and Charge Transfer Efficiency. *Physical Review Letters* **2003**, *91*, 228101. https://doi.org/10.1103/PhysRevLett.91.228101. 26

[119] Roche, S. Sequence Dependent DNA-Mediated Conduction. *Physical Review Letters* **2003**, *91*, 108101. https://doi.org/10.1103/PhysRevLett.91.108101. 26

[120] Palmero, F.; Archilla, J.F.R.; Hennig, D.; Romero, F.R. Effect of base-pair inhomogeneities on charge transport along the DNA molecule, mediated by twist and radial polarons. *New Journal of Physics* **2004**, *6*, 13. https://doi.org/10.1088/1367-2630/6/1/013.

[121] Yamada, H. Localization of electronic states in chain models based on real DNA sequence. *Physics Letters A* **2004**, *332*, 65 -- 73. https://doi.org/https://doi.org/10.1016/j.physleta.2004.09.041.

[122] Apalkov, V.M.; Chakraborty, T. Electron dynamics in a DNA molecule. *Physical Review B* **2005**, *71*, 033102. https://doi.org/10.1103/PhysRevB.71.033102.

[123] Klotsa, D.; Römer, R.A.; Turner, M.S. Electronic Transport in DNA. *Biophysical Journal* **2005**, *89*, 2187 -- 2198. https://doi.org/https://doi.org/10.1529/biophysj.105.064014.

[124] Joe, Y.S.; Lee, S.H.; Hedin, E.R. Electron transport through asymmetric DNA molecules. *Physics Letters A* **2010**, *374*, 2367 -- 2373. https://doi.org/https://doi.org/10.1016/j.physleta.2010.03.050.

[125] Yi, J. Conduction of DNA molecules: A charge-ladder model. *Physical Review B* **2003**, *68*, 193103. https://doi.org/10.1103/PhysRevB.68.193103.

[126] Caetano, R.A.; Schulz, P.A. Sequencing-Independent Delocalization in a DNA-Like Double Chain with Base Pairing. *Physical Review Letters* **2005**, *95*, 126601. https://doi.org/10.1103/PhysRevLett.95.126601.

[127] Wang, X.F.; Chakraborty, T. Charge Transfer via a Two-Strand Superexchange Bridge in DNA. *Physical Review Letters* **2006**, *97*, 106602. https://doi.org/10.1103/PhysRevLett.97.106602.

[128] Albuquerque, E.L.; Fulco, U.L.; Freire, V.N.; Caetano, E.W.S.; Lyra, M.L.; de Moura, F.A.B.F. DNA-based nanobiostructured devices: The role of quasiperiodicity and correlation effects. *Physics Reports* **2014**, *535*, 139 -- 209. https://doi.org/https://doi.org/10.1016/j.physrep.2013.10.004.

[129] Sarmento, R.G.; Mendes, G.A.; Albuquerque, E.L.; Fulco, U.L.; Vasconcelos, M.S.; Ujsághy, O.; Freire, V.N.; Caetano, E.W.S. The DNA electronic specific heat at low temperature: The role of aperiodicity. *Physics Letters A* **2012**, *376*, 2413 -- 2417. https://doi.org/https://doi.org/10.1016/j.physleta.2012.05.058.

[130] Sarmento, R.G.; Albuquerque, E.L.; Sesion, P.D.; Fulco, U.L.; de Oliveira, B.P.W. Electronic transport in double-strand poly(dG)–poly(dC) DNA segments. *Physics Letters A* **2009**, *373*, 1486 -- 1491. https://doi.org/https://doi.org/10.1016/j.physleta.2009.02.043. 26

[131] Albuquerque, E.L.; Vasconcelos, M.S.; Lyra, M.L.; de Moura, F.A.B.F. Nucleotide correlations and electronic transport of DNA sequences. *Physical Review E* **2005**, *71*, 021910. https://doi.org/10.1103/PhysRevE.71.021910.

[132] Cuniberti, G.; Maciá, E.; Rodríguez, A.; Römer, R.A. *Tight-Binding Modeling of Charge Migration in DNA Devices*. Chakraborty, T. (ed.); Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp. 1--20. https://doi.org/10.1007/978-3-540-72494-0_1. 26

[133] Ye, Y.J.; Shen, L.L. DFT approach to calculate electronic transfer through a segment of DNA double helix. *Journal of Computational Chemistry* **2000**, *21*, 1109--1117. https://doi.org/10.1002/1096-987X(200009)21:12<1109::AID-JCC7>3.0.CO;2-4. 26

[134] Ye, Y.J.; Jiang, Y. Electronic structures and long-range electron transfer through DNA molecules. *International Journal of Quantum Chemistry* **2000**, *78*, 112--130. https://doi.org/10.1002/(SICI)1097-461X(2000)78:2<112::AID-QUA5>3.0.CO;2-5.

[135] Barnett, R.N.; Cleveland, C.L.; Landman, U.; Boone, E.; Kanvah, S.; Schuster, G.B. Effect of base sequence and hydration on the electronic and hole transport properties of duplex DNA: Theory and Experiment. *The Journal of Physical Chemistry A* **2003**, *107*, 3525--3537. https://doi.org/10.1021/jp022211r.

[136] Artacho, E.; Machado, M.; Sánchez-Portal, D.; Ordejón, P.; Soler, J.M. Electrons in dry DNA from density functional calculations. *Molecular Physics* **2003**, *101*, 1587--1594. https://doi.org/10.1080/0026897031000068587.

[137] Adessi, C.; Walch, S.; Anantram, M.P. Environment and structure influence on DNA conduction. *Physical Review B* **2003**, *67*, 081405. https://doi.org/10.1103/PhysRevB.67.081405.

[138] Mehrez, H.; Anantram, M.P. Interbase electronic coupling for transport through DNA. *Physical Review B* **2005**, *71*, 115405. https://doi.org/10.1103/PhysRevB.71.115405.

[139] Voityuk, A.A. Electronic couplings and on-site energies for hole transfer in DNA: Systematic quantum mechanical/molecular dynamic study. *The Journal of Chemical Physics* **2008**, *128*, 115101. https://doi.org/10.1063/1.2841421.

[140] Kubař, T.; Woiczikowski, P.B.; Cuniberti, G.; Elstner, M. Efficient Calculation of Charge-Transfer Matrix Elements for Hole Transfer in DNA. *The Journal of Physical Chemistry B* **2008**, *112*, 7937--7947. https://doi.org/10.1021/jp801486d.

[141] Tassi, M.; Morphis, A.; Lambropoulos, K.; Simserides, C. RT-TDDFT study of hole oscillations in B-DNA monomers and dimers. *Cogent Physics* **2017**, *4*, 1361077. https://doi.org/10.1080/23311940.2017.1361077. 26

[142] Kawai, K.; Majima, T. *Increasing the hole transfer rate through DNA by chemical modification.* Akasaka, T.; Fukuzumi, A.O.S.; Kandori, H.; Aso, Y. (eds.); Springer: Tokyo, 2015. https://doi.org/https://doi.org/10.1007/978-4-431-55357-1_44. 26

[143] Lambropoulos, K.; Kaklamanis, K.; Morphis, A.; Tassi, M.; Lopp, R.; Georgiadis, G.; Theodorakou, M.; Chatzieleftheriou, M.; Simserides, C. Wire and extended ladder model predict THz oscillations in DNA monomers, dimers and trimers. *Journal of Physics: Condensed Matter* **2016**, *28*, 495101. https://doi.org/10.1088/0953-8984/28/49/495101/. 26

[144] Gutiérez, R.; Caetano, R.; Woiczikowski, P.B.; Kubař, T.; Elstner, M.; Cuniberti, G. Structural fluctuations and quantum transport through DNA molecular wires: a combined molecular dynamics and model Hamiltonian approach. *New Journal of Physics* **2010**, *12*, 023022. https://doi.org/10.1088/1367-2630/12/2/023022. 26

[145] Koslowski, T.; Jurjiu, A.; Blumen, A. Polaron Formation and Hopping Conduction in Hyperbranched Polymers: A Theoretical Approach. *The Journal of Physical Chemistry B* **2004**, *108*, 3283--3288. https://doi.org/10.1021/jp037263a. 26

[146] Koslowski, T.; Jurjiu, A.; Blumen, A. Models of Irregular Hyperbranched Polymers: Topological Disorder and Mechanical Response. *Macromol. Theory Simul.* **2006**, *15*, 538--545. https://doi.org/10.1002/mats.200600004.

[147] Jurjiu, A.; Turcu, F.; Galiceanu, M. Dynamics of a Complex Multilayer Polymer Network: Mechanical Relaxation and Energy Transfer. *Polymers* **2018**, *10*, 164. https://doi.org/10.3390/polym10020164. 26

[148] Maciá, E. Electronic structure and transport properties of double-stranded Fibonacci DNA. *Physical Review B* **2006**, *74*, 245105. https://doi.org/10.1103/PhysRevB.74.245105. 26

[149] Kundu, S.; Karmakar, S.N. Electronic specific heat of DNA: Effects of backbones and disorder. *Physics Letters A* **2015**, *379*, 1377 -- 1383. https://doi.org/https://doi.org/10.1016/j.physleta.2015.02.036.

[150] Fathizadeh, S.; Behnia, S.; Ziaei, J. Engineering DNA Molecule Bridge between Metal Electrodes for High-Performance Molecular Transistor: An Environmental Dependent Approach. *The Journal of Physical Chemistry B* **2018**, *122*, 2487--2494. https://doi.org/10.1021/acs.jpcb.7b10034. 26

[151] Sigler, L. *Fibonacci's Liber Abaci: A Translation into Modern English of Leonardo Pisano's Book of Calculation*; Springer Verlag: New York, 2003. 36

[152] Singh, P. The so-called fibonacci numbers in ancient and medieval India. *Hist. Math.* **1985**, *12*, 229 -- 244. https://doi.org/10.1016/0315-0860(85)90021-7. 36

[153] Prouhet, E. Mémoire sur les relations entre les puissances des nombres. *Comptes rendus de l' Académie des Sciences Paris* **1851**, *33*. (in French). 36

[154] Thue, A. *Selected mathematical papers of Axel Thue.* Nagell, T.; Selberg, A.; Selberg, S.; Thalberg, K. (eds.); Universitetsforlaget: Oslo, 1977. 36

[155] Morse, H.M. Recurrent Geodesics on a Surface of Negative Curvature. *Trans. Amer. Math. Soc.* **1921**, *22*, 84--100. https://doi.org/10.2307/1988844. 36

[156] Rahimi, H. Analysis of photonic spectra in Thue-Morse, double-period and Rudin-Shapiro quasiregular structures made of high temperature superconductors in visible range. *Opt. Mater.* **2016**, *57*, 264 -- 271. https://doi.org/10.1016/j.optmat.2016.04.022. 36

[157] Janot, C. *Quasicrystals: A Primer*; Clarendon Press, 1955. 37

[158] Brillhart, J.; Morton, P. A Case Study in Mathematical Research: The Golay-Rudin-Shapiro Sequence. *The American Mathematical Monthly* **1996**, *103*, 854--869. https://doi.org/10.1080/00029890.1996.12004830. 37

[159] Shapiro, H.S. Extremal problems for polynomials and power series. Master's thesis, Massachusetts Institute of Technology, 1951.

[160] Rudin, W. Some theorems on Fourier coefficients. *Proceedings of the American Mathematical Society* **1959**, *10*, 855--855. https://doi.org/10.1090/s0002-9939-1959-0116184-5. 37

[161] Cantor, G. Über unendliche, lineare Punktmannigfaltigkeiten. *Mathematische Annalen* **1883**, *21*, 545. (in German). 37

[162] Marques, M.A.L.; Ullrich, C.A.; Nogueira, F.; Rubio, A.; Burke, K.; Gross, E.K.U. *Time-Dependent Density Functional Theory.* Marques, M.A.L.; Ullrich, C.A.; Nogueira, F.; Rubio, A.; Burke, K.; Gross, E.K.U. (eds.); Springer Berlin Heidelberg, 2006. https://doi.org/10.1007/b11767107. 47

[163] Ralha, R. Perturbation Splitting for More Accurate Eigenvalues. *SIAM Journal on Matrix Analysis and Applications* **2009**, *31*, 75--91. https://doi.org/10.1137/070687049. 47

[164] Meggers, E.; Michel-Beyerle, M.E.; Giese, B. Sequence Dependent Long Range Hole Transport in DNA. *Journal of the American Chemical Society* **1998**, *120*, 12950--12955. https://doi.org/10.1021/ja983092p. 63

[165] Giese, B.; Amaudrut, J.; Köhler, A.K.; Spormann, M.; Wessely, S. Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling. *Nature* **2001**, *412*, 318--320. https://doi.org/10.1038/35085542. 63

[166] Lewis, F.D.; Wu, T.; Zhang, Y.; Letsinger, R.L.; Greenfield, S.R.; Wasielewski, M.R. Distance-Dependent Electron Transfer in DNA Hairpins. *Science* **1997**, *277*, 673--676. https://doi.org/10.1126/science.277.5326.673. 63

[167] Wan, C.; Fiebig, T.; Schiemann, O.; Barton, J.K.; Zewail, A.H. Femtosecond direct observation of charge transfer between bases in DNA. *Proceedings of the National Academy of Sciences USA* **2000**, *97*, 14052--14055. https://doi.org/10.1073/pnas.250483297. 63

[168] Takada, T.; Kawai, K.; Fujitsuka, M.; Majima, T. Direct observation of hole transfer through double-helical DNA over 100 A. *Proceedings of the National Academy of Sciences USA* **2004**, *101*, 14002--14006. https://doi.org/10.1073/pnas.0402756101. 64

[169] Mickley Conron, S.M.; Thazhathveetil, A.K.; Wasielewski, M.R.; Burin, A.L.; Lewis, F.D. Direct Measurement of the Dynamics of Hole Hopping in Extended DNA G-Tracts. An Unbiased Random Walk. *Journal of the American Chemical Society* **2010**, *132*, 14388--14390. https://doi.org/10.1021/ja106991f. 64, 65

[170] Vura-Weis, J.; Wasielewski, M.R.; Thazhathveetil, A.K.; Lewis, F.D. Efficient Charge Transport in DNA Diblock Oligomers. *Journal of the American Chemical Society* **2009**, *131*, 9722--9727. https://doi.org/10.1021/ja9015217. 65

[171] Dickerson, R.E.; Bansal, M.; Calladine, C.R.; Diekmann, S.; Hunter, W.N.; Kennard, O.; von Kitzing, E.; Lavery, R.; Nelson, H.C.M.; Olson, W.K.; et al. Definitions and nomenclature of nucleic acid structure parameters. *The EMBO Journal* **1989**, *8*, 1--4. https://doi.org/10.1002/j.1460-2075.1989.tb03339.x. 66

[172] Dickerson, R. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Research* **1989**, *17*, 1797--1803. https://doi.org/10.1093/nar/17.5.1797. 66

[173] Perez, A.; Luque, F.; Orozco, M. Frontiers in Molecular Dynamics Simulations of DNA. *Accounts of Chemical Research* **2012**, *45*, 196--205. https://doi.org/10.1021/ar2001217. 68

[174] Karplus, M.; McCammon, J. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **2002**, *9*, 646 -- 652. https://doi.org/10.1038/nsb0902-646.

[175] Hollingsworth, S.; Dror, R. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 646--652. https://doi.org/10.1016/j.neuron.2018.08.011.

[176] Dror, R.; Dirks, R.; Grossman, J.; Xu, H.; Shaw, D. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Reviews of Biophysics* **2012**, *41*, 429 -- 452. https://doi.org/10.1146/annurev-biophys-042910-155245. 68

[177] Perilla, J.; Goh, B.; Cassidy, K.; Liu, B.; Bernardi, R.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology* **2015**, *31*, 64--74. https://doi.org/10.1016/j.sbi.2015.03.007. 68

[178] Ray, A.; Di Felice, R. Protein-Mutation-Induced Conformational Changes of the DNA and Nuclease Domain in CRISPR/Cas9 Systems by Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **2020**, *124*, 2168--2179. https://doi.org/10.1021/acs.jpcb.9b07722. 68

[179] Ali, A.; Vijayan, R. Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Scientific Reports* **2020**, *10*, 14214. https://doi.org/10.1038/s41598-020-71188-3. 68

[180] Beveridge, D.; Barreiro, G.; Suzie Byun, K.; Case, D.; Cheatham, T.; Dixit, S.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J.; et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA Oligonucleotides. I. research design and results on d(CpG) steps. *Biophysical Journal* **2004**, *87*, 3799 -- 3813. https://doi.org/10.1529/biophysj.104.045252. 68

[181] Dixit, S.; Beveridge, D.; Case, D.; Cheatham, T.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J.; Osman, R.; Sklenar, H.; et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA Oligonucleotides. II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophysical Journal* **2005**, *89*, 3721--3740. https://doi.org/10.1529/biophysj.105.067397.

[182] Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T.; Case, D.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Research* **2010**, *38*, 299--313. https://doi.org/10.1093/nar/gkp834.

[183] Pasi, M.; Maddocks, J.; Beveridge, D.; Bishop, T.; Case, D.; Cheatham, T.; Dans, P.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. μABC: A systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research* **2014**, *42*, 12272--12283. https://doi.org/10.1093/nar/gku855. 68

[184] Dantas Machado, A.; Cooper, B.; Lei, X.; Di Felice, R.; Chen, L.; Rohs, R. Landscape of DNA binding signatures of myocyte enhancer factor-2B reveals a unique interplay of base and shape readout. *Nucleic Acids Research* **2020**, *48*, 8529--8544. https://doi.org/10.1093/nar/gkaa642. 68

[185] Hancock, S.; Ghane, T.; Cascio, D.; Rohs, R.; Di Felice, R.; Johnson, R. Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Research* **2013**, *41*, 6750--6760. https://doi.org/10.1093/nar/gkt357. 68

[186] Jorgensen, W.; Chandrasekhar, J.; Madura, J.D. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983**, *79*, 926 -- 935. https://doi.org/10.1063/1.445869. 68

[187] Ghane, T.; Brancolini, G.; Varsano, D.; Di Felice, R. Optical Properties of Triplex DNA from Time-Dependent Density Functional Theory. *The Journal of Physical Chemistry B* **2012**, *116*, 10693--10702. https://doi.org/10.1021/jp304818s. 68

[188] Abraham, M.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J.; Hess, B.; Lindhal, 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. 68

[189] Ivani, I.; Dans, P.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; et al. PARMBSC1: A REFINED FORCE-FIELD FOR DNA SIMULATIONS. *Nature Methods* **2016**, *13*, 55 -- 58. https://doi.org/10.1038/nmeth.3658. 68

[190] Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.; Mark, A. Peptide folding: When simulation meets experiment. *Angewandte Chemie International Edition* **1999**, *38*, 236--240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M. 68

[191] Qi, J.; Edirisinghe, N.; Rabbani, M.G.; Anantram, M.P. Unified model for conductance through DNA with the Landauer-Büttiker formalism. *Physical Review B* **2013**, *87*, 085404. https://doi.org/10.1103/PhysRevB.87.085404. 68

[192] Zhuravel, R.; Huang, H.; Polycarpou, G.; Polydorides, S.; Motamarri, P.; Katrivas, L.; Rotem, D.; Sperling, J.; Zotti, L.A.; Kotlyar, A.B.; et al. Backbone charge transport in double-stranded DNA. *Nature Nanotechnology* **2020**, *15*, 836--840. https://doi.org/10.1038/s41565-020-0741-2. 68

[193] Olson, W.K.; Bansal, M.; Burley, S.K.; Dickerson, R.E.; Gerstein, M.; Harvey, S.C.; Heinemann, U.; Lu, X.J.; Neidle, S.; Shakked, Z.; et al. A Standard Reference Frame for the Description of Nucleic Acid Base-pair Geometry. *Journal of Molecular Biology* **2001**, *313*, 229--237. https://doi.org/10.1006/jmbi.2001.4987. 70

[194] Lavery, R.; Moakher, M.; Maddocks, J.H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research* **2009**, *37*, 5917--5929. https://doi.org/10.1093/nar/gkp608. 70

[195] Mazur, J.; Jernigan, R.L. Comparison of Rotation Models for Describing DNA Conformations: Application to Static and Polymorphic Forms. *Biophysical Journal* **1995**, *68*, 1472--1489. https://doi.org/10.1016/S0006-3495(95)80320-6. 70

[196] Ussery, D. DNA Structure: A-, B- and Z-DNA Helix Families. *ENCYCLOPEDIA OF LIFE SCIENCES* **2002**. 70

[197] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical Review* **1964**, *136*, B864. https://doi.org/10.1103/PhysRev.136.B864. 73

[198] Kohn, W.; Sham, L.J. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, *140*, A1133. `https://doi.org/10.1103/PhysRev.140.A1133`.

[199] Pribram-Jones, A.; Gross, D.; Burke, K. DFT: a theory full of holes? *Annual Reviews of Physical Chemistry* **2015**, *66*, 283--304. `https://doi.org/10.1146/annurev-physchem-040214-121420`. 73

[200] Runge, E.; Gross, E.K.U. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **1984**, *52*, 997. `https://doi.org/10.1103/PhysRevLett.52.997`. 73

[201] Casida, M.; Huix-Rotllant, M. Progress in Time-Dependent Density-Functional Theory. *Annual Reviews of Physical Chemistry* **2012**, *63*, 287 -- 323. `https://doi.org/10.1146/annurev-physchem-032511-143803`.

[202] Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Reviews of Modern Physics* **2002**, *74*, 601 -- 659. `https://doi.org/10.1103/RevModPhys.74.601`. 73

[203] Marques, M.A.L.; Castro, A.; Bertsch, G.F.; Rubio, A. octopus: a first-principles tool for excited electron-ion dynamics. *Computer Physics Communications* **2003**, *151*, 60--78. `https://doi.org/10.1016/S0010-4655(02)00686-0`. 73

[204] Lopata, K.; Govind, N. Modeling Fast Electron Dynamics with Real-Time Time-Dependent Density Functional Theory: Application to Small Molecules and Chromophores. *Journal of Chemical Theory and Computation* **2011**, *7*, 1344. `https://doi.org/10.1021/ct200137z`. 73

[205] Valiev, M.; Bylaska, E.J.; Govind, N.; Kowalski, K.; Straatsma, T.P.; Van Dam, H.J.J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T.L.; et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* **2010**, *181*, 1477. `https://doi.org/10.1016/j.cpc.2010.04.018`. 73

[206] Yanai, T.; Tew, D.P.; Handy, N. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004**, *393*, 51--57. `https://doi.org/10.1016/j.cplett.2004.06.011`. 73

[207] Binkley, J.S.; Pople, J.A.; Hehre, W.J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *Journal of the American Chemical Society* **1980**, *102*, 939--947. `https://doi.org/10.1021/ja00523a008`. 73

[208] Clark, T.; Chandrasekhar, J.; Spitznagel, G.W.; Von Ragué Schleyer, P. Efficient diffuse function-augmented basis sets for anion calculations. III.

The 3-21+G basis set for first-row elements, Li-F. *Journal of Computational Chemistry* **1983**, *4*, 294--301. https://doi.org/10.1002/jcc.540040303. 73

[209] Hehre, W.J.; Ditchfield, R.; Pople, J.A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **1972**, *56*, 2257--2261. https://doi.org/10.1063/1.1677527.

[210] Hariharan, P.C.; Pople, J.A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica Chimica Acta* **1973**, *28*, 213--222. https://doi.org/10.1007/bf00533485.

[211] Ditchfield, R.; Hehre, W.J.; Pople, J.A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **1971**, *54*, 724--728. https://doi.org/10.1063/1.1674902. 73

[212] Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007--1023. https://doi.org/10.1063/1.456153. 73

[213] Kendall, R.A.; Dunning, T.H.; Harrison, R.J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *Journal of Chemical Physics* **1992**, *96*, 6796--6806. https://doi.org/10.1063/1.462569. 73

[214] Löwdin, P.O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics* **1950**, *18*, 365--375. https://doi.org/10.1063/1.1747632. 74

[215] Lambropoulos, K.; Kaklamanis, K.; Georgiadis, G.; Simserides, C. THz and above THz electron or hole oscillations in DNA dimers and trimers. *Annalen der Physik (Berlin)* **2014**, *526*, 249--258. https://doi.org/10.1002/andp.201400067. 75, 76

[216] Mantela, M.; Lambropoulos, K.; Simserides, C. Charge transport properties of ideal and natural DNA segments, as mutation detectors. *Physical Chemistry Chemical Physics* **2023**, *25*, 7750--7762. https://doi.org/10.1039/D3CP00268C. 78

[217] Shih, C.T.; Wells, S.A.; Hsu, C.L.; Cheng, Y.Y.; Römer, R.A. The interplay of mutations and electronic properties in disease-related genes. *Scientific Reports* **2012**, *2*, 272. https://doi.org/10.1038/srep00272. 78

[218] Lambropoulos, K.; Simserides, C. Periodic, quasiperiodic, fractal, Kolakoski, and random binary polymers: Energy structure and carrier transport. *Physical Review E* **2019**, *99*, 032415. https://doi.org/10.1103/PhysRevE.99.032415. 81

[219] Lambropoulos, K.; Simserides, C. Spectral and transmission properties of periodic 1D Tight-Binding lattices with a generic unit cell: an analysis within the transfer matrix approach. *Journal of Physics Communications* **2018**, *2*, 035013. https://doi.org/10.1088/2399-6528/aab065. 81, 92

[220] Lambropoulos, K.; Simserides, C. Electronic structure and charge transport properties of atomic carbon wires. *Physical Chemistry Chemical Physics* **2017**, *19*, 26890--26897. https://doi.org/10.1039/C7CP05134D. 81

[221] Landauer, R. Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction. *IBM Journal of Research and Development* **1957**, *1*, 223--231. https://doi.org/10.1147/rd.13.0223. 81

[222] Buttiker, M. Symmetry of electrical conduction. *IBM Journal of Research and Development* **1988**, *32*, 317--334. https://doi.org/10.1147/rd.323.0317.

[223] Datta, S. *Electronic Transport in Mesoscopic Systems*; Cambridge Studies in Semiconductor Physics and Microelectronic Engineering, Cambridge University Press, 1995. https://doi.org/10.1017/CBO9780511805776. 81

[224] Bębenek, A.; Ziuzia-Graczyk, I. Fidelity of DNA replication - a matter of proofreading. *Current Genetics* **2018**, *64*, 985--996. https://doi.org/10.1007/s00294-018-0820-1. 83

[225] Kunkel, T.A.; Bebenek, K. DNA Replication Fidelity. *Annual Review of Biochemistry* **2000**, *69*, 497--529, [https://doi.org/10.1146/annurev.biochem.69.1.497]. PMID: 10966467, https://doi.org/10.1146/annurev.biochem.69.1.497.

[226] Showalter, A.K.; Tsai, M.D. A Reexamination of the Nucleotide Incorporation Fidelity of DNA Polymerases. *Biochemistry* **2002**, *41*, 10571--10576, [https://doi.org/10.1021/bi026021i]. PMID: 12186540, https://doi.org/10.1021/bi026021i. 83

[227] Dashnow, H.; Lek, M.; Phipson, B.; Halman, A.; Sadedin, S.; Lonsdale, A.; Davis, M.; Lamont, P.; Clayton, J.S.; Laing, N.G.; et al. STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biology* **2018**, *19*, 121. https://doi.org/10.1186/s13059-018-1505-2. 83

[228] Chintalaphani, S.R.; Pineda, S.S.; Deveson, I.W.; Kumar, K.R. An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathologica Communications* **2021**, *9*, 98. https://doi.org/10.1186/s40478-021-01201-x. 83, 84

[229] Kratter, I.H.; Finkbeiner, S. PolyQ Disease: Too Many Qs, Too Much Function? *Neuron* **2010**, *67*, 897--899. https://doi.org/https://doi.org/10.1016/j.neuron.2010.09.012. 83

[230] MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **1993**, *72*, 971--983. https://doi.org/10.1016/0092-8674(93)90585-E. 83

[231] McColgan, P.; Tabrizi, S.J. Huntington's disease: a clinical review. *European Journal of Neurology* **2018**, *25*, 24--34, [https://onlinelibrary.wiley.com/doi/pdf/10.1111/ene.13413]. https://doi.org/https://doi.org/10.1111/ene.13413. 83

[232] Fratta, P.; Collins, T.; Pemble, S.; Nethisinghe, S.; Devoy, A.; Giunti, P.; Sweeney, M.G.; Hanna, M.G.; Fisher, E.M. Sequencing analysis of the spinal bulbar muscular atrophy CAG expansion reveals absence of repeat interruptions. *Neurobiology of Aging* **2014**, *35*, 443.e1--443.e3. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2013.07.015. 84

[233] Kuhlenbäumer, G.; Kress, W.; Ringelstein, E.B.; Stögbauer, F. Thirty-seven CAG repeats in the androgen receptor gene in two healthy individuals. *Journal of Neurology* **2001**, *248*, 23--26. https://doi.org/10.1007/s004150170265.

[234] Spada, A.R.L.; Wilson, E.M.; Lubahn, D.B.; Harding, A.E.; Fischbeck, K.H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **1991**, *352*, 77--79. https://doi.org/10.1038/352077a0. 84

[235] Sequeiros, J.; Seneca, S.; Martindale, J. Consensus and controversies in best practices for molecular genetic testing of spinocerebellar ataxias. *European Journal of Human Genetics* **2010**, *18*, 1188--1195. https://doi.org/10.1038/ejhg.2010.10. 84

[236] Zhuchenko, O.; Bailey, J.; Bonnen, P.; Ashizawa, T.; Stockton, D.W.; Amos, C.; Dobyns, W.B.; Subramony, S.H.; Zoghbi, H.Y.; Lee, C.C. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α_{1A}-voltage-dependent calcium channel. *Nature Genetics* **1997**, *15*, 62--69. https://doi.org/10.1038/ng0197-62. 84

[237] Cagnoli, C.; Stevanin, G.; Michielotto, C.; Gerbino Promis, G.; Brussino, A.; Pappi, P.; Durr, A.; Dragone, E.; Viemont, M.; Gellera, C.; et al. Large Pathogenic Expansions in the SCA2 and SCA7 Genes Can Be Detected by Fluorescent Repeat-Primed Polymerase Chain Reaction Assay. *The Journal of Molecular Diagnostics* **2006**, *8*, 128--132. https://doi.org/10.2353/jmoldx.2006.050043. 84

[238] David, G.; Abbas, N.; Stevanin, G.; Dürr, A.; Yvert, G.; Cancel, G.; Weber, C.; Imbert, G.; Saudou, F.; Antoniou, E.; et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature Genetics* **1997**, *17*, 65--70. https://doi.org/10.1038/ng0997-65. 84

[239] Qiu, C.; McCann, K.L.; Wine, R.N.; Baserga, S.J.; Hall, T.M.T. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proceedings of the National Academy of Sciences* **2014**, *111*, 18554--18559. https://doi.org/10.1073/pnas.1407634112. 85, 87

[240] Wang, R.; Wang, S.; Dhar, A.; Peralta, C.; Pavletich, N.P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **2020**, *580*, 278--282. https://doi.org/10.1038/s41586-020-2110-6. 85, 87

[241] Institute, N.H.G.R. Primers:. https://www.genome.gov/genetics-glossary/Primer A primer, as related to genomics, is a short single-stranded DNA fragment used in certain laboratory techniques, such as the polymerase chain reaction (PCR). In the PCR method, a pair of primers hybridizes with the sample DNA and defines the region that will be amplified, resulting in millions and millions of copies in a very short timeframe. Primers are also used in DNA sequencing and other experimental processes. 90

[242] Thodi, G.; Fostira, F.; Sandaltzopoulos, R.; Nasioulas, G.; Grivas, A.; Boukovinas, I.; Mylonaki, M.; Panopoulos, C.; Magic, M.B.; Fountzilasand, G.; et al. Screening of the DNA mismatch repair genes MLH1, MSH2 and MSH6in a Greek cohort of Lynch syndrome suspected families. *BMC Cancer* **2010**, *10*, 544. https://doi.org/10.1186/1471-2407-10-544. 107

[243] Maciá, E. Aperiodic crystals in biology. *Journal of Physics: Condensed Matter* **2022**, *34*, 123001. https://doi.org/10.1088/1361-648X/ac443d. 107

[244] Kreile, J.; Münzel, N.; Schweig, A.; Specht, H. Uv photoelectron spectrum of cyclobutadiene. free cyclobutadiene stable up to high temperatures. *Chemical Physics Letters* **1986**, *124*, 140--146. https://doi.org/10.1016/0009-2614(86)85133-8. 115