



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
—ΙΔΡΥΘΕΝ ΤΟ 1837—

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικό και Καποδιστριακό

Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

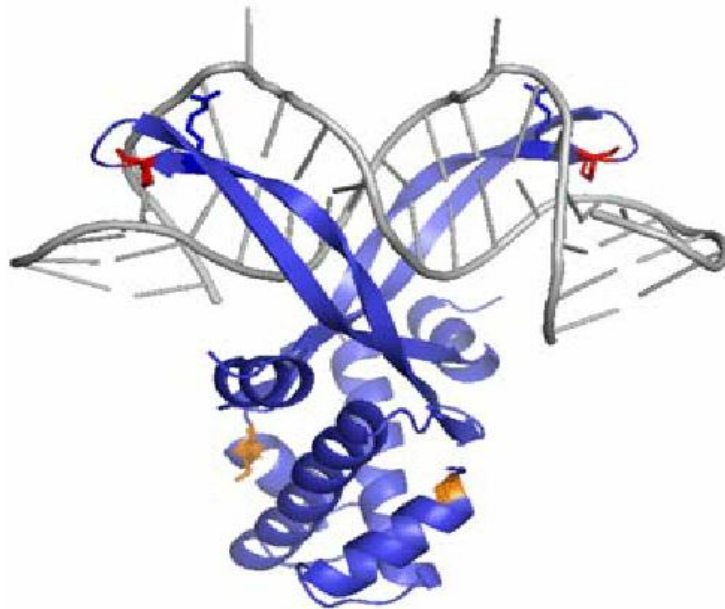
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ

ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Βιοπληροφορική ανάλυση των πρωτεϊνών της οικογένειαςHU»



ΔΗΜΗΤΡΙΟΣ ΑΝΤΩΝΟΓΙΑΝΝΑΚΗΣ

Πτυχιούχος Φυσικών Επιστημών, Ε.Α.Π

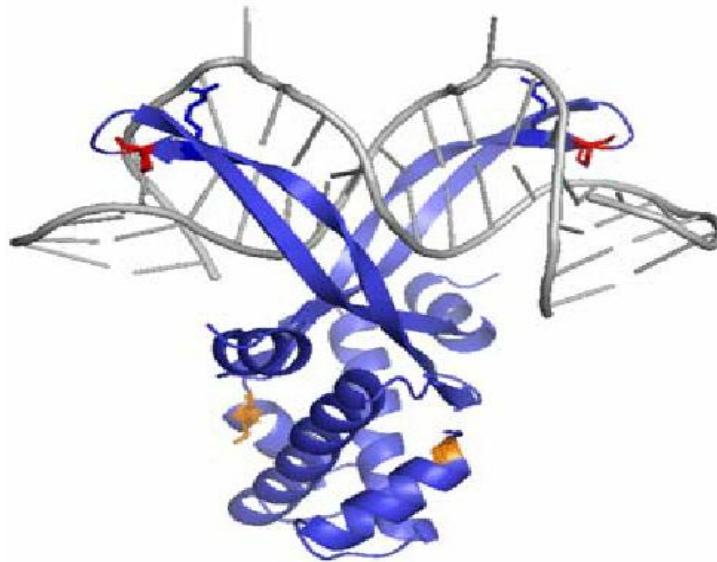
ΑΘΗΝΑ 2018



HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
— EST. 1837 —

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY
MASTER IN «BIOINFORMATICS»

Master Diploma Thesis
«Bioinformatic analysis of HU protein family»



DIMITRIOS ANTONOGIANNAKIS
BSc in Natural Sciences, Hellenic Open University (HOU)
A T H E N S 2 0 1 8



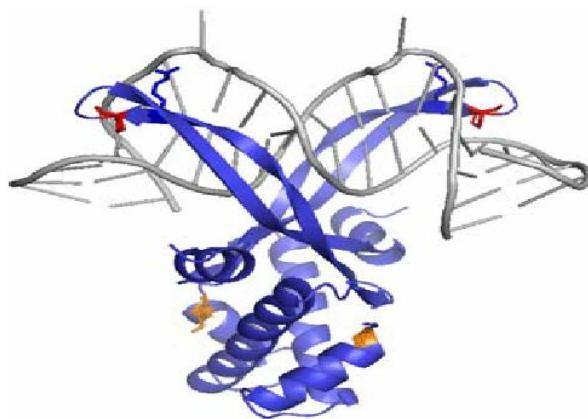
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
— ΙΔΡΥΘΕΝ ΤΟ 1837 —

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ
ΣΠΟΥΔΩΝ «ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Βιοπληροφορική ανάλυση των πρωτεϊνών της οικογένειαςHU»



Καθηγητής Κωνσταντίνος Βοργίας (Επιβλέπων)

Τομέας Βιολογίας,

Τμήμα Βιοχημείας και Μοριακής Βιολογίας, ΕΚΠΑ

Αναπληρωτής Καθηγητής Αλέξανδρος Γεωργακίλας

Τομέας Φυσικής,

Τμήμα ΣΕΜΦΕ, ΕΜΠ

Επίκουρη Καθηγήτρια Βασιλική Οικονομίδου

Τομέας Βιολογίας,

Τμήμα Κυτταρικής Βιολογίας και Βιοφυσικής, ΕΚΠΑ

Στον γιο μου, Μάξιμο,

για όλες εκείνες τις φορές που ρώτησε...

“Μπαμπά πότε θα τελειώσεις το σχολείο σου;”

Ευχαριστίες

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε στα πλαίσια του ΠΜΣ Βιοπληροφορικής του τμήματος Βιολογίας του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών, υπό την επίβλεψη του Καθηγητή Δρ. Κωνσταντίνου Βοργιά.

Ευχαριστώ θερμά τον Καθηγητή κ. Βοργιά για την καθοδήγησή του τόσο στη συγκεκριμένη εργασία, όσο και σε όλη την πορεία της συνεργασίας μας τα τελευταία δύο χρόνια.

Ακόμα, ευχαριστώ θερμά τον Δρ. Νικόλαο Παπανδρέου κατά την διάρκεια της εργασίας μου και για την τεράστια υπομονή του. Επίσης, θα ήθελα να ευχαριστήσω την Δρ. Β. Οικονομίδου και τον Δρ. Α. Γεωργακίλα που δέχτηκαν με χαρά να είναι μέλη της τριμελούς επιτροπής μου.

Τέλος, θα ήθελα να πω ένα μεγάλο ευχαριστώ στη σύντροφό μου και τον γιό μου που είναι πάντα δίπλα μου και με στηρίζουν σε κάθε μου βήμα, με υπέρμετρη κατανόηση και αγάπη.

Περιεχόμενα:

1. Εισαγωγή	5
1. Το βακτηριακό χρωμόσωμα και οι πρωτεΐνες HU.....	6
1.2 Ο λειτουργικός ρόλος της HU.....	9
1.3 Δομικά δεδομένα της πρωτεΐνης HU.....	11
1.4 Οι πρωτεΐνες HU ως μοντέλο μελέτης θερμοσταθερότητας των πρωτεϊνών.....	15
1.5 Σταθερότητα πρωτεϊνών.....	16
1.6 Θερμοσταθερότητα πρωτεϊνών.....	19
1.7 Μηχανισμοί θερμοσταθερότητας των πρωτεϊνών HU.....	23
2. Στόχοι	25
3. Μεθοδολογία	27
3.1 Βάσεις δεδομένων που χρησιμοποιήθηκαν.....	27
3.2 Κατασκευή-αξιολόγηση πρωτεϊνικών μοντέλων με την τεχνική του Homology Modelling.....	39
3.3 Κατασκευή-αξιολόγηση ομοδιμερών ή ετεροδιμερών μοντέλων με την τεχνική του Homology Modelling.....	51
3.4 Εντοπισμός σήματος διμερισμού των μοντέλων HU.....	59
3.5 Υπολογισμός ενδομοριακών και διαμοριακών αλληλεπιδράσεων.....	64
4. Αποτελέσματα	67
4.1 Συλλογή-αξιολόγηση αμινοξικών ακολουθιών από βάσεις δεδομένων.....	67
4.2 Συλλογή-αξιολόγηση γνωστών δομών από βάσεις δεδομένων.....	69
4.3 Στοίχιση πρωτεϊνικών αλληλουχιών.....	72
4.4 Κατασκευή πρωτεϊνικών μοντέλων με την τεχνική του Homology Modeling.....	74
4.5 Αξιολόγηση πρωτεϊνικών μοντέλων μέσω WhatIF.....	76
4.6 Ενεργειακή ελαχιστοποίηση μέσω Chimera.....	78
4.7 Κατασκευή ομοδιμερών & ετεροδιμερών μοντέλων HU.....	80
4.8 Αξιολόγηση ομο/ετερο διμερών πρωτεϊνικών μοντέλων.....	84
4.9 Μελέτη της επίδρασης αλληλουχίας διμερισμού στην σταθερότητα των HU.....	86
4.10 Συγκριτική ανάλυση διαφόρων DS και συσχέτιση σταθερότητας.....	88
4.11 Συγκριτική ανάλυση ασθενών αλληλεπιδράσεων στις HU.....	93
4.12 Συσχετισμός θερμοκρασίας ανάπτυξης βακτηρίων με τις ασθενείς αλληλεπιδράσεις των HU.....	95
5. Συζήτηση	111
6. Βιβλιογραφία	115
7. Appendix	122

1.Εισαγωγή

Γενικά

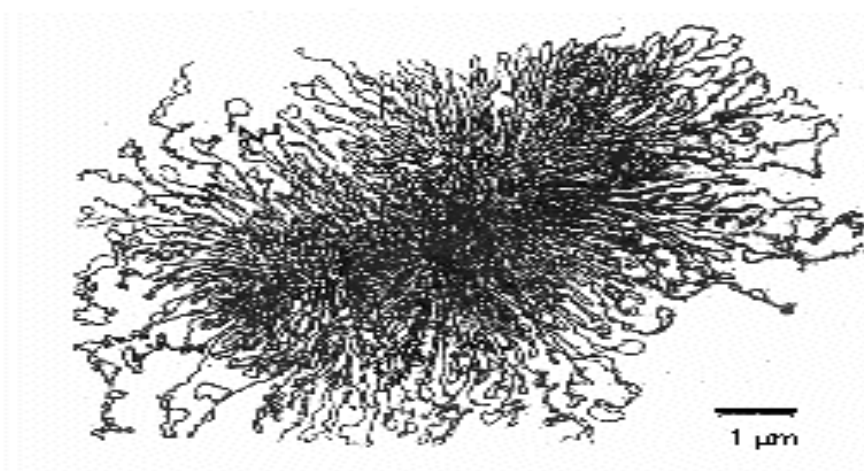
Το βακτηριακό χρωμόσωμα περιλαμβάνει όλα τα γονίδια ενός βακτηρίου τα οποία είναι διατεταγμένα σε ένα υπερελικώμενο κυκλικό μόριο DNA στο οποίο οι διάφορες περιοχές του πρέπει να είναι προσβάσιμες κάθε στιγμή από μηχανισμούς μεταγραφής, αναγνώρισης και ρύθμισης του κυττάρου. Στο κέντρο αυτού του βακτηριακού χρωμοσώματος υπάρχει ένα συμπυκνωμένο κυκλικό χρωμόσωμα το οποίο σχετίζεται με 10-20 DNA binding proteins που όλες μαζί απαρτίζουν το βακτηριακό νουκλεοειδές [1].

Οι DNA binding proteins είναι μια πρωτεϊνική οικογένεια που αποτελείται από δύο ομάδες πρωτεϊνών: (α) HU που θα αναλύσουμε περαιτέρω και (β) την υποομάδα IHF(Integration Host Factor) [2].

Οι HU είναι μια ομάδα μικρών πρωτεϊνών συνήθως 90 αμινοξικών καταλοίπων η οποία μοιάζει με την ευκαρυωτική H2B. Η HU ενεργεί όμοια με την τελευταία, ελαττώνοντας την αρνητική υπερελίκωση στο κυκλικό DNA, με την βοήθεια της τοποισομεράσης I. Η ονομασία HU προκύπτει από το στέλεχος *Escherichia coli* (*E. coli*) U93 από όπου απομονώθηκε για πρώτη φορά από την ενώ το γράμμα H προστέθηκε εκ των υστέρων όταν αποσαφηνίστηκε η ομοιότητα της με τις ευκαρυωτικές ιστόνες. Στα εντεροβακτήρια (π.χ. *E. coli*) η HU εμφανίζεται κυρίως ως ετεροδιμερές (2 γονίδια), ενώ στα υπόλοιπα ευβακτήρια ως ομοδιμερές (1 γονίδιο) [3, 4]. Στο *E. coli* οι δυο υπομονάδες HUα και HUβ έχουν συνήθως 90 κατάλοιπα, παρουσιάζουν ομοιότητα 70% και κωδικοποιούνται από δυο γονίδια *hupA* και *hupB* αντίστοιχα [5, 6]. Για την ονοματολογία της στα υπόλοιπα βακτήρια χρησιμοποιείται ως πρώτο γράμμα το H το οποίο και ακολουθείται από 1-4 γράμματα, 1 για το γένος και 1-3 για το είδος (π.χ. HUBst για τον οργανισμό *Bacillus stearothermophilus*).

Το βακτηριακό χρωμόσωμα και οι πρωτεΐνες HU

Όλα τα γονίδια ενός βακτηρίου περιέχονται σε ένα υπερελικωμένο, κυκλικό μόριο DNA το οποίο πρέπει να πακεταριστεί με κατάλληλο τρόπο μέσα στο κύτταρο ώστε οι διάφορες περιοχές του να είναι ανά πάσα στιγμή προσιτές προς χρήση από τους μηχανισμούς αναγνώρισης, ρύθμισης, μεταγραφής, κ.α. του κυττάρου. Δύο είναι τα βασικά χαρακτηριστικά του βακτηριακού χρωμοσώματος: α) 45 ± 10 διακριτές τοπολογικά μονάδες (domains or loops) μεγέθους περίπου 100 kbp οι οποίες μπορούν να διατηρούν διαφορετικά επίπεδα υπερελίκωσης [7]. β) στο κέντρο υπάρχει μια πυκνή και πολύπλοκη δομή που αποτελείται από πρωτεΐνες και RNA (scaffold). Το συμπυκνωμένο λοιπόν κυκλικό χρωμόσωμα σχετίζεται με 10-20 DNA binding proteins που όλες μαζί συνιστούν το βακτηριακό νουκλεοειδές [7]. Η σχετική σύσταση του νουκλεοειδούς αλλά και η δομή του DNA ποικίλουν ανάλογα με τις συνθήκες της κυτταρικής αύξησης [8].



Εικόνα 1 Απεικόνιση του βακτηριακού χρωμοσώματος από τον οργανισμό *Escherichia coli* (*E. coli*).

Οι πρωτεΐνες που ανήκουν στην οικογένεια των “histone-like DNA binding proteins” είναι, εκτός των HU που περιγράφονται παρακάτω, οι IHF (Integration Host Factor), HNS, H, HLP1, FIS. [2]. Στην *E. coli* η πρωτεΐνη HU αναγνωρίστηκε αρχικά ως μεταγραφικός ενεργοποιητής πολλών γονιδίων βακτηριοφάγων λ [9]. Επειδή η πρώτη απομόνωση της πρωτεΐνης από *E. coli* έγινε από το στέλεχος U93 ονομάστηκε

παράγοντας U ενώ το γράμμα H προστέθηκε όταν έγινε σαφές ότι μοιάζει με τις ευκαρυωτικές ιστόνες. Συγκεκριμένα η πρωτεΐνη αυτή έχει ασυνήθιστη αμινοξική σύσταση για πρωτεΐνη του *E. coli* ενώ μετά από σύγκριση βρέθηκε ότι προέρχεται από την ευκαρυωτική ιστόνη H2B [9]. Απαντάται στα αρνητικά και θετικά κατά Gram βακτήρια, σε ευκαρυωτικές οργανέλες [10] καθώς και σε έναν ιό [11] και ένα βακτηριοφάγο [12]. Είναι μια άφθονη (30.000 αντίγραφα/κύτταρο), μικρή (19 kDa), βασική (pI=10,5) πρωτεΐνη.

Στα εντεροβακτήρια (π.χ. *E. coli*) η HU εμφανίζεται κυρίως ως ετεροδιμερές (2 γονίδια), ενώ στα υπόλοιπα ευβακτήρια απαντάται ως ομοδιμερές (1 γονίδιο) [3, 4]. Στο *E. coli* οι δυο υπομονάδες HU α (ή NS2 ή HU2) και HU β (ή NS1 ή HU1) έχουν 90 κατάλοιπα, παρουσιάζουν ομοιότητα 70% και κωδικοποιούνται από δυο γονίδια *hupA* και *hupB* αντίστοιχα [5, 6]. Για την ονοματολογία της στα υπόλοιπα βακτήρια χρησιμοποιείται ως πρώτο γράμμα το H το οποίο και ακολουθείται από 1-4 γράμματα, 1 για το γένος και 1-3 για το είδος (π.χ. HUBst για τον οργανισμό *Bacillus stearothermophilus*).

In vivo αλλά και in vitro η HU στο *E. coli* εμφανίζεται σε 3 διμερείς μορφές HU α_2 , HU β_2 , HU $\alpha\beta$ [3], που η παρουσία τους ποικίλει κατά την κυτταρική ανάπτυξη. Έτσι η HU α_2 εμφανίζεται-συσσωρεύεται κατά τη φάση διαίρεσης των κυττάρων, ενώ η HU $\alpha\beta$ εμφανίζεται κατά το μεταβατικό στάδιο οπότε και κυριαρχεί κατά το τέλος του εκθετικού σταδίου ανάπτυξης και εν συνεχεία κατά το στάδιο σταθεροποίησης [13]. Μόνο το 5% των ομοδιμερών κατά το στάδιο της σταθεροποίησης ανιχνεύονται σε μορφή HU β_2 [3]. Το μικρό αυτό ποσοστό δικαιολογείται από την πρόταση ότι τα HU β_2 μπορεί να είναι δηλητηριώδη για τα κύτταρα, κάτι που αποδεικνύεται από το γεγονός η HU β_2 αποικοδομείται από την Lonprotease στο *hupA*-mutant. Στο μεταβατικό στάδιο τα ομοιομερή ανταλλάσσουν τις αλυσίδες έτσι ώστε να δημιουργηθούν τα ετεροδιμερή με κάποιο μοριακό μηχανισμό που μας είναι άγνωστος [13]. Εύκολα λοιπόν συμπεραίνεται ότι πιθανότατα υπάρχει διαφορετική βιολογική λειτουργία των 3 διμερών [13]. In vitro το HU α_2 και το HU $\alpha\beta$ προσδένονται με την ίδια συγγένεια σε DNA που περιέχει μονόκλιωνα σπασίματα ενώ αντιθέτως το HU β_2 έχει χαμηλότερη συγγένεια πρόσδεσης. Επίσης έχει δείχτει ότι ο κυτταρικός κύκλος μπορεί να αλλάξει από μια μη κανονική αναλογία HU διμερών

[14] ενώ ακόμα και τα τρία διμερή προσδένονται με συγκρίσιμες συγγένειες με το σταυρόμορφο (cruciform) DNA [15].

Εκτός όμως από τα βακτήρια, η πρωτεΐνη HU έχει βρεθεί και στο αρχαιοβακτήριο *Thermoplasma acidophilum* [16]. Φυλογενετικά τα αρχαιοβακτήρια βρίσκονται ανάμεσα στα προκαρυωτικά και τα ευκαρυωτικά καθώς δανείζονται βιοχημικά χαρακτηριστικά και από τις δυο μεγάλες αυτές ομάδες. Η HU από το *T. acidophilum* αντανακλά αυτή ακριβώς τη θέση των αρχαιοβακτηρίων καθώς η αμινοξική της αλληλουχία εμφανίζει ομολογία με την HU των βακτηρίων αλλά και με τις ευκαρυωτικές ιστόνες H2A και H3 [17].

Ο λειτουργικός ρόλος της HU

Με βάση τις μέχρι τώρα *in vitro* έρευνες, διάφορες λειτουργίες έχουν αποδοθεί στις HU πρωτεΐνες. Ακολουθούν ορισμένα χαρακτηριστικά που αφορούν την ικανότητα της HU για πρόσδεση στο DNA και στα οποία βασίζεται ο φυσιολογικός της ρόλος.

1. Η HU ρυθμίζει την αρνητική υπερέλικωση του DNA. Μπορεί να εισάγει αρνητικές στροφές υπερέλικας σε χαλαρό κυκλικό πλασμίδιο παρουσία τοποισομεράσης I. Γενικά συμβάλει στη δημιουργία και στη διατήρηση υπερελικωμένων μορφών DNA [18-21].
2. Όπως οι ευκαρυωτικές ιστόνες, προσδένεται στο DNA και προκαλεί τη δημιουργία δομών όπως τα νουκλεοσώματα [3].
3. Συμμετέχει σε πολλές κυτταρικές διεργασίες όπως για παράδειγμα στην μεταγραφή του DNA [22] και στην προστασία από την γάμμα ή UV ακτινοβολία [23, 24] καθώς επίσης και στην επιδιόρθωση του DNA [25].
4. Προσδένεται σε δίκλωνο και μονόκλωνο DNA και RNA και σε υβρίδια DNA-RNA μόρια χωρίς να αναγνωρίζει συγκεκριμένη αλληλουχία. Παρουσιάζει μεγαλύτερη συγγένεια για μονόκλωνο συγγένεια DNA και RNA.
5. Σε μοριακό επίπεδο, έχει δείχτει ότι η HU σχηματίζει σύμπλοκα χαμηλής ή υψηλής συγγένειας με το DNA. Στα χαμηλής συγγένειας σύμπλοκα ($K_A=10^7 M^{-1}$) η HU προσδένεται στο DNA με στοιχειομετρία 1 διμερές/9bp χωρίς κάποια ειδικότητα ως προς την ακολουθία [18], ενώ στα υψηλής συγγένειας σύμπλοκα η HU αναγνωρίζει ευέλικτες μορφές όπως θηλιές (stem loops) και DNA με μονόκλινα σπασίματα (nicks and gaps) [22, 26] καθώς και RNA [27].
6. Εκτός από την πρόσδεση στο DNA, στο *E. coli* έχει δείχτει ότι καταστολή των δυο αλληλόμορφων γονιδίων επηρεάζει έντονα την ανάπτυξη του μεταλλαγμένου βακτηρίου και σχετίζεται ακόμα με διαταραχές στην κυτταρική διαίρεση, δημιουργία απύρηνων βακτηρίων, θερμοευαίσθητο (heat-lethal) φαινότυπο και υπερευαίσθησία στην ιονίζουσα ακτινοβολία [23]. Βακτήρια *E. coli* με κατασταλαμένο το γονίδιο *hurA* παρουσιάζουν έναν «πιο ήπιο» φαινότυπο ενώ βακτήρια με κατασταλαμένο το *hurB* παρουσιάζουν φαινότυπο άγριου τύπου. Σημαντική επίσης παρατήρηση είναι ότι τα βακτήρια με κατασταλαμένο το ένα ή

και τα δυο αλληλόμορφα είναι ευαίσθητα σε ψυχρό ή θερμικό στρές (cold/heat-shock stress).

7. Προσδένεται κατά προτίμηση σε ολιγονουκλεοτίδια πλούσια σε GC.
8. Μπορεί να σχηματίζει ανώτερης τάξης δομές όπως διμερή, τριμερή και τετραμερή, εκμεταλλεζόμενη αλληλεπιδράσεις μεταξύ των πρωτεϊνών.
9. Η HU εμφανίζει θερμοσταθερότητα. Σε πειράματα όπου η πρωτεΐνη βρισκόταν σε διάλυμα NaCl 0,4 M και ακολούθησε θέρμανση έως τους 100°C και κατόπιν ψύξη, η πρωτεΐνη επανέκτησε την ικανότητα της να προσδένεται στο DNA [9]. Βιοφυσικές μελέτες δείχνουν ότι το ετεροδιμερές (όπου αυτό υπάρχει) είναι σταθερότερο από τα ομοδιμερή. Σε αυτές τις μελέτες το ομοδιμερές HU β_2 απέτυχε να αποδιαταχτεί τόσο αποτελεσματικά όσο το ετεροδιμερές και το ομοδιμερές HU α_2 [28].

Συνοψίζοντας, η επικρατούσα άποψη για τον ρόλο των HU πρωτεϊνών *in vivo* είναι ότι μέσω της ιδιότητάς τους να προκαλούν κάμψη στο DNA και να διατηρούν υπερελικωμένες δομές, διευκολύνουν ειδικές αλληλεπιδράσεις DNA-πρωτεϊνών που απαιτούνται για τη ρύθμιση της αντιγραφής, του γενετικού ανασυνδυασμού και της γονιδιακής έκφρασης.

Δομικά Δεδομένα των HU

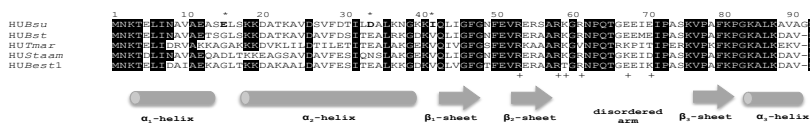
Οι HU απαντώνται στα αρνητικά και θετικά κατά Gram βακτήρια, σε ευκαρυωτικά οργανίδια [10, 29], σε ένα ιο [11] και σε ένα βακτηριοφάγο [12]. Επίσης έχει βρεθεί και στο αρχαιοβακτήριο *Thermoplasma acidophilum* [16] το οποίο είναι στα όρια μεταξύ αρχαιοβακτηρίων και ευβακτηρίων. Η HU από το *Thermoplasma acidophilum* εμφανίζει ομολογία στην πρωτοταγή δομή τόσο με τα βακτήρια, όσο και με τις ευκαρυωτικές ιστόνες H2A και H3 [17].

Ενώ η πρωτοταγής δομή της HU έχει προσδιορισθεί σε παραπάνω από 100 βακτήρια, μόνο από λίγους μικροοργανισμούς έχει λυθεί η τρισδιάστατη δομή της, τόσο μέσω κρυσταλλογραφίας ακτινών-X όσο και σε διάλυμα με την μέθοδο του Nuclear Magnetic Resonance (NMR). Συγκεκριμένα έχουν δημοσιευθεί οι δομές για τα βακτήρια *E.coli*, *Borrelia burgdorferi*, *Bacillus stearothermophilus*, *Anabaenasp*, *Mycobacterium tuberculosis* και *Thermotoga maritima*.

Ο *Bacillus stearothermophilus* ήταν ο πρώτος οργανισμός στον οποίο έγινε ο προσδιορισμός της πρωτοταγής αμινοξικής ακολουθίας [30], κρυσταλλώθηκε [31] και στην συνέχεια προσδιορίστηκε η τρισδιάστατη δομή της μέσω κρυσταλλογραφίας ακτινών-X, αρχικά με διακριτικότητα 3 Å [32] και αργότερα με διακριτικότητα 2,1 Å [33]. Αργότερα κλωνοποιήθηκε το πρώτο γονίδιο, *hubst*, και προσδιορίστηκε η νουκλεοτιδική αλληλουχία [34]. Τέλος έχει προσδιοριστεί η δομή της HUBst σε διάλυμα μέσω NMR [35], με τα αποτελέσματα να συνάδουν με αυτά που προκύπτουν από την κρυσταλλογραφία ακτινών-X [16].

Η δομή του μονομερούς της HUBst που είναι η πλέον μελετημένη απεικονίζεται στο παρακάτω σχήμα και αποτελείται από 3 διακριτά στοιχεία: (α) το HTH (helix-turn-helix), (β) DS (Dimerization Signal) και (γ) DBD (DNA binding domain).

Figure 3 Supplement



Εικόνα 2. Στοίχιση των HUBsu, HUBst, HUTmar και HUStaam όπου απεικονίζεται η δευτεροταγής δομή (α-έλικες, β-πτυχωτές επιφάνειες και το DSL: ακολουθία υπεύθυνη για το διμερισμό των πρωτεϊνών HU και IHF)

Η περιοχή HTH της πρωτεΐνης HU αποτελείται από δύο α-έλικες (α₁: αμινοξικά κατάλοιπα 3-13, α₂: αμινοξικά κατάλοιπα 21-37) οι οποίες συνδέονται με μια β-στροφή (αμινοξικά κατάλοιπα 14-20) σχηματίζοντας μια υπερδευτεροταγή δομή σχήματος V, στην βάση της οποίας οι 2 έλικες έρχονται σε επαφή μέσω δυο συντηρημένων αμινοξικών καταλοίπων Ala11 και Ala 21 [15].

Η περιοχή κοντά στο καρβοξυτελικό άκρο του μορίου αποτελείται κυρίως από μια αντιπαράλληλη β-πτυχωτή επιφάνεια η οποία εντοπίζεται στο πάνω μέρος του σχηματισμού V, καλύπτοντας τις α-έλικες και ενώνεται με την α₂ έλικα του αμινοτελικού άκρου μέσω μιας στροφής που έχει συντηρημένο ένα κατάλοιπο Gly39. Η εν λόγω αντιπαράλληλη β-πτυχωτή επιφάνεια αποτελείται από 3 β-κλώνους (β₁: 40-44, β₂: 48-51, β₃: 78-83). Οι κλώνοι β₁ και β₂ συνδέονται με μια πολύ συντηρημένη στροφή ειδικότερα στις θέσεις 46 και 48. Οι κλώνοι β₂ και β₃ συνδέονται μέσω μιας ευέλικτης δομής που προεξέχει από το κυρίως σώμα του μορίου (arm region). Ουσιαστικά οι κλώνοι β₂ και β₃ επεκτείνονται μέχρι τα αμινοξικά κατάλοιπα 53-56 και 73-76 αντίστοιχα, σχηματίζοντας μια δίκλωνη αντιπαράλληλη β-πτυχωτή επιφάνεια.

Στα αμινοξικά κατάλοιπα Ala57 και Pro72 οι κλώνοι εκτείνονται προς το εξωτερικό της δομής και συνεχίζουν ως δίκλωνη αντιπαράλληλη β-πτυχωτή επιφάνεια μέχρι τα κατάλοιπα Asn62 και Glu67. Μεταξύ των καταλοίπων 62 και 67 υπάρχει μια β-στροφή. Η όλη δομή χαρακτηρίζεται ως β-hairpin arm ή β-ribbon arm. Το τελευταίο στοιχείο δευτεροταγούς δομής είναι μια μικρή α-έλικα (α_3 : 84-89) στο καρβοξυτελικό άκρο [15].

Τα δυο μονομερή έρχονται σε επαφή για το σχηματισμό της διμερούς. Η βάση του μορίου αποτελείται από τις δυο υπερδευτεροταγείς δομές σχήματος V που σχηματίζουν οι τέσσερις α-έλικες (2 από κάθε μονομερές), δημιουργώντας μια πλατφόρμα τεσσάρων α-ελίκων. Ουσιαστικά οι α_2 έλικες βρίσκονται υπό γωνία επαπτόμενες μέσω της συντηρημένης Ala24 [15] και οι α_1 βρίσκονται σε κάθε πλευρά. Η βάση αυτή καλύπτεται από τις δυο τρίκλωνες β-πτυχωτές επιφάνειες αλλά λόγω της απόστασης των δυο β_3 κλώνων και του προσανατολισμού τους δεν σχηματίζουν μια β-πτυχωτή επιφάνεια έξι κλώνων [15]. Ουσιαστικά τα καρβοξυτελικά άκρα των μονομερών δεν αλληλεπιδρούν αλλά αποκλίνουν σχηματίζοντας δυο ευέλικτα προεξέχοντα άκρα (β-arms) κατάλληλα για την αλληλεπίδραση με τη διπλή έλικα του DNA. Οι δύο α_3 έλικες διευθετούνται οριζοντίως σε αντίθετες πλευρές των β-πτυχωτών επιφανειών στην εξωτερική επιφάνεια του μορίου.

Το διμερές παρουσιάζει τυπικά δομικά χαρακτηριστικά διαλυτών πρωτεϊνών με υδρόφιλη επιφάνεια και έναν υδρόφοβο πυρήνα στο σχηματισμό του οποίου συμμετέχουν κατάλοιπα και των δυο μονομερών. Τα κατάλοιπα που είναι εξαιρετικά συντηρημένα και υπεύθυνα για τη σταθεροποίηση του διμερούς είναι τέσσερις Phe στις θέσεις (29, 47, 50, 79). Επομένως οι δυο υπομονάδες συγκρατούνται με υδρόφοβες αλληλεπιδράσεις, και η πρωτεΐνη υπάρχει σε διάλυμα σε μορφή διμερούς ή τετραμερούς. Τα προεξέχοντα άκρα του μορίου, βραχίονες (β-arms) μαζί με τους συμμετρικά τοποθετημένους β_3 κλώνους ανάμεσα τους δημιουργούν μια κοιλότητα διαμέτρου $\sim 25\text{\AA}$, η οποία είναι γεωμετρικά συμπληρωματική με τη διπλή έλικα B-DNA [15]. Επίσης, η αμινοξική αλληλουχία των προεξέχόντων άκρων είναι αρκετά συντηρημένη [15], υποδηλώνοντας το ζωτικό λειτουργικό της ρόλο, και περιέχει τέσσερα από τα πέντε κατάλοιπα Arg του κάθε μονομερούς (στις θέσεις 53, 55, 58,

61), τρία από τα οποία (στις θέσεις 55, 58, 61) έχει αποδειχθεί ότι συμμετέχουν στην πρόσδεση των HU με διπλό κλώνο DNA [36]. Επίσης, η αλληλεπίδραση HU-DNA εξαρτάται από συγκεντρώσεις άλατος [3]. Οι παρατηρήσεις αυτές υποδεικνύουν, και τα πειραματικά στοιχεία υποστηρίζουν, ότι η πρωτεΐνη προσδένεται με ηλεκτροστατικές αλληλεπιδράσεις στον φωσφορικό σκελετό του DNA (και του RNA) [15]. Πρέπει να τονιστεί επίσης πως οι προεξέχοντες βραχίονες εμφανίζουν υψηλή κινητικότητα και ευλυγισία σε σχέση με το υπόλοιπο σώμα του μορίου.

Συνοψίζοντας, το μονομερές μόριο των HU μπορεί να χωριστεί σε τρεις δομικά και λειτουργικά διακριτές περιοχές [37]. Το καρβοξυτελικό άκρο του μορίου αποτελεί τη θέση πρόσδεσης στο DNA (DNA-binding domain, DBD) είναι εξαιρετικά συντηρημένο και εμφανίζει υψηλή περιεκτικότητα σε βασικά αμινοξέα. Το κεντρικό τμήμα του μορίου περιέχει την αλληλουχία “GFGXF” (κατάλοιπα 46-50), κοινή στην οικογένεια των HU. Η αλληλουχία αυτή δημιουργεί έναν υδρόφοβο πυρήνα στην επιφάνεια μεταξύ των δυο υπομονάδων, και επειδή συμβάλλει στο σχηματισμό πολύ σταθερού διμερούς ονομάζεται σήμα διμερισμού (Dimerization Signal, DS). Το αμινοτελικό άκρο του μορίου περιλαμβάνει το δομικό μοτίβο έλικα - στροφή - έλικα (helix – turn – helix motif) μαζί με στοιχεία β-πτυχωτής επιφάνειας. Το τμήμα αυτό συνιστά το κυρίως σώμα της πρωτεΐνης και είναι λιγότερο συντηρημένο όσον αφορά στην πρωτοταγή του δομή σε σύγκριση με το υπόλοιπο μόριο [15].

Οι πρωτεΐνες HU ως μοντέλο μελέτης της θερμοσταθερότητας πρωτεϊνών

Οι πρωτεΐνες HU είναι ένα ιδανικό μοντέλο για μελέτες θερμοσταθερότητας πρωτεϊνών διότι:

1. Είναι γνωστή η πρωτοταγής δομή από περίπου 150 βακτήρια.
2. Η τρισδιάστατη δομή αρκετών HU έχει λυθεί από αρκετούς οργανισμούς μέσω κρυσταλλογραφίας ακτινών-X και μάλιστα σε αρκετά υψηλή διακριτικότητα (~2 Å)
3. Οι πρωτεΐνες HU εμφανίζουν θερμοσταθερότητα που συνάδει σε κάποιο βαθμό με την θερμοκρασία ανάπτυξης των μικροοργανισμών που προέρχονται [38].
4. Αρκετές HU πρωτεΐνες έχουν μελετηθεί θερμοδυναμικά και λόγω της θερμικής τους αντιστρεψιμότητας μας δίνουν την δυνατότητα να κάνουμε πλήρη πειράματα θερμοδυναμικής σταθερότητας [28].
5. Παρατηρείται μεγάλο ποσοστό ομολογίας μέχρι 85% μεταξύ διαφόρων HU αλληλουχιών από διαφορετικούς οργανισμούς με μεγαλώνουν σε πολύ διαφορετικά περιβάλλοντα, γεγονός που μας επιτρέπει να χρησιμοποιήσουμε συγκεκριμένα μόρια HU ως πρότυπα μόρια για μεταλλαξιγένεση και δομική μοντελοποίηση,
6. Τα γονίδια πολλών HU έχουν κλωνοποιηθεί, εκφράζονται ετερόλογα και σε μεγάλες ποσότητες ανασυνδυασμένης πρωτεΐνης που μπορούμε να καθαρίσουμε σχετικά εύκολα και γρήγορα.
7. Όλες οι HU πρωτεΐνες είναι σε διμερή μορφή άρα μπορούμε να μελετήσουμε τους μηχανισμούς διμερισμού και σταθεροποίησης των διμερών.

Μετά από όλα τα παραπάνω είναι εμφανές ότι η οικογένεια των HU πρωτεϊνών αποτελεί πρότυπο σύστημα μελέτης της θερμοσταθερότητας τόσο σε δομικό όσο και σε θερμοδυναμικό επίπεδο.

Σταθερότητα των πρωτεϊνών

Η σταθερότητα των πρωτεϊνών αναφέρεται στην ικανότητα τους να διατηρούν την φυσιολογική τρισδιάστατη δομή τους κάτω από διάφορες συνθήκες από κανονικές μέχρι ακραίες συνθήκες (θερμοκρασία, αλατότητα, οξύτητα, πίεση κλπ) ή ακόμη και χημικούς αποδιατακτικούς ή άλλους φυσικούς παράγοντες [39]. Η φυσιολογική στερεοδιαμόρφωση είναι στην πραγματικότητα ένα άθροισμα διαμορφώσεων της δομής και οι πειραματικά προσδιορισμένες δομές αποτελούν μια μέση διαμόρφωση ή ένα στιγμιότυπο διαμορφώσεων. Από την άλλη πλευρά, η αποδιαταγμένη κατάσταση αποτελείται επίσης από πολλές τυχαίες διαμορφώσεις [39].

Η αύξηση θερμοκρασίας, η αλλαγή pH, η εφαρμογή εξωτερικής πίεσης, η προσθήκη χημικών αποδιατακτικών παραγόντων (π.χ. ουρία, χλωριούχο γουανιλίδιο κ.λ.π.) είναι μερικοί από τους παράγοντες που προκαλούν την αποδιάταξη πρωτεϊνών (denaturation), δηλαδή την απώλεια της τρισδιάστατης δομής τους. Η αλληλεπίδραση ασθενών δεσμών των αποδιαταγμένων πρωτεϊνών όμως εξακολουθεί να υφίσταται, αλλά αυτή τη φορά με τα μόρια του διαλύτη. Είναι αυτοί οι δεσμοί που αντικαθίστανται από ενδομοριακές αλληλεπιδράσεις κατά την αναδίπλωση των αποδιαταγμένων δομών. Βεβαίως συμβαίνει και το αντίστροφο. Όπως είναι αναμενόμενο άλλωστε τα πρωτεϊνικά μόρια στην αποδιαταγμένη τους μορφή εμφανίζουν μεγαλύτερους βαθμούς ελευθερίας καθώς και μεγαλύτερη εντροπία. Και οι δύο αυτοί παράμετροι μειώνονται στην διπλωμένη κατάσταση των πρωτεϊνών.

Πειραματικά η μέτρηση της μεταβολής της σταθερότητας επιτυγχάνεται έμμεσα, με συσχέτιση της δομής με την σταθερότητα. Οι δομικές μεταβολές στα πρωτεϊνικά μόρια ανιχνεύονται με τη χρήση τεχνικών Πυρηνικού Μαγνητικού Συντονισμού (Nuclear Magnetic Resonance), Διαφορική Σάρωση Θερμιδομετρίας (Differential Scanning Calorimetry), Κυκλικού διχρωισμού, Φθορισμομετρία και άλλες μοντέρνες βιοφυσικές μεθόδους.

Η φυσική και η αποδιαταγμένη κατάσταση μιας πρωτεΐνης είναι θερμοδυναμικά καλά ορισμένες και πρωτεϊνική σταθερότητα μπορεί να εκφραστεί ως η διαφορά της

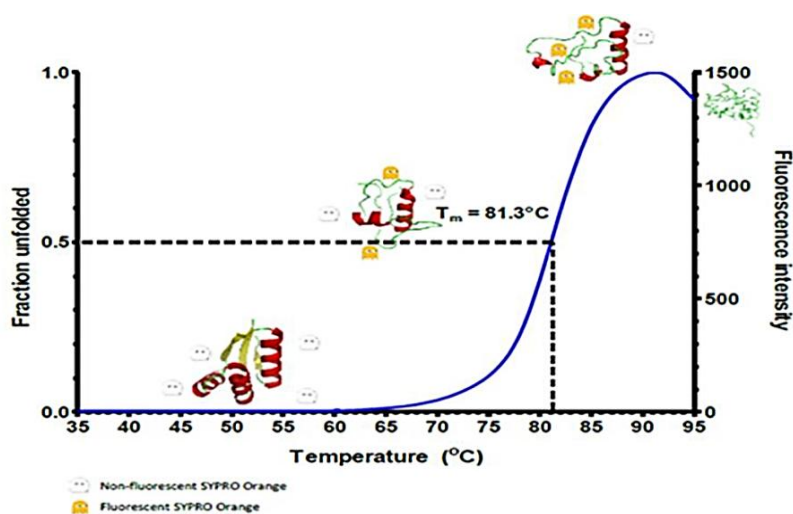
ελεύθερης ενέργειας του Gibbs (Gibbs free energy (ΔG), $\Delta G = \Delta G_{\text{unfold}} - \Delta G_{\text{fold}}$) των δύο αυτών καταστάσεων. Η ισορροπία μεταξύ της φυσικής και της αποδιαταγμένης κατάστασης εξαρτάται από την θερμοκρασία. Το διάγραμμα της ΔG_{unfold} συναρτήσει της θερμοκρασίας είναι μια παραβολική καμπύλη, η οποία διέρχεται από το 0 τόσο σε χαμηλή, όσο και σε υψηλή θερμοκρασία και παίρνει μέγιστη τιμή, όταν η θερμοκρασία έχει τιμή περίπου 275K. Τα σημεία στα οποία $\Delta G_{\text{unfold}} = 0$ ορίζουν τις θερμοκρασίες αποδιάταξης για μια πρωτεΐνη κάτω από συγκεκριμένες συνθήκες [39].

Η ελεύθερη ενέργεια σταθεροποίησης ΔG είναι συνάρτηση θερμοδυναμικών μεταβλητών όπως θερμοκρασία, πίεση, συγκέντρωση αποδιατακτικών παραγόντων, pH, πρόσδεση ουσιών μεγαλύτερης συγγένειας στην αποδιατεταγμένη μορφή, ιοντική ισχύ κλπ. Η σταθερότητα της πρωτεΐνης πρέπει να επιτρέπει όχι μόνο την αναδίπλωση και διατήρηση της μορφής της αλλά και την ευελιξία στην στερεοδιάταξη της προκειμένου να εκτελούνται οι λειτουργίες της. Η τρισδιάστατη δομή είναι η διάταξη των πρωτεϊνικών μορίων στον χώρο η οποία καθορίζει την λειτουργία τους. Στην διαμόρφωση της δομής αυτής συμβάλλουν και ασθενείς δυνάμεις όπως δεσμοί υδρογόνου, Van der Waals, ηλεκτροστατικές και υδρόφοβες αλληλεπιδράσεις. Αξιοσημείωτο είναι το γεγονός ότι παρόλο που η ενέργεια των μεμονωμένων δεσμών μεταξύ ατόμων είναι μικρή (0,1-12Kcal/mol) η συνολική ενέργεια των δεσμών είναι εξαιρετικά μεγάλη, δεδομένου ότι προκύπτει ως το άθροισμα των ενεργειών των ατόμων που την απαρτίζουν.

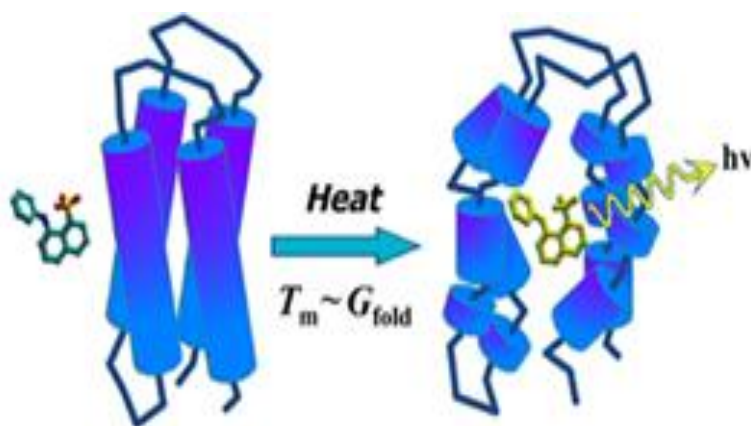
Πρωτεϊνικά μόρια μεγάλου μοριακού βάρους έχει παρατηρηθεί ότι κατά την επίδραση αποδιατακτικών παραγόντων, καταβυθίζονται σχηματίζοντας ιζήματα καθιστώντας έτσι αδύνατη την αντιστρεψιμότητα της ανάκτησης της αρχικής τρισδιάστατης δομής τους. Αυτό δεν συμβαίνει με πρωτεϊνικά μόρια μικρού μοριακού βάρους, τα οποία κατά την απομάκρυνση του διαλύτη, ανακτούν ξανά την αρχική τρισδιάστατη δομή τους.

Κατά την διάρκεια της μετάβασης από την αποδιατεταγμένη στην διατεταγμένη μορφή, υπάρχει ένα σημείο όπου ο αριθμός των αποδιατεταγμένων μορίων ισούται με τον αριθμό των διατεταγμένων μορίων. Σε αυτό το σημείο η θερμοκρασία ονομάζεται

θερμοκρασία τήξης (melting temperature $-T_m$) ή θερμοκρασία αποδιατάξης (denaturation temperature).



Εικόνα 3 Χαρακτηριστικό διάγραμμα της μετάβασης από την διπλωμένη στην αποδιατεταγμένη μορφή μιας πρωτεΐνης με τη χρήση φθορισμομετρίας και ορισμός της θερμοκρασίας τήξης (T_m)



Εικόνα 4 Χαρακτηριστικό διάγραμμα της ανίχνευσης της μετάβασης από την διπλωμένη στην αποδιατεταγμένη μορφή μιας πρωτεΐνης με τη χρήση φθορίζουσας ουσίας στις μετρήσεις μέσω διαφορικής φθορισμομετρίας

Θερμοσταθερότητα πρωτεϊνών

Με τον όρο θερμοσταθερότητα ορίζουμε το δυναμικό διατήρησης της χημικής και δομικής οντότητας μιας πολυπεπτιδικής αλυσίδας κάτω από ακραίες θερμοκρασιακές συνθήκες. Η θερμοσταθερότητα των πρωτεϊνών από διάφορους ακραιόφιλους οργανισμούς αποτελεί ερευνητικό πόλο έλξης, αλλά οι μοριακοί μηχανισμοί που σχετίζονται με την θερμοσταθερότητα των πρωτεϊνών είναι ακόμη υπό διερεύνηση και προβλέπεται μακρύς δρόμος.

Η βιόσφαιρα του πλανήτη μας, περιέχει οργανισμούς που μπορούν να επιβιώνουν και να αναπαράγονται στις πιο ακραίες συνθήκες [40, 41]. Από τους θερμοπίδακες της Ισλανδίας έως τον παγετώνα Perito Moreno της Αργεντινής, από την Τάφρο των Μαρριανών με τα 1086 bar πίεσης έως την κορυφή του Έβερεστ και από υδροβιότοπους που η τιμή του pH ξεκινούν από pH=1 (όξινοι βιότοποι) και φτάνει έως pH=11 (αλκαλικοί βιότοποι) [25], έχουν καταγραφεί μορφές ζωής. Οι οργανισμοί κατέφεραν έτσι όχι μόνο να επιβιώσουν σε τέτοιες συνθήκες αλλά και να «προσαρμοστούν» σε αυτές τις ακραίες συνθήκες για την μετέπειτα επιβίωση τους. Στην περίπτωση της θερμοκρασίας, της πίεσης, αλλά και των ακραίων τιμών του pH, η προσαρμογή των οργανισμών με στόχο την επιβίωση εικάζεται ότι έχει επέλθει μέσω μεταλλάξεων, ενώ κι εξωγενείς παράγοντες ή κυτταρικά συστατικά (μεταβολίτες) συνετέλεσαν πιθανότατα σε αυτή την προσαρμογή [24]. Οι οργανισμοί που επιβιώνουν σε αυτές τις ακραίες συνθήκες, ονομάζονται ακραιόφιλοι.

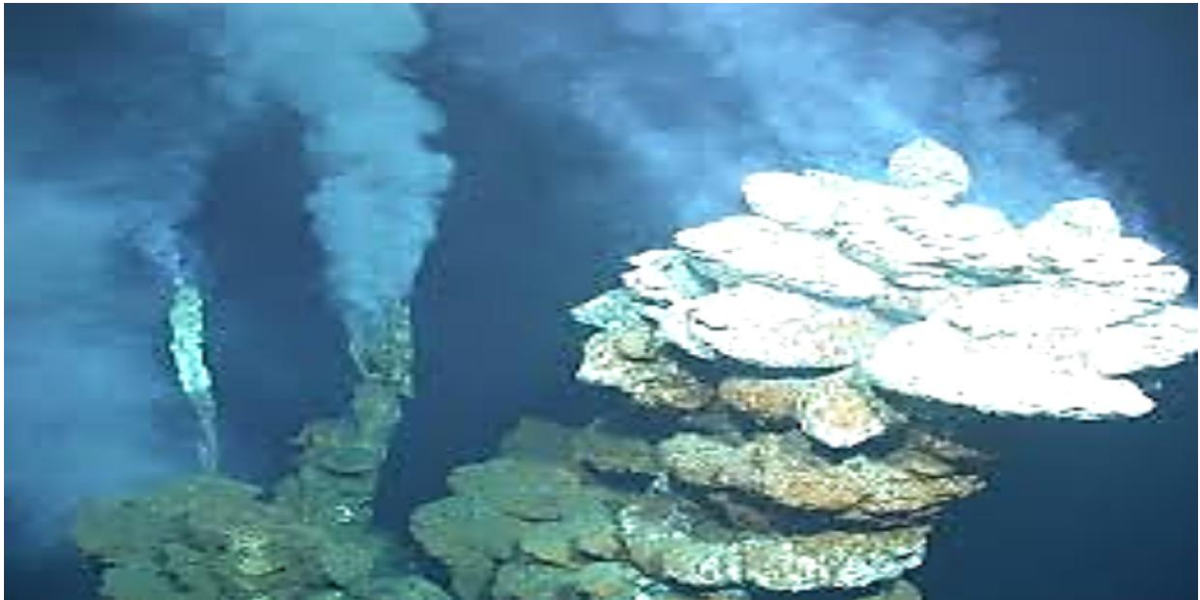
Η ταξινόμηση των ακραιόφιλων γίνεται βάσει του θερμοκρασιακού περιβάλλοντος που επιβιώνουν και έχουν την βέλτιστη ανάπτυξη, και αυτή ορίζεται ως Growth Organism Temperature (GOT). Έτσι προκύπτουν 4 υποκατηγορίες ακραιόφιλων.

1. Ψυχρόφιλοι με $-5^{\circ}\text{C} < \text{GOT} < 20^{\circ}\text{C}$ [42]
2. Μεσόφιλοι με $15^{\circ}\text{C} < \text{GOT} < 45^{\circ}\text{C}$ [43]
3. Θερμόφιλοι με $45^{\circ}\text{C} < \text{GOT} < 80^{\circ}\text{C}$ [27]
4. Υπερθερμόφιλοι με $\text{GOT} > 80^{\circ}\text{C}$ [44]

Η θερμοκρασία που είναι συμβατή με την προκαρυωτική ζωή εκτείνεται από $-20\text{ }^{\circ}\text{C}$ μέχρι και τουλάχιστον $113\text{ }^{\circ}\text{C}$ [39, 45]. Η ποικιλία των θερμοκρασιακών συνθηκών τόσο των Bacteria όσο και των Archaea είναι τεράστια πάνω στη γη [46]. Αν και οι περισσότεροι οργανισμοί αναπτύσσονται σε θερμοκρασίες που κυμαίνονται από 20 έως $50\text{ }^{\circ}\text{C}$ (μεσόφιλοι), μερικά αρχαία (archaea) και όχι λίγα βακτήρια (bacteria) έχουν βρεθεί ότι είναι ικανά να αντέχουν θερμοκρασίες που πλησιάζουν ή και ξεπερνούν τους $100\text{ }^{\circ}\text{C}$ και ανήκουν στους υπερθερμόφιλους οργανισμούς [47].



Εικόνα 5 Φωτογραφία από τα υψίπεδα των Ιμαλαίων οπού απεικονίζεται το ακραιόφιλο περιβάλλον



Εικόνα 6 Φωτογραφία από το βαθύτερό σημείο του πλανήτη όπου παρόλο που εκτιμάται ότι η πίεση σε αυτά τα βάθη φτάνει τα 1086bar υπάρχει ζωή.



Εικόνα 7 Φωτογραφία από τον παγετώνα Perito Moreno στην Παταγονία, δεδομένου ότι οι θερμοκρασίες αγγίζουν τους -70°C έχουν ληφθεί δείγματα οργανισμών.

Τα τελευταία 30 χρόνια που ξεκίνησε η μελέτη της θερμοσταθερότητας των πρωτεϊνών, οι ερευνητές δεν έχουν εντοπίσει άλλες χημικές ιδιότητες σε επίπεδο αμινοξέων ή καινούργιες ομοιοπολικές αλληλεπιδράσεις ή δομικά μοτίβα που να εξηγούν την δυνατότητα αυτών των μορίων να λειτουργούν σε τέτοιες ακραίες συνθήκες [44]. Αντίθετα, παρατηρήθηκαν μικρές επανακατανομές των ίδιων ενδομοριακών αλληλεπιδράσεων που είναι απαραίτητες για την σταθεροποίηση των πρωτεϊνών σε μέτριες συνθήκες να απαιτούνται για την διατήρηση της δομικής ενότητας σε πολύ ζεστά ή πολύ κρύα περιβάλλοντα [28]. Αξιοσημείωτο είναι το

γεγονός ότι ο πεπτιδικός δεσμός που συνδέει τα μεμονωμένα αμινοξικά κατάλοιπα στην πολυπεπτιδική αλυσίδα και καθορίζει την πρωτοταγή δομή των πρωτεϊνών είναι σχετικά αμετάβλητος από το εύρος των θερμοκρασιών που μπορούν να υπάρξουν στην βίωση. Έτσι τα επίπεδα οργάνωσης της δομής που είναι ευαίσθητα στην θερμοκρασία είναι η δευτεροταγής, η τριτοταγής και η τεταρτοταγής δομή των πρωτεϊνών που κατά κύριο λόγο σχηματίζονται και σταθεροποιούνται από μη-ομοιοπολικές αλληλεπιδράσεις οι οποίες καθορίζουν την λειτουργία των πρωτεϊνών σε όλα αυτά τα εύρη θερμοκρασιών [48]. Σε πρωτεΐνες που είναι προσαρμοσμένες σε διαφορετικά θερμοκρασιακά περιβάλλοντα παρατηρούνται διαφοροποιήσεις σε όλη την έκταση της δομής την πρωτεΐνης στο χώρο: στο τοπικό δίπλωμα της πολυπεπτιδικής αλυσίδας, σε δευτεροταγή και υπερδευτεροταγή δομικά στοιχεία, δομικές περιοχές και υπομονάδες [41,49]. Έχει βρεθεί ότι η θερμική σταθερότητα επιτυγχάνεται κυρίως από μια σειρά μικρών πρόσθετων βελτιώσεων σε πολλές περιοχές του μακρομορίου χωρίς όμως αξιοσημείωτες αλλαγές στην συνολική τρισδιάστατη δομή της πρωτεΐνης [40,50], που επηρεάζουν τις θερμοδυναμικές ιδιότητες των πρωτεϊνών [42,51]. Οι θερμοφιλικές πρωτεΐνες διατηρούν το βασικό και χαρακτηριστικό για την συγκεκριμένη αυτή πρωτεϊνική οικογένεια δίπλωμα και μια σχετική ευελιξία ώστε να παραμείνουν λειτουργικές σε αυτή την θερμοκρασία [44, 52]. Αυτό εικάζεται ότι επιτυγχάνεται με αλλαγές στην αμινοξική σύσταση των πρωτεϊνών και κάποιες δομικές τροποποιήσεις μέσω της αλλαγής κάποιων αμινοξέων [24,28]. Έτσι η μοριακή προσαρμογή σε ακραίες θερμοκρασίες εξαρτάται αφενός από αμοιβαίες αλλαγές στην τοπική και συνολική κατανομή των αμινοξέων στην αλληλουχία ενός πρωτεϊνικού μορίου και αφετέρου στους άλλους παράγοντες που ενδεχομένως συμβάλλουν στην διατήρηση των θερμοφίλων οργανισμών [24,34] και για το λόγο αυτό στους θερμοφίλους οργανισμούς, κατά σύμβαση θεωρούνται όλες οι πρωτεΐνες θερμοσταθερές (αν και η θερμοσταθερότητα διαφέρει λίγο από πρωτεΐνη σε πρωτεΐνη στον ίδιο οργανισμό και στην ίδια θερμοκρασία περιβάλλοντος) [53]. Εν κατακλείδι, μπορούμε να ορίσουμε ότι η θερμοσταθερότητα των πρωτεϊνών είναι το αποτέλεσμα ενδομοριακών αλληλεπιδράσεων που είναι χαρακτηριστικές για κάθε είδος πρωτεϊνικής οικογένειας [34, 37].

Μηχανισμοί θερμοσταθερότητας των πρωτεϊνών ΗΥ

1. Μοριακό πακετάρισμα με υψηλότερη αποδοτικότητα (κυρίως μέσω των Van der Waals αλληλεπιδράσεων), υψηλότερη πυκνότητα πακεταρίσματος του υδρόφοβου πυρήνα, υψηλότερη υδροφοβικότητα και μείωση του όγκου και του αριθμού των κοιλοτήτων ή κενών χώρων .
2. Μείωση των ευέλικτων περιοχών όπως π.χ σταθεροποίηση των βρόγχων μέσω αυξημένου περιεχομένου αυτών σε προλίνη, μείωση του μήκους των βρόγχων μεταξύ στοιχείων δευτεροταγούς δομής και επέκταση των δευτεροταγών δομών όπως π.χ. αυξημένος μήκος α-ελίκων και β-πτυχωτών επιφανειών.
3. Διαφορετική αμινοξική σύσταση, όπως π.χ. μεγαλύτερο ποσοστό προλίνης σε βρόχους .
4. Πρόσθετες ηλεκτροστατικές αλληλεπιδράσεις που είναι αποτέλεσμα περισσότερων ή πιο ισχυρών δεσμών άλατος ή δικτύων δεσμών άλατος .
5. Βελτιωμένους δεσμούς υδρογόνου γεωμετρικά ή βελτιωμένα δίκτυα δεσμών υδρογόνου.
6. Περισσότερες δισουλφιδικές γέφυρες [33], αλλά όχι στα υπερθερμόφιλα που η κυστεΐνη δεν υφίστανται.
7. Μεγαλύτερη ανθεκτικότητα σε χημικές τροποποιήσεις .
8. Υψηλότερο ισοηλεκτρικό σημείο [43]
9. Αντικατάσταση μεγάλων αμινοξικών καταλοίπων με μικρότερα ή/και τοποθέτηση υδροφοβικών καταλοίπων σε θέσεις μερικώς εκτεθειμένες στην επιφάνεια .
10. Διαφορετικές αναλογίες εκτεθειμένων/θαμμένων καταλοίπων και/ή υδρόφοβες/πολικές περιοχές και/ή πολικά/επιφανειακά κατάλοιπα και/ή μικρότερη αναλογία επιφάνειας/όγκο.
11. Λιγότερα θερμοευαίσθητα κατάλοιπα [52, 54].

12. Τάση προς μεγαλύτερες καταστάσεις ολιγομερισμού (ίσως με πιο ισχυρές δυνάμεις μεταξύ δομικών περιοχών και μεταξύ υπομονάδων .
13. Ειδικές τοπικές αλληλεπιδράσεις[55].
14. Μείωση αριθμού ή μήκους των βρόχων (ειδικά των επιφανειακών βρόχων)
βελτίωση των αλληλουχιών των β-στροφών.

2. Στόχοι της διπλωματικής εργασίας

Στο μέρος της εισαγωγής, έγινε μια αποτύπωση των πληροφοριών σχετικά με τις πρωτεΐνες HU, οι οποίες έχουν αναλυθεί μετά από επίπονες και χρονοβόρες διαδικασίες. Στο δεύτερο μισό του περασμένου αιώνα παρόλο που έγιναν πολυάριθμες προσπάθειες, δεν υπήρχε το ανάλογο τεχνολογικό υπόβαθρο ώστε αφενός να μειωθεί ο χρόνος των πειραμάτων και αφετέρου να γίνει επεξεργασία του μεγάλου όγκου πληροφορίας που πρόεκυπτε από τα εν λόγω πειράματα. Η απάντηση δεν άργησε να έρθει. Με την εκθετική αύξηση στο πεδίο της πληροφορικής εμφανίστηκε η πρώιμη τότε Υπολογιστική Βιολογία που εξελίχθηκε σε μερικά χρόνια στην επιστήμη της Βιοπληροφορικής. Η Βιοπληροφορική ως ένα υβρίδιο Βιολογίας και Πληροφορικής έχει καταλυτική επίδραση ως προς την εξοικονόμηση χρόνου από την εξαγωγή και ανάλυση αποτελεσμάτων των *in vitro*, *in vivo* και *in situ* πειραμάτων. Εκτός όμως από αυτό, στις μέρες μας που η ταχύτητα επεξεργασίας έχει εκτοξευτεί, προσφέρει την δυνατότητα να δημιουργηθούν μοντέλα και να εκτελεστούν αναρίθμητα πειράματα προσομοίωσης και κατασκευής θεωρητικών μοντέλων μέσω αλγορίθμων, τα οποία ονομάζονται *in silico*. Τα εν λόγω πειράματα προσφέρουν εξοικονόμηση φυσικών και ανθρώπινων πόρων οι οποίοι μπορούν να καταναλωθούν ποικιλοτρόπως.

Σκοπός της παρούσης διπλωματικής εργασίας είναι η αξιοποίηση των δεδομένων για τις πρωτεΐνες HU με στόχο:

1. Εφαρμογή διαφορετικών βιοπληροφοριακών εργαλείων ώστε να αξιοποιήσουμε τα υφιστάμενα δομικά δεδομένα (π.χ κατατεθειμένες δομές HU πρωτεϊνών μέσω κρυσταλλογραφιών με ακτίνες X σε βάσεις δεδομένων κ.α)
2. Επέκταση των γνωστών δομών μας σε όλες την οικογένεια των HU μέσω μοντελοποίησης βάση ομολογίας.
3. Πλήρης σύγκριση των παραγόμενων δομών σε επίπεδο ενδομοριακών και διαμοριακών επιδράσεων και άλλων δομικών χαρακτηριστικών.

4. Συσχετισμός με θερμοδυναμικά δεδομένα (πειραματικά και υπολογιστικά) και διερεύνηση πιθανών συσχετισμών των τελευταίων με το GOT των βακτηρίων από όπου προέρχονται.
5. Προσπάθεια συσχέτισης δομικών και φυσικών χαρακτηριστικών με τη θερμοσταθερότητα της οικογένειας των HU πρωτεϊνών.

3. Μεθοδολογία

3.1 Βάσεις δεδομένων που χρησιμοποιήθηκαν

Για την επίτευξη των στόχων της παρούσας διπλωματικής έγινε σε πρώτη φάση η δημιουργία βάση δεδομένων με την συλλογή των αμινοξικών ακολουθιών των πρωτεϊνών της οικογενείας HU από τις βάσεις δεδομένων, GenBank και Uniprot.

Η **GenBank** (<http://www.ncbi.nlm.nih.gov/GenBank/index.html>) αποτελεί μια βάση δεδομένων νουκλεοτιδικών αλληλουχιών που είναι ελεύθερα διαθέσιμη στην επιστημονική κοινότητα [57-59]. Βρίσκεται υπό την αιγίδα του Εθνικού Ινστιτούτου Υγείας των Η.Π.Α. Η κύρια πηγή της πληροφορίας που περιέχεται στην GenBank προέρχεται από απευθείας υποβολές δεδομένων (πρωτογενείς καταχωρήσεις) όπως προκύπτουν από πειραματικές διεργασίες διαφόρων ερευνητικών ομάδων. Τα νεοεισερχόμενα δεδομένα υφίστανται επεξεργασία και προστίθενται σχόλια (annotation) για την διευκόλυνση των ερευνητών. Ανά τακτά χρονικά διαστήματα τα ήδη κατατεθειμένα δεδομένα επανεξετάζονται και γίνονται διορθώσεις αν προκύπτουν νέα δεδομένα σχετικά με τις εγγραφές. Η διαδικασία κατάθεσης των δεδομένων μπορεί να πραγματοποιηθεί πολύ γρήγορα μέσω του διαδικτύου με την συμπλήρωση κατάλληλης φόρμας και στη συνέχεια οι υπεύθυνοι της βάσης αναλαμβάνουν το σχολιασμό της εγγραφής και τη δημοσιοποίηση της στην βάση. Έχει διαπιστωθεί ότι ο αριθμός των νουκλεοτιδικών βάσεων που περιέχονται στην GenBank αυξάνεται με γεωμετρική πρόοδο.

NCBI Resources How To Sign in to NCBI

GenBank Protein Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Other

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan 41(D1):D36-42). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file format.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with [Entrez Nucleotide](#), which is divided into three divisions: [CoreNucleotide](#) (the main collection), [dbEST](#) (Expressed Sequence Tags), and [dbGSS](#) (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using [BLAST](#) (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see [BLAST info](#) for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using [NCBI e-utils](#).
- The ASN.1 and flatfile formats are available at NCBI's anonymous FTP server: <ftp://ftp.ncbi.nlm.nih.gov/ncbi-asn1> and <ftp://ftp.ncbi.nlm.nih.gov/genbank>.

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

Confidentiality

Some authors are concerned that the appearance of their data in GenBank prior to publication will compromise their work. GenBank will, upon request, withhold release of new submissions for a specified period of time. A date must be specified; we can not hold a sequence indefinitely pending publication. However, if a paper citing the sequence or accession number is published prior to the specified date, the sequence will be released upon publication. In order to prevent the delay in the appearance of published sequence data, we urge authors to inform us of the appearance of the published data. As soon as it is available, please send the full publication data—all authors, title, journal, volume, pages and date—to the following address: update@ncbi.nlm.nih.gov

Privacy

If you are submitting human sequences to GenBank, do not include any data that could reveal the personal identity of the source. GenBank

GenBank Resources

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)
- [Search GenBank](#)
- [Update GenBank Records](#)

Εικόνα 8 Απεικόνιση της επιφάνειας εργασίας από την GenBank. Στο πάνω μέρος, αριστερά απεικονίζεται η ζητούμενη κατηγορία της ακολουθίας π.χ. Protein, και στο δεξί μέρος είναι κενό, ώστε να περιγράψουμε την αναζήτηση μας, π.χ. "Histone like DNA binding protein".

Στην συνέχεια συλλέχθηκαν δεδομένα από την παραπάνω βάση με την παρακάτω διαδικασία:

1. Αρχικά αναζητήθηκαν σε επίπεδο χρήστη, οι πρωτεΐνες HU, γράφοντας στο πεδίο αναζήτησης της βάσης δεδομένων "Histone like DNA binding protein". Όπως είναι φυσικό όμως, το αποτέλεσμα αυτής της αναζήτησης δεν αντικατοπτρίζει την πραγματική πληροφορία, καθώς μεγάλος όγκος της τελευταίας χάνεται.
2. Στη συνέχεια, τα πρωτογενή δεδομένα συλλέχθηκαν τοπικά στον υπολογιστή. Ακολούθως αναπτύξαμε ένα πρόγραμμα γραμμένο σε γλώσσα προγραμματισμού Perl (perl script). Η γλώσσα Perl επιλέχθηκε αφενός για την φιλικότητα προς τον προγραμματιστή, αφετέρου για την υψηλή ταχύτητα της και τέλος γιατί είναι ιδανική για σαρώσεις που αφορούν ομοιότητες με προκαθορισμένους χαρακτήρες κειμένου (pattern matching) σε μεγάλο όγκο αρχείων.
3. Χρησιμοποιώντας ως είσοδο το αρχείο των αποτελεσμάτων της βάση δεδομένων GenBank, αυτό σαρώθηκε μέσω του Script. Στο κώδικα χρησιμοποιήθηκαν ως Keywords με την διαδικασία του pattern matching, το πεδίο "Locus" και "Accession" προκειμένου να δοθεί το όνομα της καταχώρησης και στην συνέχεια στο πεδίο "Organism", δόθηκαν τα keywords "Bacteria", "Archaea" και "Eukaryota" ώστε να διαχωριστούν οι προκαρυωτικοί και τα αρχαία από τους ευκαρυωτικούς οργανισμούς. Επιπλέον χρησιμοποιήθηκε και η ανάλογη παράμετρος απαρίθμησης προκειμένου να καταμετρηθούν. Ο απόσπασμα του κώδικα σε γλώσσα προγραμματισμού Perl που αναπτύχθηκε και χρησιμοποιήθηκε παρατίθεται παρακάτω:

```

1 #!/usr/bin/perl
2 print "Πλικτρολογιστε το path του arxeiou pou epithimeite....\n";
3 $a=<STDIN>;
4 open FILE,"$a" or die "Cannot open file";
5 open NCBI_IHF_RESULTS, ">NCBI HU RESULTS.txt";
6 $counter=0;
7 $prok=0;
8 $eyk=0;
9 $svar=0;
10 $unk=0;
11 $arc=0;
12 print NCBI_IHF_RESULTS "////Arxeio Apotelesmatwn NCBI-GENBANK HU proteinwn Dimitriou Antonogiannaki///\n\n";
13 while (<FILE>)
14 {
15     if($_ =~/^LOCUS (\S+)/g)
16     {
17         $loc=$1;
18         print NCBI_IHF_RESULTS "$loc\n";
19     }
20     elsif($_ =~/^ACCESSION (\S+)/g)
21     {
22         $acc=$1;
23         print NCBI_IHF_RESULTS "$acc\n";
24     }
25
26     elsif($_ =~/^ ORGANISM (\S+) (\S+)/g)
27     {
28         $tax=$1;
29         $taxa=$2;
30         print NCBI_IHF_RESULTS "$tax $taxa\n";
31     }
32
33     elsif($_ =~/^ (\S+) (\;) (\S+) (\;) (\S+) (\;) (\S+) (\;) (\S+) (\;) (\S+) (\;)/g)
34     {
35         @seq=("$1", "$2", "$3", "$4", "$5", "$6", "$7", "$8", "$9", "$10");
36         print NCBI_IHF_RESULTS "@seq \n";
37
38
39         if($seq[1] eq 'Bacteria')
40         {
41             $prok++;
42         }
43         elsif($seq[1] eq 'Eukaryota')
44         {

```

Εικόνα 9 Χαρακτηριστικό τμήμα του κώδικα Perl που αναπτύχθηκε για τις ανάγκες της παρούσας εργασίας. Ευδιάκριτα είναι τα Keywords, η μεταβλητή του μετρητή, καθώς και το όνομα του αρχείου εξόδου των αποτελεσμάτων.


```

38
39     if($seq[$i] eq 'Bacteria')
40     {
41         $prok++;
42     }
43     elsif($seq[$i] eq 'Eukaryota')
44     {
45
46         $eyk++;
47         print NCBI_IHF_RESULTS "o organismos anhkei sta $seq[$i] \n";
48     }
49     elsif($seq[$i] eq 'Archaea')
50     {
51
52         $sarc++;
53     }
54     else
55     {
56         $dif=$seq[$i];
57         push(@organism,$dif);
58         $svar++;
59     }
60
61
62 }
63 elsif($_ =~ /^(\s+)/g)
64 {
65     $send=$1;
66     if($send eq '/')
67     {
68         $counter++;
69         print NCBI_IHF_RESULTS"$counter\n";
70     }
71 }
72 }
73
74 }
75
76 $svar=$counter-$prok-$eyk;
77 $spro=$prok+$svar;
78 print NCBI_IHF_RESULTS "\n\nVrethikan $spro prokarywtikoi $sarc organismoi pou anikoun sta archaia kai $eyk eykarywtikoi organismoi\n";
79
80
81

```

Εικόνα 9 (συνέχεια) Χαρακτηριστικό τμήμα του κώδικα Perl που αναπτύχθηκε για τις ανάγκες της παρούσας εργασίας. Ευδιάκριτα είναι τα Keywords, η μεταβλητή του μετρητή, καθώς και το όνομα του αρχείου εξόδου των αποτελεσμάτων.

```
345866 AAW04719
345867 AAW04719
345868 161913
345869 AAT16432
345870 AAT16432
345871 161914
345872 AF320916_1
345873 AAK08220
345874 Lactococcus lactis
345875 Bacteria ; Firmicutes ; Bacilli ; Lactobacillales ; Streptococcaceae ;
345876 161915
345877 AF320915_3
345878 AAK08219
345879 Lactococcus lactis
345880 Bacteria ; Firmicutes ; Bacilli ; Lactobacillales ; Streptococcaceae ;
345881 161916
345882 AJL25163
345883 AJL25163
345884 161917
345885 AF320913_3
345886 AAK08213
345887 Lactococcus lactis
345888 Bacteria ; Firmicutes ; Bacilli ; Lactobacillales ; Streptococcaceae ;
345889 161918
345890 AF320914_3
345891 AAK08216
345892 Lactococcus lactis
345893 Bacteria ; Firmicutes ; Bacilli ; Lactobacillales ; Streptococcaceae ;
345894 161919
345895 ACJ92684
345896 ACJ92684
345897
345898
345899
345900 Vrethikan 120009 prokarywtikoi 391 organismoi pou anikoun sta archaia kai 38399 eukarywtikoi
345901
```

Normal text file length: 6.514.135 lines: 345.901 Ln: 345.900 Col: 81 Sel: 0|0 Windows (CR LF) UTF-8 INS

Εικόνα 10. Αποτέλεσμα αρχείου εξόδου NcbihuResults για τον διαχωρισμό των πρωτεϊνών HU σε προκαρυωτικούς, ευκαρυωτικούς και οργανισμούς που ανήκουν στα αρχαία.

Όσον αφορά τη βάση δεδομένων πρωτεϊνικών ακολουθιών **UniProt** (<http://www.uniprot.org>) [60, 61], αποτελεί στην ουσία μια προσπάθεια ενοποίησης των βάσεων δεδομένων πρωτεϊνικών ακολουθιών **SwissProt** [62], **TrEMBL** [63] και **PIR-PSD** [64]. Τα ιδρυτήρια που συμμετέχουν σε αυτή την προσπάθεια είναι το European Bioinformatics Institute (EBI), το Swiss Institute of Bioinformatics (SIB) και η Protein Information Resource (PIR). Ο βασικός φορέας χρηματοδότησης της προσπάθειας αυτής είναι το Εθνικό Ινστιτούτο Υγείας (NIH) των ΗΠΑ. Ο στόχος αυτής της βάσης είναι να συγκεντρωθεί όσο το δυνατόν περισσότερη πληροφορία για τις πρωτεϊνικές ακολουθίες η οποία ήταν διασπαρμένη στις παραπάνω βάσεις εμπλουτίζοντας το σχολιασμό κάθε εγγραφής και ελαχιστοποιώντας τα λάθη που μπορεί να περιέχονται.

Η **SWISS-PROT** (<http://www.expasy.ch/sprot/>), είναι μια βάση δεδομένων πρωτεϊνικών ακολουθιών που ιδρύθηκε το 1986 και στις μέρες μας συντηρείται από το Ελβετικό Ινστιτούτο Βιοπληροφορικής (Swiss Institute of Bioinformatics) σε συνεργασία με το Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής (European Bioinformatics Institute). Στις καταχωρήσεις της SWISSPROT εκτός από την ακολουθία υπάρχουν και συμπληρωματικά σχόλια όπως, βιβλιογραφικές αναφορές, γενικά στοιχεία δευτεροταγούς δομής, σύνδεσμοι σε άλλες βάσεις δεδομένων σχετικές με κάθε εγγραφή καθώς και σημειώσεις για τη βιολογική λειτουργία (αν είναι γνωστή) και άλλες χρήσιμες πληροφορίες.

Η βάση δεδομένων πρωτεϊνικών ακολουθιών **TrEMBL** περιέχει εγγραφές που είναι αποτέλεσμα της μετάφρασης εγγραφών της EMBL (nucleotide sequence database) σε πολυπεπτιδικές αλυσίδες. Όταν οι εγγραφές υποστούν περαιτέρω έλεγχο (δευτερογενείς καταχωρήσεις) ενσωματώνονται στην SWISS-PROT. Είναι κατανοητό ότι τα δεδομένα της βάσης TrEMBL λόγω του τρόπου με τον οποίο παράγονται δεν μπορούν να εκτιμηθούν ως πολύ αξιόπιστα. Για το λόγο αυτό χρησιμοποιήσαμε μόνο τα δεδομένα της βάσης UniProt που αναφέρονται στο κομμάτι της SwissProt. Αν και η βάση SwissProt θεωρείται υψηλής αξιοπιστίας, εντούτοις η εμφάνιση λαθών στις καταχωρήσεις δεν είναι κάτι σπάνιο. Ένας προσεκτικότερος

έλεγχος, που αφορούσε το μήκος των αμινοξικών αλληλουχιών, ήταν αναγκαίος και ανέδειξε πιθανά λάθη σε καταχωρήσεις.

Αφού μελετήσαμε τις βάσεις δεδομένων, μπορούμε τώρα να περιγράψουμε διεξοδικότερα τον τρόπο ανάκτησης πρωτεϊνικών ακολουθιών για την βάση δεδομένων Uniprot.

The screenshot shows the UniProtKB search results page. The search query is "histone like dna binding protein". The results are filtered to show 109,441 entries. A table of results is displayed, including entry names, protein names, gene names, organisms, and lengths. A sidebar on the left offers various filters such as "Reviewed (5,015)", "Unreviewed (104,426)", "Popular organisms", "Subcellular location", and "Search terms".

Entry	Entry name	Protein names	Gene names	Organism	Length
P9WMK7	DBH_MYCTU	DNA-binding protein HU homolog	hup hlp, hupB, lbp21, Rv2986c, MTCY349.01	Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)	214
P0ACF8	HNS_ECOLI	DNA-binding protein H-NS	hns bglY, cur, drdX, hnsA, msyA	Escherichia coli (strain K12)	137
P0A1S2	HNS_SALTY	DNA-binding protein H-NS	hns hnsA, osmZ, STM1751	Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)	137
P03685	NP_BPPH2	Histone-like protein p6	6	Bacillus phage phi29 (Bacteriophage phi-29)	104
P0AEU7	SKP_ECOLI	Chaperone protein Skp	skp hlpA, ompH, b0178, JW0173	Escherichia coli (strain K12)	161
Q16576	RBBP7_HUMAN	Histone-binding protein RBBP7	RBBP7 RBAP46	Homo sapiens (Human)	425
Q60973	RBBP7_MOUSE	Histone-binding protein RBBP7	Rbbp7 Rbap46	Mus musculus (Mouse)	425
Q92830	KAT2A_HUMAN	Histone acetyltransferase KAT2A	KAT2A GCNS, GCNSL2	Homo sapiens (Human)	837
Q9H9B1	EHMT1_HUMAN	Histone-lysine N-methyltransferase ...	EHMT1 EUHMTASE1, GLP, KIAA1876, KMT1D	Homo sapiens (Human)	1,298
Q9JHD2	KAT2A_MOUSE	Histone acetyltransferase KAT2A	Kat2a Gcn5l2	Mus musculus (Mouse)	830
Q14493	SLBP_HUMAN	Histone RNA hairpin-binding protein	SLBP HBP	Homo sapiens (Human)	270
P34470	HCP3_CAEEL	Histone H3-like centromeric protein...	hcp-3 F58A4.3	Caenorhabditis elegans	288
P97440	SLBP_MOUSE	Histone RNA hairpin-binding protein	Slbp Hbp	Mus musculus (Mouse)	

Εικόνα 12 Επιφάνεια εργασίας από την Uniprot. Στο πάνω μέρος, απεικονίζεται η ζητούμενη κατηγορία της ακολουθίας ώστε να περιγράψουμε την αναζήτηση μας, π.χ. histone like DNA binding protein.

Συνεπώς, σε αναλογία με την GenBank, έτσι και εδώ αναζητήθηκαν σε επίπεδο χρηστή, οι πρωτεΐνες HU, γράφοντας την οικογένεια στο πεδίο αναζήτησης της βάσης δεδομένων. Ακολούθως δημιουργήθηκε ένα πρόγραμμα γραμμένο σε γλώσσα προγραμματισμού Perl (perl script). Χρησιμοποιώντας ως είσοδο το αρχείο των αποτελεσμάτων της βάση δεδομένων Uniprot, αυτό σαρώθηκε μέσω του Script. Στο κώδικα χρησιμοποιήθηκαν ως Keywords με την διαδικασία του pattern matching, το πεδίο "ID" και "AC" προκειμένου να δοθεί το όνομα της καταχώρησης και στην συνέχεια στο πεδίο "OC", δόθηκαν τα keywords "Bacteria", "Eukaryota" και "Archea" ώστε να διαχωριστούν οι πρωτεΐνες που ανήκουν σε προκαρυωτικούς, σε ευκαρυωτικούς οργανισμούς αλλά και σε ιούς. Ακολούθως στο πεδίο "ID" χρησιμοποιήθηκε και το Keyword "Reviewed" προκειμένου να εξακριβωθεί αν πρόκειται για δευτερογενή αρχεία (reviewed) ή για πρωτογενή (unreviewed). Τέλος, όπως ήταν αναμενόμενο χρησιμοποιήθηκε και η ανάλογη παράμετρος απαρίθμησης προκειμένου να καταμετρηθούν τα αποτελέσματα.

```

4  open FILE,"$a" or die "Cannot open file";
5  open UNIPROT_HU_RESULTS, ">RESULTS.txt";
6  $counter=0;
7  $seul=0;
8  $spre=0;
9  $mic=0;
10 $arc=0;
11 $rev=0;
12 $unrev=0;
13 print UNIPROT_HU_RESULTS "////Αρχείο Αποτελεσμάτων UNIPROT HU proteinων Dimitriou Antonogiannaki//\n\n";
14 while (<FILE>)
15 {
16     if ($ =~ /^ID (\S+) (\S+) (\S+)/g)
17     {
18         $review=$3;
19         #print "$review";
20         #print "$rev";
21         if ($3 eq 'Reviewed:')
22         {
23             $rev++;
24         }
25         else
26         {
27             $unrev++;
28         }
29         @id=("$1","$2","$3");
30         print UNIPROT_HU_RESULTS "@id\n";
31     }
32     elsif ($ =~ /^AC (\S+) (\S+)/g)
33     {
34         @ac=("$1","$2");
35         print UNIPROT_HU_RESULTS "@ac\n";
36     }
37     elsif ($ =~ /^OC (\S+) (\S+) (\S+)/g)
38     {
39         @oc=("$1","$2","$3","$4");
40         print UNIPROT_HU_RESULTS "@oc\n";
41         if ($oc[1] eq 'Bacteria:')
42         {
43             $arc++;
44         }
45     }
46 }

```

Εικόνα 13. Χαρακτηριστικό τμήμα του κώδικα Perl που χρησιμοποιήθηκε. Ευδιάκριτα είναι τα Keywords που επιλέχθηκαν στον εν λόγω κώδικα.

```

40 {
41     @oc=("$1", "$2", "$3", "$4");
42     print UNIPROT_HU_RESULTS "@oc\n";
43     if($oc[$1] eq 'Bacteria;')
44     {
45         $pro++;
46     }
47     elsif ($oc[$1] eq 'Eukaryota;')
48     {
49
50         $euk++;
51         print UNIPROT_HU_RESULTS "Η hu anikei se $oc[$1] organismo kai einai h $count kataxwrisi \n";
52     }
53
54     elsif ($oc[$1] eq 'Archea;')
55     {
56
57         $arc++;
58     }
59
60 }
61 elsif($_ =~ /^OC (\S+) (\S+) (\S+) (\S+)/g)
62 {
63     @oca=("$1", "$2", "$3", "$4");
64     print UNIPROT_HU_RESULTS "@oca\n";
65 }
66
67 elsif($_ =~ /^(\S+)/g)
68 {
69     $end=$1;
70     if($end eq '/')
71     {
72         $counter++;
73         print UNIPROT_HU_RESULTS "$counter\n";
74     }
75 }
76 }
77
78 }
79 $mic=$counter-$pro-$euk-$vir;
80 print UNIPROT_HU_RESULTS "\n\nTa Reviewed arxeia einai $rev kai ta Unreviewed einai $unrev\n\nTa prokaryotika einai $pro enw

```

Εικόνα 13 (συνέχεια) Χαρακτηριστικό τμήμα του κώδικα Perl που χρησιμοποιήθηκε. Ευδιάκριτα είναι τα Keywords, το όνομα του αρχείου εξόδου καθώς και η διαμόρφωση των αποτελεσμάτων.

```

Eπεξεργασία:bt x RESULTS bt x Uniprothu.pl x Uniprothu.pl x RESULTS HU UNIPROT.bt x
33925 A0A0Q0CUK1_9PSED Unreviewed;
33926 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
33927 109430
33928 A0A176U9F8_9FIRM Unreviewed;
33929 Bacteria; Firmicutes; Clostridia; Clostridiales.
33930 109431
33931 A0A0D5WVX8_9ENTR Unreviewed;
33932 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
33933 109432
33934 I6ZE89_MYCWM Unreviewed;
33935 Bacteria; Tenericutes; Mollicutes; Mycoplasmataceae;
33936 109433
33937 X8I4N3_9FUSO Unreviewed;
33938 Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
33939 109434
33940 A0A142X6R9_9PLAN Unreviewed;
33941 Bacteria; Planctomycetes; Planctomycetia; Planctomycetales;
33942 109435
33943 A6H060_FLAPJ Unreviewed;
33944 Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales;
33945 109436
33946 X6PXX0_9BACT Unreviewed;
33947 Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
33948 109437
33949 F0IG78_9FLAO Unreviewed;
33950 Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales;
33951 109438
33952 A0A0G1XHQ6_9BACT Unreviewed;
33953 109439
33954 X6PYH1_9BACT Unreviewed;
33955 Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
33956 109440
33957 A0A1B2N5X5_9GAMM Unreviewed;
33958 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
33959 109441
33960 A0A133Y0C8_9BACT Unreviewed;
33961 Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales.
33962
33963
33964
33965 Ta Reviewed arxia einai 5015 kai ta Unreviewed einai 104426
33966 Ta prokaryotika einai 56972 enw ta eukaryotika einai 50631 kai ta archaia einai 1101 kai ta axarakthrista 1838
Normal text file length:1.176.870 lines:33.966 Ln:33.927 Col:7 Sel:0|0 Windows(CRLF) UTF-8 INS

```

Εικόνα 14 Αποτέλεσμα αρχείου εξόδου UniprothuResults για τον διαχωρισμό των πρωτεϊνών HU από προκαρυωτικούς σε ευκαρυωτικούς οργανισμούς και αρχαία.

```

Ene&grypaic.bat RESULTS.txt Uniprothu.pl Uniprothu.pl RESULTS HU UNPROT.txt
27880 5009
27881 UBN1_HUMAN Reviewed;
27882 Q9NPG3; B7Z6D3;
27883 Eukaryota; Metazoa; Chordata; Craniata;
27884 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27885 Mammalia; Eutheria; Euarchontoglires; Primates;
27886 5010
27887 UBP22_MOUSE Reviewed;
27888 Q5DU02; Q3TU34;
27889 Eukaryota; Metazoa; Chordata; Craniata;
27890 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27891 Mammalia; Eutheria; Euarchontoglires; Glires;
27892 Muroidea; Muridae; Murinae; Mus;
27893 5011
27894 UBP16_MOUSE Reviewed;
27895 Q99LG0; G5E860;
27896 Eukaryota; Metazoa; Chordata; Craniata;
27897 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27898 Mammalia; Eutheria; Euarchontoglires; Glires;
27899 Muroidea; Muridae; Murinae; Mus;
27900 5012
27901 UBP16_RAT Reviewed;
27902 Eukaryota; Metazoa; Chordata; Craniata;
27903 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27904 Mammalia; Eutheria; Euarchontoglires; Glires;
27905 Muroidea; Muridae; Murinae; Rattus.
27906 5013
27907 UBP36_HUMAN Reviewed;
27908 Q9P275; Q05C98;
27909 Eukaryota; Metazoa; Chordata; Craniata;
27910 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27911 Mammalia; Eutheria; Euarchontoglires; Primates;
27912 5014
27913 UHRF1_XENTR Reviewed;
27914 Eukaryota; Metazoa; Chordata; Craniata;
27915 H hu anikei se Eukaryota; organismo kai einai h kataxwrisi
27916 Amphibia; Batrachia; Anura; Pipioidea;
27917 5015
27918
27919
27920 Ta Reviewed arxia einai 5015 kai ta Unreviewed einai 0
27921 Ta prokaryotika einai 714 enw ta eukaryotika einai 4124 kai ta arxia einai 0 kai ta axarakthrista 177
Normal text file length:1.034.139 lines:27.921 Ln:1 Col:1 Sel:0|0 Windows(CRLF) ANSI INS

```

Εικόνα 15 Αποτέλεσμα αρχείου εξόδου UniprothuResults για τον διαχωρισμό των Reviewed πρωτεϊνών HU από προκαρυωτικούς, ευκαρυωτικούς οργανισμούς καθώς και αρχαία.

3.2 Κατασκευή και αξιολόγηση πρωτεϊνικών μοντέλων μέσω ομολογίας με πρότυπες δομές πρωτεϊνών (Homology Modeling)

Η κρυσταλλογραφία ακτινών X και η φασματοσκοπία NMR είναι οι βασικότερες μέθοδοι που μας προσφέρουν αναλυτική δομική πληροφορία. Επειδή οι μέθοδοι αυτές απαιτούν κρυστάλλωση ή οποία δεν είναι πάντα εφικτή ή για το NMR το μέγεθος του βιομορίου είναι σημαντικός παρεμποδιστικός παράγοντας η παραγωγή πειραματικών δεδομένων δεν είναι πάντοτε δυνατή. Ελλείψει πειραματικών δεδομένων, η κατασκευή μοντέλων με βάση την γνωστή τρισδιάστατη δομή ομόλογων πρωτεϊνών είναι ο αξιόπιστος τρόπος για την συλλογή δομικών πληροφοριών. Συγκρίσεις τρισδιάστατων δομών ομόλογων πρωτεϊνών έχουν δείξει ότι η τρισδιάστατη δομή κατά την διάρκεια της εξέλιξης έχει συντηρηθεί καλύτερα από την πρωτοταγή δομή, ενώ υπάρχουν βάσεις δεδομένων που περιέχουν τα αποτελέσματα των συγκρίσεων μεταξύ των πειραματικά εντοπισμένων δομών. Όλα αυτά υποδεικνύουν την δυνατότητα της κατασκευής προτύπων βάσει ομολογίας.

Με τον όρο **δημιουργία προτύπων πρωτεϊνών με ομολογία (homology modeling)** αναφερόμαστε στην πρόβλεψη της τριτοταγούς δομής μιας άγνωστης πρωτεΐνης με βάση την πειραματικά προσδιορισμένη δομή της ομόλογης. Ο όρος ομολογία στις πρωτεΐνες έχει ποικίλους ορισμούς. Ο όρος **ομόλογες πρωτεΐνες** εξελικτικά εντοπίζει αυτές τις ακολουθίες που έχουν κοινό πρόγονο. Ομόλογες (παρόμοιες) δομικά πρωτεΐνες είναι αυτές που ανεξάρτητα από την προέλευση τους έχουν περισσότερο από 30% σύμπτωση αμινοξικών καταλοίπων στην πρωτοταγή τους δομή.

Που βασίζεται η δημιουργία προτύπων πρωτεϊνών βάσει ομολογίας πρωτοταγούς δομής (homology modeling)

Η δημιουργία μοντέλων πρωτεϊνικών δομών βάσει ομολογίας με γνωστές πειραματικά προσδιορισμένες δομές (homology modeling) μπορεί να είναι επιτυχής υπό την προϋπόθεση η ομοιότητα σε πρωτοταγή δομή να είναι $> 25\%$. Έχει δειχτεί ότι πρωτεΐνες με ποσοστό ομοιότητας της τάξης του 20% έχουν μεγάλες δομικές διαφορές διότι το ποσοστό αυτό αγγίζει τα όρια της τυχαιότητας. Όσο πιο υψηλή είναι η διακριτικότητα μίας λυμένης κρυσταλλογραφικά δομής που θα χρησιμοποιηθεί ως μήτρα τόσο πιο ακριβές είναι και το παραγόμενο μοντέλο του. Η μέθοδος περιλαμβάνει:

1. Πρόβλεψη της δευτεροταγούς δομής
2. Στοιχισή ακολουθιών
3. Έλεγχος δομικών κατηγοριών και οικογενειών
4. Υπέρθεση δομών και αντικατάσταση δομικών στοιχείων

Η ομόλογη προτυποποίηση είναι μία κλιμακωτή-σταδιακή διεργασία. Σε όλα τα στάδια πρέπει να γίνουν επιλογές που ποτέ δεν είναι γνωστό αν είναι οι βέλτιστες και έτσι ένα μεγάλο μέρος από την διαδικασία αφορά την λήψη αποφάσεων από παρόμοιες επιλογές. Επίσης όλα τα μοντέλα που κατασκευάζονται περιέχουν λάθη. Έτσι η πιστοποίηση των δεδομένων που χρησιμοποιούνται αλλά και παρέχονται από τις μεθόδους προτυποποίησης είναι απαραίτητη για την κατάληξη σε ένα αξιόπιστο αποτέλεσμα.

Εισαγωγικά στοιχεία για την PDB

Στα πλαίσια των στόχων της διπλωματικής εργασίας, σε πρώτη φάση έπρεπε να ελέγξουμε τον αριθμό των πειραματικά προσδιορισμένων δομών της πρωτεΐνης HU που υπάρχουν κατατεθειμένες στη βάση δεδομένων PDB (Protein Data Bank). Η Protein Data Bank (www.rcsb.org) αποτελεί τη μοναδική βάση όπου είναι κατατεθειμένες οι τρισδιάστατες δομές βιολογικών μακρομορίων. Ιδρύθηκε το 1971 στα Brookhaven National Laboratories (BNL) των ΗΠΑ και περιελάμβανε 7 δομές μακρομορίων όπως αυτές προέκυψαν από κρυσταλλογραφικές μελέτες. Ο ρυθμός αύξησης των νέων εγγραφών στη δεκαετία του '70 ήταν πολύ μικρός. Από το 1980 και μετά λόγω της τεχνολογικής εξέλιξης σε κάθε στάδιο του προσδιορισμού δομών βιομορίων, ο ρυθμός προσθήκης δεδομένων στην PDB αυξήθηκε εκθετικά. Πλέον στην βάση περιλαμβάνονται και δομές όπως προκύπτουν με φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (NMR). Οι εγγραφές στην PDB εκτός από τις συντεταγμένες των ατόμων που απαρτίζουν τη δομή περιλαμβάνουν και επιπρόσθετα βοηθητικά στοιχεία όπως βιβλιογραφικές αναφορές, λεπτομέρειες για τον προσδιορισμό της δομής καθώς και άλλα στοιχεία που προκύπτουν από τη συγκεκριμένη δομή. Κάθε δομή προτού διατεθεί στο κοινό υφίσταται έλεγχο για την ορθότητα της με τη χρήση ειδικού λογισμικού. Στη συνέχεια εφόσον περάσει τις δοκιμές με επιτυχία αποκτά ένα χαρακτηριστικό κωδικό και προστίθεται στη βάση. Ο έλεγχος ορθότητας της δομής γίνεται μέσω του ADIT Validation Server (διακομιστή) (<http://deposit.rcsb.org/validate/>).

Παρακάτω θα αναλυθεί η διαδικασία καθώς και τα βήματα της δημιουργίας προτύπων πρωτεϊνών με ομολογία (homology modeling). Ως παράδειγμα παρατίθενται τα στιγμιότυπα δημιουργίας *in silico* μοντέλου και ομοδιμερές της πρωτεϊνικής ακολουθίας με αριθμό καταχώρησης (AccessionID) BG8X11 της βάσης δεδομένων Uniprot η οποία ανήκει στην οικογένεια των HU μέσω του Pymol 2.0, το οποίο είναι ένα προσθετό εργαλείο του γνωστού Pymol, και μας δίνει την δυνατότητα διεπαφής με τα γνωστά προγράμματα Βιοπληροφορικής (Blast, Modeller, ClustalW).

Προϋποθέσεις αποδοχής ενός μοντέλου

Το ποσοστό ομολογίας μεταξύ της άγνωστης πρωτεΐνης στόχος (target) και πρωτεΐνης με γνωστή στερεοδομή, γνωστή ως μήτρα (template) πρέπει να είναι τουλάχιστον 25-30%. Ελέγχεται λοιπόν η πρωτεΐνης στόχος μέσω των βάσεων δεδομένων των ακολουθιών με προγράμματα όπως το FASTA ή BLAST που είναι ιδανικά για τον εντοπισμό πρωτεϊνικών ομολογιών. Συστήνεται η αναζήτηση να γίνει και στην PDB ώστε να αποφανθούμε αν υπάρχει κάποια δομή που να πληρεί τα κριτήρια ώστε να λειτουργήσει ως δομή - οδηγός (template) για την κατασκευή του πρωτεϊνικού μοντέλου στόχου. Εάν το ποσοστό ομολογίας είναι μικρότερο από 25%, τότε απαιτούνται άλλες τεχνικές για την κατασκευή του μοντέλου. Βασικό κριτήριο για αυτό θα πρέπει να είναι ο συνδυασμός της χαμηλότερης δυνατής τιμής E-Value με την μέγιστη δυνατή τιμή (query coverage) και φυσικά πάνω από 30% στην Ομοιότητα (identity).

Στοίχιση των ακολουθιών

Η στοίχιση των αμινοξικών ακολουθιών επιτυγχάνεται με διάφορα προγράμματα (Align, ClustalW) που βασίζονται σε ένα πίνακα με πιθανότητες μεταλλαγής ενός αμινοξέος σε ένα άλλο σύμφωνα με παρατηρούμενες αλλαγές σε ένα ή περισσότερα γονιδιώματα. Στους πίνακες αυτούς μπορούν να συμπεριληφθούν και προτιμήσεις που βασίζονται σε φυσικά ή χημικά χαρακτηριστικά των αμινοξέων (υδροφοβικότητα, φορτίο, όγκος κλπ.). Η ευθυγράμμιση των ακολουθιών μπορεί να βελτιστοποιηθεί με την χρήση πληροφοριών από πολλές ομόλογες ευθυγραμμίσεις. Η μεθοδολογία που ακολουθείται είτε βασίζεται σε προφίλ ακολουθιών, είτε σε τεχνολογία «ύφανσης» (threading). Αυτή γίνεται δια μέσου βάσεων δεδομένων που επιτρέπουν τον εύκολο, ευέλικτο και γρήγορο εντοπισμό στερεοδομικών πληροφοριών και είναι βασική για τον εντοπισμό του ομόλογου προτύπου.

Η αμινοξική ακολουθία της πρωτεΐνης - στόχου, αφού ελεγχθεί για πιθανά λάθη και κατόπιν εισήχθη και η ακολουθία της δομής - οδηγού. Μέσω της διεπαφής του προγράμματος ClustalW, το οποίο εμπεριέχεται στην λίστα εργαλείων του Pymod

2.0, εκτελέστηκε στοίχιση μεταξύ των δύο ανωτέρω ακολουθιών. Το CLUSTALW είναι το πιο διαδεδομένο πρόγραμμα στοίχισης βιολογικών ακολουθιών. Το πρόγραμμα χρησιμοποιεί πίνακές σκορ (scoring matrix) για να κάνει σταδιακή στοίχιση πρωτεϊνικών ή νουκλεοτιδικών (DNA) ακολουθιών. Στην δημιουργία μοντέλων χρησιμοποιήθηκαν οι προκαθορισμένες (default) ρυθμίσεις όσο αφορά την επιλογή πίνακα σκορ (scoring matrix selection) - στην περίπτωση μας BLOSUM - , και τις ποινές για το κενό (gap open in penalty) και την επέκταση (gap extension penalty) - στην περίπτωση μας με τιμές 10 και 0.2 αντίστοιχα -.

Κατασκευή μοντέλου

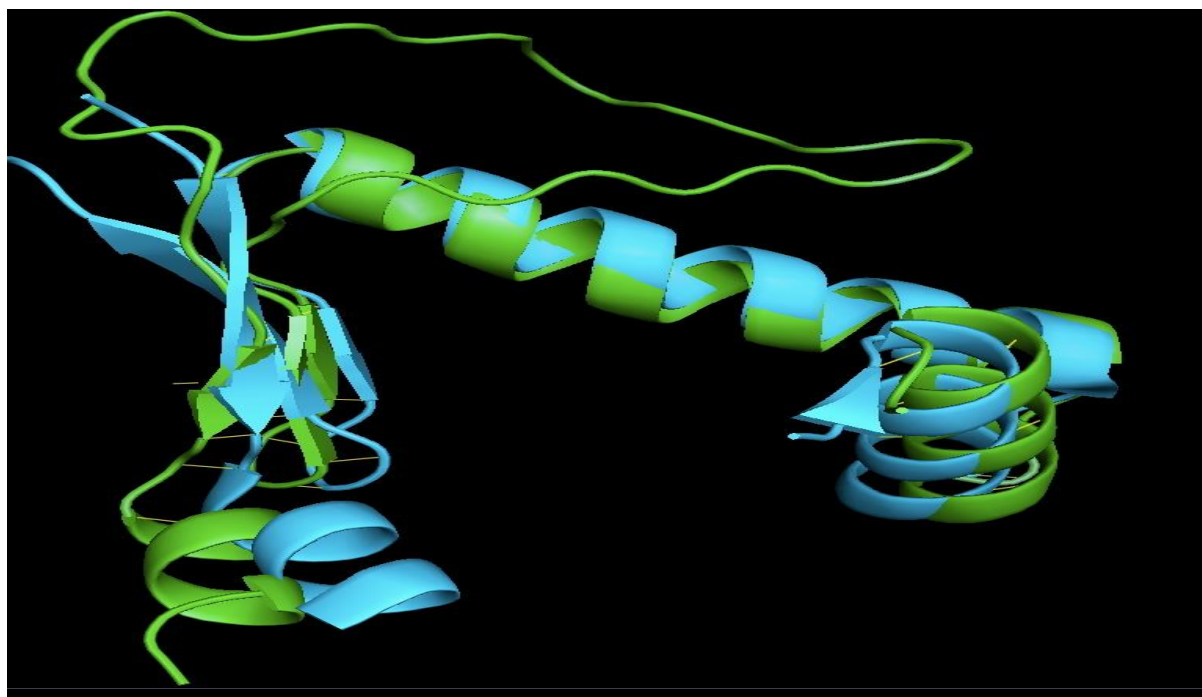
Το επόμενο βήμα είναι η κατασκευή του σκελετού που σχηματίζεται με τον υπολογισμό των ατομικών θέσεων του βιομορίου στόχου με βάση τις θέσεις των αντίστοιχων ατόμων του ομόλογου πειραματικά προσδιορισμένου μορίου. Όταν περισσότερα από μία ομόλογες τρισδιάστατες δομές είναι διαθέσιμες τότε οι θέσεις προσδιορίζονται με ποσοστά συμμετοχής ή βάρη που υπολογίζονται από το ποσοστό ομολογίας στην τοπική ομολογία της ακολουθίας. Το πιο διαδεδομένο πρόγραμμα είναι το Modeller το οποίο είναι γραμμένο σε γλώσσα Python και δύναται να κατασκευάσει μοντέλα ακόμη και με ακολουθίες που έχουν κάτω από 20% ομοιότητα. Αρχικά όπως φαίνεται στην παρακάτω εικόνα, κατά την διαδικασία, στον χρήστη ζητείται να επιλέξει την δομή-οδηγό (template).

Στην συνέχεια, στο πεδίο Options, ζητούνται να προσδιοριστούν:

1. Ο αριθμός των μοντέλων που επιθυμεί να κατασκευάσει
2. Αν θα συμπεριληφθούν ή θα εξαιρεθούν τα ετεροάτομα
3. Να προσδιοριστεί το επίπεδο απεικόνισης
4. Αν θα χρησιμοποιηθεί η επιπρόσθετη ενεργεία ελαχιστοποίησης
5. Με τι τρόπο θα γίνει η χρωματική απεικόνιση του νέου μοντέλου

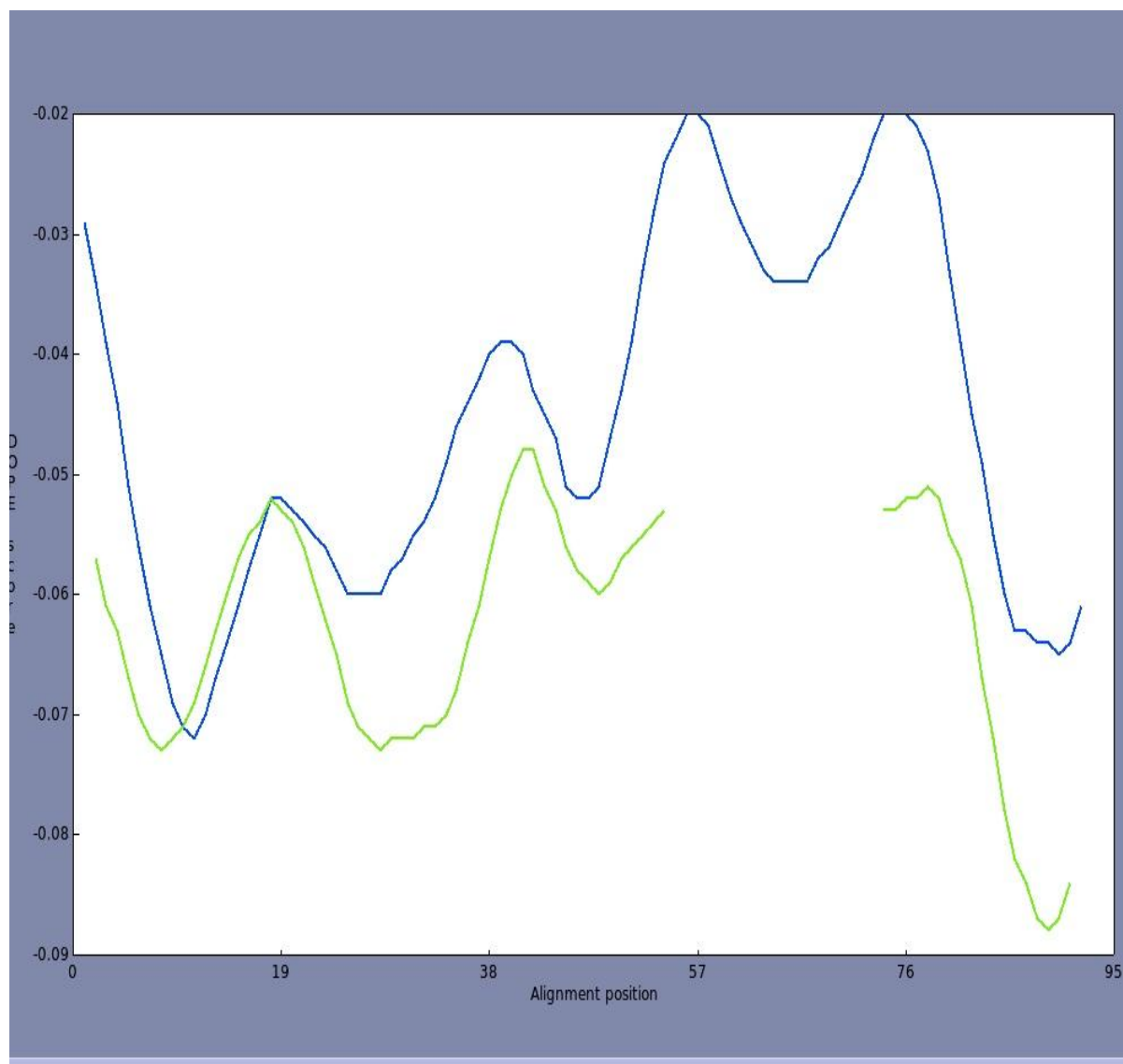
6. Αν θα γίνει η όχι υπέρθεση του νέου μοντέλου, στην δομή - οδηγό

Ρυθμίζοντας, όλες τις ανωτέρω παραμέτρους, και χρησιμοποιώντας τις προεπιλεγμένες ρυθμίσεις, καταλήξαμε στην κατασκευή του νέου μοντέλου με όνομα αρχείου 1_sp_ B8GX11_Chain_A.pdb. Παρακάτω απεικονίζεται ένα μοντέλο, που έχουν συμπεριληφθεί τα ετεροάτομα, δεν έχει χρησιμοποιηθεί επιπρόσθετη ενέργεια ελαχιστοποίησης και η επιλογή χρώματος είναι προεπιλεγμένη από τον αλγόριθμο.



Εικόνα 16 Υπέρθεση του νέου θεωρητικού μοντέλου B8GX11 (πράσινο χρώμα) με την δομή οδηγό 4YFY_A (γαλάζιο χρώμα)

Το Output του Modeller εκτός από το αρχείο .pdb μας δίνει και δύο άλλα αρχεία. Το πρώτο είναι το DOPE (Discrete Optimized Protein Energy) [65] Profile. Στο οποίο απεικονίζεται το διάγραμμα των τιμών DOPE σε σχέση με τα αμινοξικά κατάλοιπα. Με γαλάζιο χρώμα απεικονίζεται το διάγραμμα της δομής-οδηγού, ενώ με πράσινό το διάγραμμα του νέου μοντέλου.



Εικόνα 17 Διαγράμματα DOPE Score του μοντέλου BG8X11 (μπλε χρώμα) και της δομής οδηγού 4YEW_A (πράσινο χρώμα). Παρατηρείται η διακοπή του διαγράμματος στην έλλειψη αμινοξικών καταλοίπων της δομής-οδηγού με βάση το αρχείο .pdb

Βελτιστοποίηση μοντέλου

Η κατασκευή ενός μοντέλου από μόνη της δεν είναι αρκετή, προκειμένου να χαρακτηριστεί το υφιστάμενο μοντέλο ως αξιόλογο. Για να γίνει αυτό θα πρέπει να βελτιστοποιηθεί η δομή του. Όλα τα ομόλογα μοντέλα περιέχουν λάθη όπως πλευρικές αλυσίδες σε λάθος στερεοδομή ή κακώς τοποθετημένες στροφές. Αν τα λάθη είναι τοπικά και δεν συμμετέχουν σε σημαντικές για την λειτουργία περιοχές της πρωτεΐνης τότε δεν δημιουργούν σημαντικό πρόβλημα. Η πιστοποίηση της ποιότητας του μοντέλου είναι από τα σημαντικότερα στάδια της κατασκευής του.

Υπάρχουν δύο κύριοι τρόποι διόρθωσης των λαθών

1. Προσδιορισμός της ενέργειας του συστήματος με την χρήση της μεθόδου ελαχιστοποίησης της ενέργειας ή μοριακή δυναμική.
2. Προσδιορισμός δεικτών κανονικότητας που υποδεικνύουν πόσο κοντά στην πραγματικότητα είναι ορισμένα βασικά χαρακτηριστικά του μοντέλου.

Ο βασικός στόχος είναι η ανάπτυξη κριτηρίων με διακριτική ικανότητα μεταξύ ενός καλού ή κακού μοντέλου. Ο υπολογισμός της εσωτερικής ενέργειας είναι ένα καλό κριτήριο για τον εντοπισμό τάσεων στην στερεοδομή και αποκλίσεων από τις φυσιολογικές τιμές μηκών δεσμών ή γωνιών αλλά δεν μπορεί να διακρίνει πρόβλημα μεταξύ μίας ελικοειδούς δομής από μία β-πτυχωτή επιφάνεια. Ο υπολογισμός της ολικής (ελεύθερης ενέργειας) θα ήταν ένα καλό κριτήριο αλλά δεν υπάρχουν ακριβείς τρόποι υπολογισμού της με ακρίβεια.

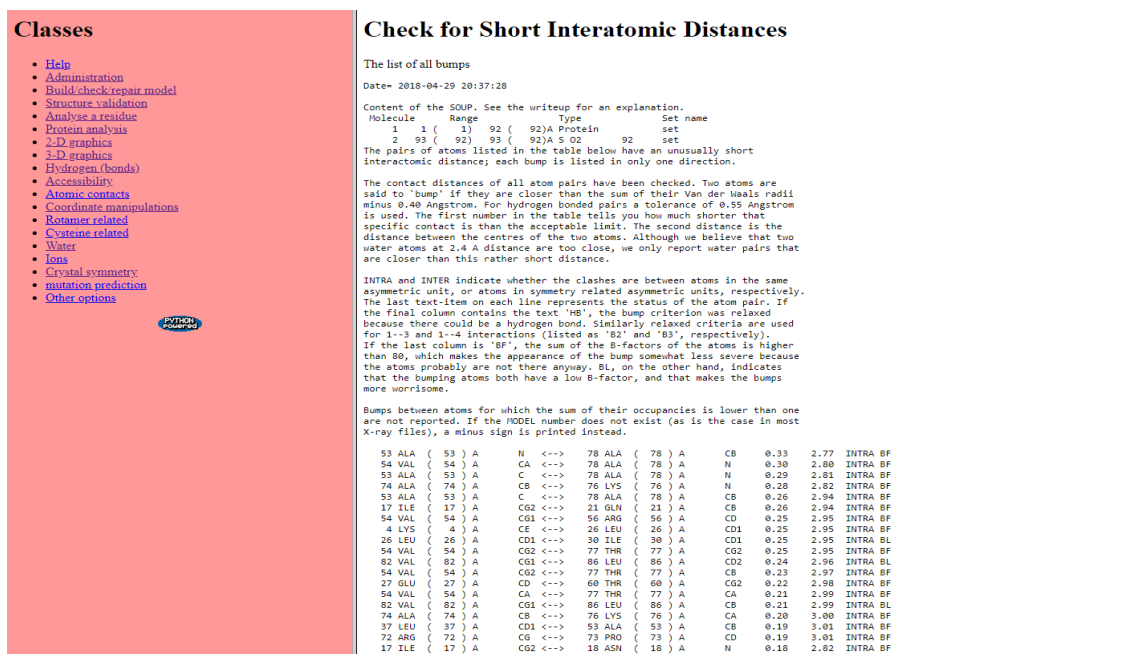
Απαραίτητα κριτήρια είναι οι δείκτες κανονικότητας. Αυτά είναι:

1. Έλεγχοι διέδρων γωνιών, επιπέδων γωνιών και μηκών δεσμών.
2. Κατανομές πολικών και μη πολικών αμινοξέων στο εσωτερικό και εξωτερικό της πρωτεΐνης.
3. Δυναμικά διαλυτότητας.
4. Ενέργεια μετάλλαξης.
5. Κανόνες πακεταρίσματος αμινοξέων και δομικών στοιχείων.
6. Επαφές ατόμων ή ομάδων που δεν παρατηρούνται στις πρωτεϊνικές βάσεις.
7. Μελέτη της επιφάνειας του μοντέλου για εσοχές και του εσωτερικού για κενές κοιλότητες.

Σύμφωνα με τα παραπάνω, για να ελέγξουμε την πιστότητα του μοντέλου που κατασκευάστηκε παραπάνω χρησιμοποιήθηκε το πρόγραμμα Chimera.

Αξιολόγηση του μοντέλου που υπολογίστηκε

Μετά το πέρας της κατασκευής του νέου μοντέλου χρησιμοποιούμε το πρόγραμμα τρισδιάστατης απεικόνισης PyMol μέσω του οποίου ανοίγουμε το αρχείο 1_sp_B8GX11_Chain_A.pdb. και σαρώνουμε οπτικά το μοντέλο για τυχόν ανωμαλίες στις τιμές των αποστάσεων μεταξύ των ατόμων. Σε συνέχεια του μακροσκοπικού οπτικού έλεγχου, εκτελούμε και έναν ακόμη έλεγχο για να επιβεβαιώσουμε και μικροσκοπικά τυχόν ατέλειες του νέου μοντέλου μας. Για να γίνει κάτι τέτοιο θα πρέπει να ελεγχτούν οι ενδό/δια μοριακές αποστάσεις της νέας δομής. Για δομές υψηλής ποιότητας, διαπιστώθηκε ότι κανένα ζεύγος μη συνδεδεμένων με υδρογόνο ατόμων θα πρέπει να έχει απόσταση μεταξύ των ατόμων μεγαλύτερη από 0,4 Å, μικρότερη από το άθροισμα των δύο ακτινών Van der Waals. Για άτομα δεσμευμένα με υδρογόνο αυτό το όριο βρέθηκε να είναι 0,55 Angstrom. Για τον σκοπό αυτό δεν φαντάζει κάτι πιο κατάλληλο, από τον διακομιστή του WhatIF. Το WhatIF είναι γραμμένο σε γλώσσα προγραμματισμού Python και είναι ιδανικό για εκτίμηση ενδοατομικών-διαμοριακών αποστάσεων. Εκεί το εν λόγω μοντέλο ελέγχεται για ενδομοριακές αλληλεπιδράσεις και μετριούνται οι προαναφερθείσες αποστάσεις.



Classes

- [Help](#)
- [Administration](#)
- [Build check/repair model](#)
- [Structure validation](#)
- [Analyse a residue](#)
- [Protein analysis](#)
- [2-D graphics](#)
- [3-D graphics](#)
- [Hydrogen \(bonds\)](#)
- [Accessibility](#)
- [Atomic contacts](#)
- [Coordinate manipulations](#)
- [Rotamer related](#)
- [Cysteine related](#)
- [Water](#)
- [Ions](#)
- [Crystal symmetry](#)
- [Mutation prediction](#)
- [Other options](#)

Check for Short Interatomic Distances

The list of all bumps

Date= 2018-04-29 20:37:28

Content of the SOUP. See the writup for an explanation.

Molecule	Range	Type	Set name
1	1 (1)	92 (92)	Protein set
2	93 (93)	93 (93)	S O2 set

The pairs of atoms listed in the table below have an unusually short interatomic distance; each bump is listed in only one direction.

The contact distances of all atom pairs have been checked. Two atoms are said to "bump" if they are closer than the sum of their Van der Waals radii minus 0.40 Angstrom. For hydrogen bonded pairs a tolerance of 0.55 Angstrom is used. The first number in the table tells you how much shorter that specific contact is than the acceptable limit. The second distance is the distance between the centres of the two atoms. Although we believe that two water atoms at 2.4 A distance are too close, we only report water pairs that are closer than this rather short distance.

INTRA and INTER indicate whether the clashes are between atoms in the same asymmetric unit, or atoms in symmetry related asymmetric units, respectively. The last text-item on each line represents the status of the atom pair. If the final column contains the text 'HB', the bump criterion was relaxed because there could be a hydrogen bond. Similarly relaxed criteria are used for 1-3 and 1-4 interactions (listed as 'B2' and 'B3', respectively). If the last column is 'BF', the sum of the B-factors of the atoms is higher than 80, which makes the appearance of the bump somewhat less severe because the atoms probably are not there anyway. BL, on the other hand, indicates that the bumping atoms both have a low B-factor, and that makes the bumps more worrisome.

Bumps between atoms for which the sum of their occupancies is lower than one are not reported. If the MODEL number does not exist (as is the case in most X-ray files), a minus sign is printed instead.

Atom 1	Atom 2	Distance (Å)	Acceptable (Å)	Status
53 ALA (53) A	N <-> 78 ALA (78) A	0.33	2.77	INTRA BF
54 VAL (54) A	CA <-> 78 ALA (78) A	0.30	2.80	INTRA BF
53 ALA (53) A	C <-> 78 ALA (78) A	0.29	2.81	INTRA BF
74 ALA (74) A	CB <-> 76 LYS (76) A	0.28	2.82	INTRA BF
53 ALA (53) A	C <-> 78 ALA (78) A	0.26	2.94	INTRA BF
17 ILE (17) A	CG2 <-> 21 GLN (21) A	0.26	2.94	INTRA BF
54 VAL (54) A	CG1 <-> 56 ARG (56) A	0.25	2.95	INTRA BF
4 LYS (4) A	CE <-> 26 LEU (26) A	0.25	2.95	INTRA BF
26 LEU (26) A	CD1 <-> 30 ILE (30) A	0.25	2.95	INTRA BL
54 VAL (54) A	CG2 <-> 77 THR (77) A	0.25	2.95	INTRA BF
82 VAL (82) A	CG1 <-> 86 LEU (86) A	0.24	2.96	INTRA BL
54 VAL (54) A	CG2 <-> 77 THR (77) A	0.23	2.97	INTRA BF
27 GLU (27) A	CD <-> 60 THR (60) A	0.22	2.98	INTRA BF
54 VAL (54) A	CA <-> 77 THR (77) A	0.21	2.99	INTRA BF
82 VAL (82) A	CG1 <-> 86 LEU (86) A	0.21	2.99	INTRA BL
74 ALA (74) A	CB <-> 76 LYS (76) A	0.20	3.00	INTRA BF
37 LEU (37) A	CD1 <-> 53 ALA (53) A	0.19	3.01	INTRA BF
72 ARG (72) A	CG <-> 73 PRO (73) A	0.19	3.01	INTRA BF
17 ILE (17) A	CG2 <-> 18 ASN (18) A	0.18	2.82	INTRA BF

Εικόνες 19-20 Στιγμιότυπο αποτελεσμάτων του WhatIF μέσω του οποίου ελέγχονται οι ενδομοριακές αποστάσεις. Στην Εικόνα 20 φαίνεται το Score 7.939.

Classes

- [Help](#)
- [Administration](#)
- [Build/check/repair model](#)
- [Structure validation](#)
- [Analyse a residue](#)
- [Protein analysis](#)
- [2-D graphics](#)
- [3-D graphics](#)
- [Hydrogen \(bonds\)](#)
- [Accessibility](#)
- [Atomic contacts](#)
- [Coordinate manipulations](#)
- [Rotamer related](#)
- [Systeme related](#)
- [Water](#)
- [Ions](#)
- [Crystal symmetry](#)
- [Mutation prediction](#)
- [Other options](#)

PYTHON
Powered

```

53 ALA ( 53 ) A      C <--> 78 ALA ( 78 ) A      CA  0.16  3.04  INTRA BF
8 VAL ( 8 ) A       CG1 <--> 9 THR ( 9 ) A       N   0.16  2.84  INTRA BL
71 ASN ( 71 ) A     NO2 <--> 73 PRO ( 73 ) A     CD  0.16  2.94  INTRA BF
78 ALA ( 78 ) A     C <--> 79 ARG ( 79 ) A     CG  0.16  2.94  INTRA BF
29 PHE ( 29 ) A     CE1 <--> 33 VAL ( 33 ) A     CG2 0.14  3.06  INTRA BF
80 PHE ( 80 ) A     CE2 <--> 82 VAL ( 82 ) A     CG2 0.14  3.06  INTRA BL
4 LYS ( 4 ) A       HE <--> 26 LEU ( 26 ) A     CD1  0.12  2.98  INTRA BF
79 ARG ( 79 ) A     CB <--> 88 PHE ( 88 ) A     N   0.12  2.58  INTRA BF
42 ASP ( 42 ) A     CG <--> 52 LYS ( 52 ) A     CG  0.11  3.09  INTRA BL
52 LYS ( 52 ) A     N <--> 79 ARG ( 79 ) A     O   0.11  2.59  INTRA BF
74 ALA ( 74 ) A     CA <--> 75 SER ( 75 ) A     C   0.10  2.90  INTRA BF
72 ARG ( 72 ) A     N <--> 73 PRO ( 73 ) A     CD  0.09  2.91  INTRA BF
43 VAL ( 43 ) A     CG1 <--> 51 PHE ( 51 ) A     CB  0.09  3.11  INTRA BF
53 ALA ( 53 ) A     O <--> 78 ALA ( 78 ) A     CB  0.08  2.72  INTRA BF
53 ALA ( 53 ) A     O <--> 77 THR ( 77 ) A     C   0.07  2.73  INTRA BF
5 ALA ( 5 ) A       O <--> 8 VAL ( 8 ) A       CG1 0.06  2.74  INTRA BL
43 VAL ( 43 ) A     CG1 <--> 51 PHE ( 51 ) A     CD2 0.06  3.14  INTRA BF
78 ALA ( 78 ) A     C <--> 79 ARG ( 79 ) A     CA  0.06  2.24  INTRA BF
54 VAL ( 54 ) A     CB <--> 78 ALA ( 78 ) A     O   0.06  2.74  INTRA BF
42 ASP ( 42 ) A     OD2 <--> 52 LYS ( 52 ) A     CG  0.06  2.74  INTRA BL
37 LEU ( 37 ) A     CB <--> 53 ALA ( 53 ) A     CB  0.06  3.14  INTRA BF
74 ALA ( 74 ) A     CB <--> 75 SER ( 75 ) A     C   0.06  2.94  INTRA BF
43 VAL ( 43 ) A     O <--> 51 PHE ( 51 ) A     N   0.05  2.65  INTRA BF
53 ALA ( 53 ) A     O <--> 78 ALA ( 78 ) A     N   0.05  2.65  INTRA BF
84 GLU ( 84 ) A     O <--> 88 SER ( 88 ) A     N   0.05  2.65  INTRA BL
37 LEU ( 37 ) A     CD1 <--> 56 ARG ( 56 ) A     NH2 0.05  3.05  INTRA BF
8 VAL ( 8 ) A       O <--> 12 ALA ( 12 ) A     N   0.05  2.65  INTRA BL
54 VAL ( 54 ) A     CB <--> 78 ALA ( 78 ) A     N   0.04  3.06  INTRA BF
43 VAL ( 43 ) A     N <--> 51 PHE ( 51 ) A     O   0.04  2.66  INTRA BF
54 VAL ( 54 ) A     N <--> 78 ALA ( 78 ) A     N   0.03  2.82  INTRA BF
63 ASN ( 63 ) A     CA <--> 64 PRO ( 64 ) A     CD  0.03  2.77  INTRA BF
55 THR ( 55 ) A     OG1 <--> 76 LYS ( 76 ) A     CG  0.02  2.78  INTRA BF
44 ARG ( 44 ) A     CG <--> 59 THR ( 59 ) A     CG2 0.01  3.19  INTRA BF
24 ASP ( 24 ) A     O <--> 28 ALA ( 28 ) A     N   0.01  2.69  INTRA BL
52 LYS ( 52 ) A     C <--> 53 ALA ( 53 ) A     CA  0.01  2.29  INTRA BF
12 ALA ( 12 ) A     O <--> 16 GLY ( 16 ) A     N   0.01  2.69  INTRA BL
39 SER ( 39 ) A     N <--> 40 GLY ( 40 ) A     N   0.01  2.59  INTRA BL
27 GLU ( 27 ) A     OE1 <--> 60 THR ( 60 ) A     CG2 0.01  2.79  INTRA BF

```

The bumps have been binned in 5 categories ranging from 'please look at till must fix'. Additionally, the integrated sum of all bumps, the squared sum of all bumps, and these latter two values normalized by the number of contacts are listed too for comparison purposes between, for example, small and large proteins.

```

Total bump value: 7.939
Total bump value per residue: 0.663
Total number of bumps: 61
Total squared bump value: 1.527
Total number of bumps in the mildest bin: 52
Total number of bumps in the second bin: 9
Total number of bumps in the middle bin: 0
Total number of bumps in the fourth bin: 0
Total number of bumps in the worst bin: 0

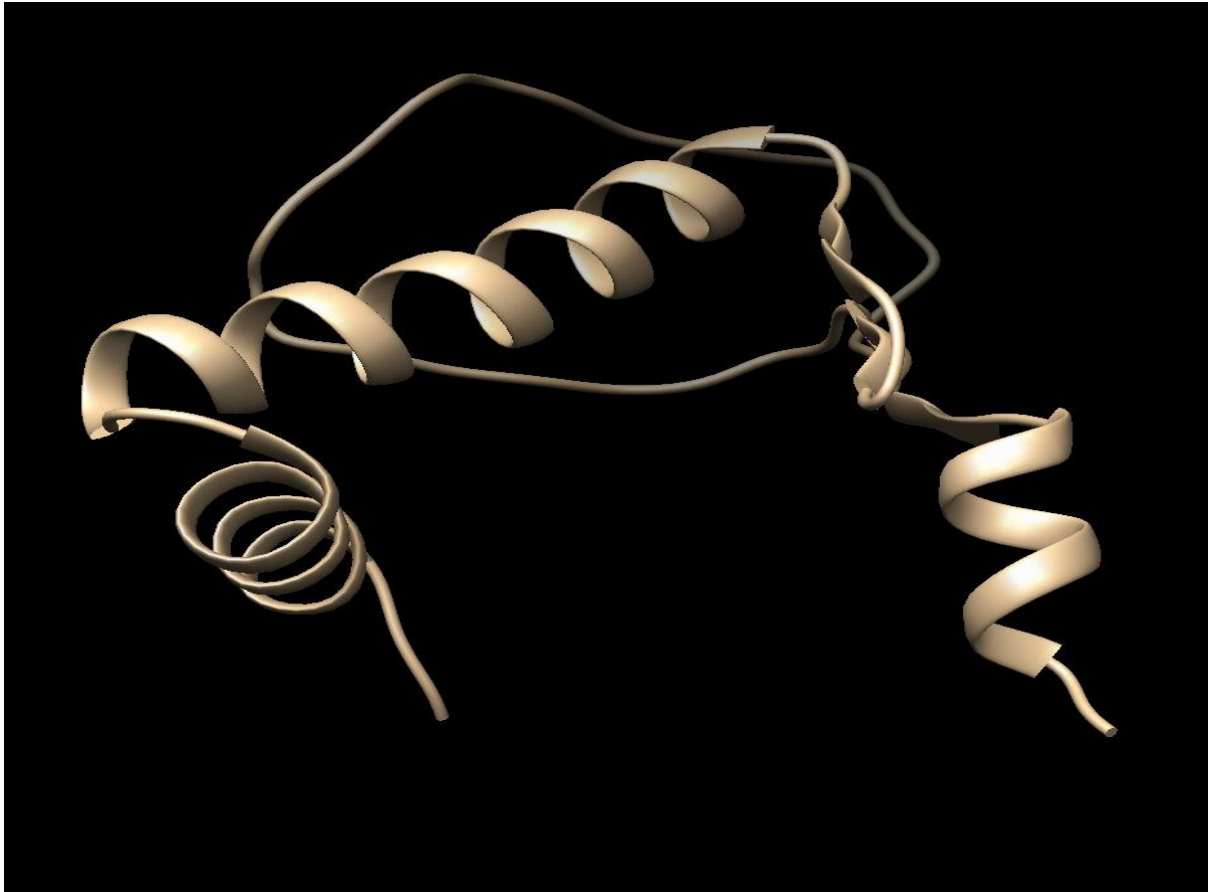
```

If you have detected any error, or have any question or suggestion, please send an Email to Gert Vriend.
 Roland Krause, Maarten L Hekkelman, Jens E Nielsen. [Gert Vriend](#).

Last modified Sun Apr 29 20:37:31 2018

Εικόνες 19-20 Στιγμιότυπο αποτελεσμάτων του WhatIF μέσω του οποίου ελέγχονται οι ενδομοριακές αποστάσεις. Στην Εικόνα 20 φαίνεται το Score 7.939.

Αφού ολοκληρωθεί η αναφορά του WhatIF, στην συνέχεια εισάγουμε το αρχείο με το νέο μοντέλο στο πρόγραμμα Chimera [66] όπου εκτελείται τοπικά και βρίσκεται στην σελίδα (<https://www.cgl.ucsf.edu/chimera>). Ανάμεσα στις πολυάριθμες λειτουργίες και επιλογές του Chimera είναι και η βελτιστοποίηση των νέων δομών μέσω της ενεργειακής ελαχιστοποίησης. Έτσι αφού εισάγουμε το αρχείο 1_sp_B8GX11_Chain_A.pdb στο Chimera επιλέγουμε το μονοπάτι Tools->Structure Editing ->Minimize Structure, βελτιστοποιώντας έτσι την νέα δομή. Εδώ αξίζει να αναφερθεί ότι υπήρξαν και μοντέλα όπου ετεροάτομα τους (Ni,Cl) στην δομή προκάλεσαν πρόβλημα κατά την βελτιστοποίηση. Το συγκεκριμένοι πρόβλημα αντιμετωπίστηκε με το να μορφοποιηθεί κατάλληλα το αρχείο .pdb μέσω notepad, προκειμένου να αφαιρεθούν τα ετεροάτομα και να εκτελεστεί εκ νέου η βελτιστοποίηση η οποία και ήταν επιτυχής.



Εικόνα 21 Στιγμιότυπο από το πρόγραμμα Chimera στο οποίο απεικονίζεται το νεοσύστατο μοντέλο στο οποίο θα επέλθει δομική βελτιστοποίηση του τελευταίου μέσω ενεργειακής ελαχιστοποίησης δυναμικού.

Ακολούθως επαναλαμβάνουμε την διαδικασία μικροσκοπικού έλεγχου μέσω του διακομιστή του WhatIF, και παραθέτουμε το αρχείο αποτέλεσμα των μετά την εισαγωγή του μοντέλου στο Chimera, προκείμενου να παρατηρήσουμε τις βελτιώσεις στις αποστάσεις που πραγματοποιήθηκαν.

Classes

- [Help](#)
- [Administration](#)
- [Build/check/repair model](#)
- [Structure validation](#)
- [Analyse a residue](#)
- [Protein analysis](#)
- [2-D graphics](#)
- [3-D graphics](#)
- [Hydrogen \(bonds\)](#)
- [Accessibility](#)
- [Atomic contacts](#)
- [Coordinate manipulations](#)
- [Rotamer related](#)
- [Cysteine related](#)
- [Water](#)
- [Ions](#)
- [Crystal symmetry](#)
- [mutation prediction](#)
- [Other options](#)



Check for Short Interatomic Distances

The list of all bumps

Date= 2018-04-29 20:51:36

The pairs of atoms listed in the table below have an unusually short interatomic distance; each bump is listed in only one direction.

The contact distances of all atom pairs have been checked. Two atoms are said to 'bump' if they are closer than the sum of their Van der Waals radii minus 0.40 Angstrom. For hydrogen bonded pairs a tolerance of 0.55 Angstrom is used. The first number in the table tells you how much shorter that specific contact is than the acceptable limit. The second distance is the distance between the centres of the two atoms. Although we believe that two water atoms at 2.4 A distance are too close, we only report water pairs that are closer than this rather short distance.

INTRA and INTER indicate whether the clashes are between atoms in the same asymmetric unit, or atoms in symmetry related asymmetric units, respectively. The last text-item on each line represents the status of the atom pair. If the final column contains the text 'HB', the bump criterion was relaxed because there could be a hydrogen bond. Similarly relaxed criteria are used for 1--3 and 1--4 interactions (listed as 'B2' and 'B3', respectively). If the last column is 'BF', the sum of the B-factors of the atoms is higher than 80, which makes the appearance of the bump somewhat less severe because the atoms probably are not there anyway. BL, on the other hand, indicates that the bumping atoms both have a low B-factor, and that makes the bumps more worrisome.

Bumps between atoms for which the sum of their occupancies is lower than one are not reported. If the MODEL number does not exist (as is the case in most X-ray files), a minus sign is printed instead.

```
4 LYS ( 4 ) A    NZ <-> 27 GLU ( 27 ) A    OE1  0.13  2.57 INTRA BF
4 LYS ( 4 ) A    NZ <-> 27 GLU ( 27 ) A    CD   0.08  3.02 INTRA BF
65 ARG ( 65 ) A  NH1 <-> 68 GLU ( 68 ) A    OE1  0.07  2.63 INTRA BF
72 ARG ( 72 ) A  N   <-> 73 PRO ( 73 ) A    CD   0.03  2.97 INTRA BF
35 ASP ( 35 ) A  OD2 <-> 38 LYS ( 38 ) A    NZ   0.01  2.69 INTRA BL
```

The bumps have been binned in 5 categories ranging from 'please look at' till 'must fix'. Additionally, the integrated sum of all bumps, the squared sum of all bumps, and these latter two values normalized by the number of contacts are listed too for comparison purposes between, for example, small and large proteins.

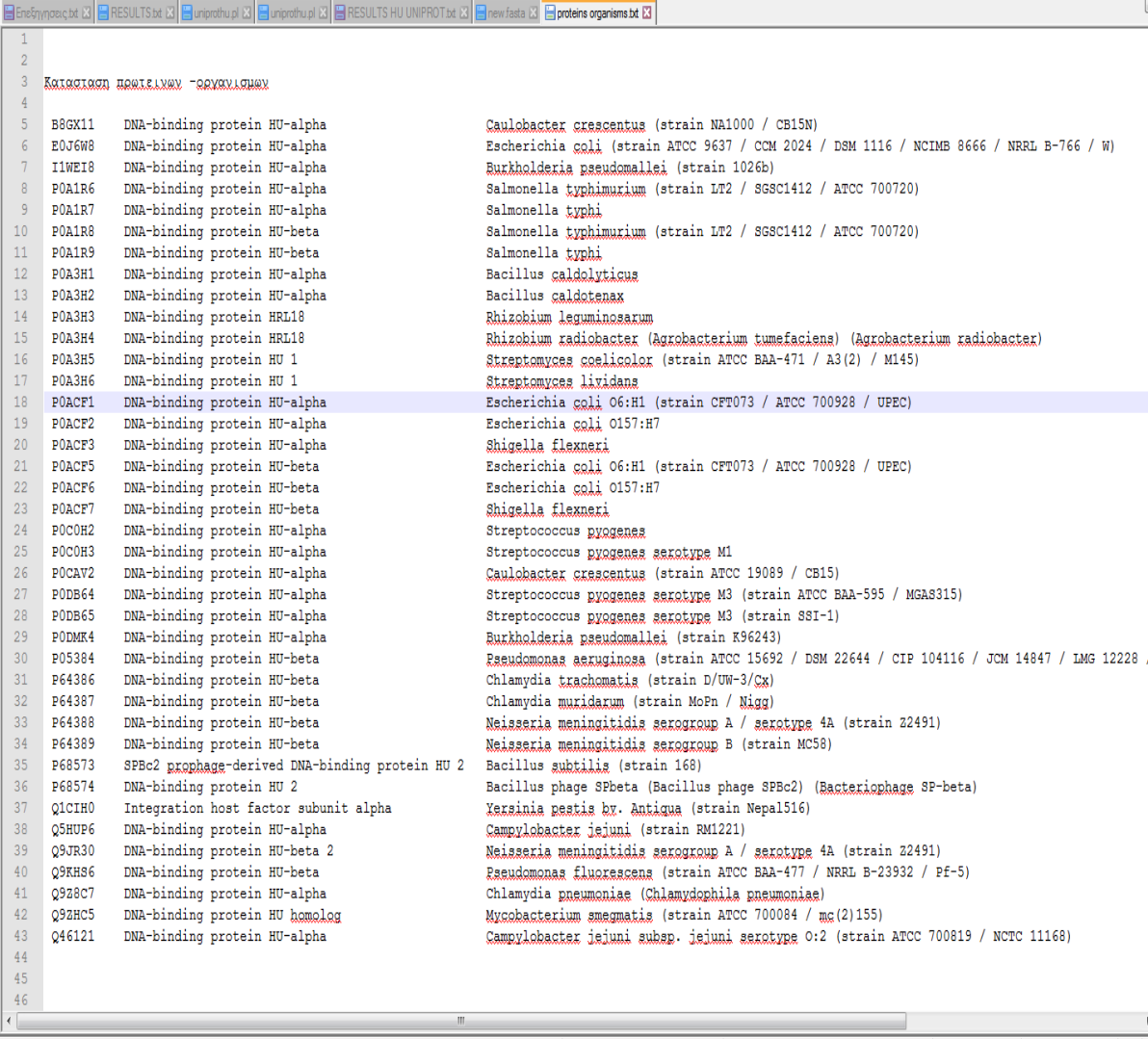
```
Total bump value: 0.329
Total bump value per residue: 0.054
Total number of bumps: 5
Total squared bump value: 0.030
Total number of bumps in the mildest bin: 5
Total number of bumps in the second bin: 0
Total number of bumps in the middle bin: 0
Total number of bumps in the fourth bin: 0
Total number of bumps in the worst bin: 0
```

If you have detected any error, or have any question or suggestion, please send an Email to Gert Vriend.
Roland Krause, Maarten L Hekkelman, Jens E Nielsen, [Gert Vriend](#).

Εικόνα 22 Στιγμιότυπο αποτελεσμάτων του WhatIF μέσω του οποίου ελέγχονται οι σύντομες ενδομοριακές αποστάσεις μετά την βελτιστοποίηση που έχει επέλθει μέσω του Chimera. Στο κάτω μέρος φαίνεται το Score 0.329.

3.3 Κατασκευή ομοδιμέρων ή ετεροδιμερών μοντέλων μέσω Modeler

Όπως έχει αναφερθεί και στην εισαγωγή, η πρωτεΐνη HU μπορεί να υφίσταται σε μορφή ομοδιμερούς ή ετεροδιμερούς. Κατά την αναζήτηση μας, δημιουργήσαμε μια λίστα. Εμφανίζεται η αλυσίδα της κάθε καταχώρησης σε σχέση με τον οργανισμό που ανήκει ώστε να αποφανθούμε για τον ομοδιμερισμό (αν προκύψει ένα στέλεχος αλυσίδας A ή B μεμονωμένα ανά οργανισμό) ή τον ετεροδιμερισμό τους (αν προκύψουν στελέχη A και B στον ίδιο οργανισμό).



```
1
2
3 Κατασκευή πρωτεϊνών -οργανισμών
4
5 B8GX11 DNA-binding protein HU-alpha Caulobacter crescentus (strain NA1000 / CB15N)
6 E0J6W8 DNA-binding protein HU-alpha Escherichia coli (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B-766 / W)
7 I1WEI8 DNA-binding protein HU-alpha Burkholderia pseudomallei (strain 1026b)
8 P0A1R6 DNA-binding protein HU-alpha Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
9 P0A1R7 DNA-binding protein HU-alpha Salmonella typhi
10 P0A1R8 DNA-binding protein HU-beta Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)
11 P0A1R9 DNA-binding protein HU-beta Salmonella typhi
12 P0A3H1 DNA-binding protein HU-alpha Bacillus caldolyticus
13 P0A3H2 DNA-binding protein HU-alpha Bacillus caldotenax
14 P0A3H3 DNA-binding protein HRL18 Rhizobium leguminosarum
15 P0A3H4 DNA-binding protein HRL18 Rhizobium radiobacter (Agrobacterium tumefaciens) (Agrobacterium radiobacter)
16 P0A3H5 DNA-binding protein HU 1 Streptomyces coelicolor (strain ATCC BAA-471 / A3(2) / M145)
17 P0A3H6 DNA-binding protein HU 1 Streptomyces lividans
18 P0ACF1 DNA-binding protein HU-alpha Escherichia coli O6:H1 (strain CPT073 / ATCC 700928 / UPEC)
19 P0ACF2 DNA-binding protein HU-alpha Escherichia coli O157:H7
20 P0ACF3 DNA-binding protein HU-alpha Shigella flexneri
21 P0ACF5 DNA-binding protein HU-beta Escherichia coli O6:H1 (strain CPT073 / ATCC 700928 / UPEC)
22 P0ACF6 DNA-binding protein HU-beta Escherichia coli O157:H7
23 P0ACF7 DNA-binding protein HU-beta Shigella flexneri
24 P0C0H2 DNA-binding protein HU-alpha Streptococcus pyogenes
25 P0C0H3 DNA-binding protein HU-alpha Streptococcus pyogenes serotype M1
26 P0CAV2 DNA-binding protein HU-alpha Caulobacter crescentus (strain ATCC 19089 / CB15)
27 P0DB64 DNA-binding protein HU-alpha Streptococcus pyogenes serotype M3 (strain ATCC BAA-595 / MGA8315)
28 P0DB65 DNA-binding protein HU-alpha Streptococcus pyogenes serotype M3 (strain SSI-1)
29 P0DMR4 DNA-binding protein HU-alpha Burkholderia pseudomallei (strain K96243)
30 P05384 DNA-binding protein HU-beta Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 /
31 P64386 DNA-binding protein HU-beta Chlamydia trachomatis (strain D/UW-3/Cx)
32 P64387 DNA-binding protein HU-beta Chlamydia muridarum (strain MoPn / Nigg)
33 P64388 DNA-binding protein HU-beta Neisseria meningitidis serogroup A / serotype 4A (strain Z2491)
34 P64389 DNA-binding protein HU-beta Neisseria meningitidis serogroup B (strain MC58)
35 P68573 SPBc2 phage-derived DNA-binding protein HU 2 Bacillus subtilis (strain 168)
36 P68574 DNA-binding protein HU 2 Bacillus phage SPbeta (Bacillus phage SPBc2) (Bacteriophage SP-beta)
37 Q1CIH0 Integration host factor subunit alpha Yersinia pestis bv. Antiqua (strain Nepal516)
38 Q5HVP6 DNA-binding protein HU-alpha Campylobacter jejuni (strain RM1221)
39 Q9JR30 DNA-binding protein HU-beta 2 Neisseria meningitidis serogroup A / serotype 4A (strain Z2491)
40 Q9KH86 DNA-binding protein HU-beta Pseudomonas fluorescens (strain ATCC BAA-477 / NRRL B-23932 / Pf-5)
41 Q9Z8C7 DNA-binding protein HU-alpha Chlamydia pneumoniae (Chlamydomphila pneumoniae)
42 Q9ZHC5 DNA-binding protein HU homolog Mycobacterium smegmatis (strain ATCC 700084 / mc(2)155)
43 Q46121 DNA-binding protein HU-alpha Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11168)
44
45
46
```

Εικόνα 25 Αποτύπωση αρχείου στο οποίο απεικονίζονται οι πρωτεϊνικές ακολουθίες με τους αντίστοιχους οργανισμούς στους οποίους εμπεριέχονται.

Αξιοσημείωτο είναι εδώ το γεγονός, ότι σε έναν μεγάλο αριθμό templates, υπήρξε αλλαγή σε σχέση με αυτά που χρησιμοποιήθηκαν στα μονομερή. Και αυτό έγινε διότι ενώ εμείς θέλαμε να ομοδιμερίσουμε ένα στέλεχος HU το template της αρχικής *insilico* κατασκευής του μονομερούς δεν ήταν ομοδιμερές. Η λύση που δώσαμε ήταν να εκτελέσουμε εκ νέου BLAST ώστε να προκύψει το καταλληλότερο ομοδιμερές ή ετεροδιμερές template με το βέλτιστο δυνατό score που να πληρεί τις προϋποθέσεις.

Description B8GX11					Max score	Total score	Query cover	E value	
Select seq	pdb 4YEW A	Chain A, Huab-19bp			80.1	80.1	95%	4e-21	
Select seq	pdb 2097 B	Chain B, Crystal Structure Of E. Coli Hu Heterodimer			79.3	79.3	95%	1e-20	
Select seq	pdb 1MUL A	Chain A, Crystal Structure Of The E. Coli Hu Alpha2 Protein			79.0	79.0	94%	1e-20	
Select seq	pdb 4YFY A	Chain A, Huaa-20bp			79.0	79.0	94%	1e-20	
Select seq	pdb 1HUE A	Chain A, Histone-Like Protein			78.6	78.6	94%	2e-20	
Description E0J6W8					Max score	Total score	Query cover	E value	Iden
Select seq	pdb 4YFY A	Chain A, Huaa-20bp			179	179	100%	3e-60	100%
Select seq	pdb 1MUL A	Chain A, Crystal Structure Of The E. Coli Hu Alpha2 Protein			179	179	100%	3e-60	100%
Select seq	pdb 4YFH A	Chain A, Hu38-20bp			176	176	100%	5e-59	98%
Select seq	pdb 4YF0 A	Chain A, Hu38-19bp			175	175	100%	7e-59	98%
Select seq	pdb 2097 B	Chain B, Crystal Structure Of E. Coli Hu Heterodimer			123	123	98%	3e-38	70%
Description I1WEI8					Max score	Total score	Query cover	E v	
Select seq	pdb 4QUN A	Chain A, Crystal Structure Of Apo Nucleoid Associated Protein, Sav1473			97.4	97.4	96%	7e	
Select seq	pdb 1HUE A	Chain A, Histone-Like Protein			97.1	97.1	96%	8e	
Select seq	pdb 2097 B	Chain B, Crystal Structure Of E. Coli Hu Heterodimer			95.9	95.9	97%	2e	
Select seq	pdb 4YEW A	Chain A, Huab-19bp			93.6	93.6	97%	2e	
Select seq	pdb 1P51 A	Chain A, Anabaena Hu-Dna Cocystal Structure (Ahu6)			92.8	92.8	96%	4e	
Description O68451					Max score	Total			
Select seq	pdb 1IHF B	Chain B, Integration Host FactorDNA COMPLEX			48.9	4			

Εικόνα 26 Αποτύπωση αποτελεσμάτων του νέου BLAST που εκτελέστηκε. Παρατηρούμε ότι επιλέγεται η πρώτη καταλληλότερη ΟΜΟΔΙΜΕΡΗ δομή-οδηγός (Στην περίπτωση της BG8X11 η δομή 1HUE).

Για τον ομο/ετερο-διμερισμό τροποποιήθηκαν και χρησιμοποιήθηκαν υπορουτίνες του αλγόριθμου του γνωστού Modeller, οι οποίες είναι γραμμένες σε γλώσσα Python. Η ονομασία των αρχείων ομο ή έτερο διμερών που πρόεκυψαν ονομάστηκαν homomodel ή heteromodel αντίστοιχά (Στην περίπτωση που παραδείγματός μας b8gx11-1huehomomodel). Το template έχει αλλάξει από την δομή 4YEW_A που είχε αρχικά, με την δομή 1HUE που πρόεκυψε λόγω του νέου BLAST.

```

1  # Comparative modeling by the automodel class
2  #
3  # Demonstrates how to build multi-chain models, and symmetry restraints
4  #
5  from modeller import *
6  from modeller.automodel import * # Load the automodel class
7
8  log.verbose()
9
10 # Override the 'special_restraints' and 'user_after_single_model' methods:
11 class MyModel(automodel):
12     def special_restraints(self, aln):
13         # Constrain the A and B chains to be identical (but only restrain
14         # the C-alpha atoms, to reduce the number of interatomic distances
15         # that need to be calculated):
16         s1 = selection(self.chains['A']).only_atom_types('CA')
17         s2 = selection(self.chains['B']).only_atom_types('CA')
18         self.restraints.symmetry.append(symmetry(s1, s2, 1.0))
19     def user_after_single_model(self):
20         # Report on symmetry violations greater than 1A after building
21         # each model:
22         self.restraints.symmetry.report(1.0)
23
24 env = environ()
25 # directories for input atom files
26 env.io.atom_files_directory = ['.', 'C:\Users\dimitris\Desktop\omodimer\ldim1.pdb']
27
28 # Be sure to use 'MyModel' rather than 'automodel' here!
29 a = MyModel(env,
30             alnfile = 'twochain.ali', # alignment filename
31             knowns = 'ldim', #codes of the templates
32             sequence = 'dimseq' ) # code of the target
33
34 a.starting_model= 1 # index of the first model
35 a.ending_model = 1 # index of the last model
36 # (determines how many models to calculate)
37 a.make() # do comparative modeling

```

Εικόνα 27 Υπορουτίνα (script) του Modeller για την κατασκευή ομοδιμερών και ετεροδιμερών μοντέλων.

Αξιολόγηση των *in silico* κατασκευασμένων ομο/έτερο διμερών μοντέλων

Σε αντιστοιχία με τα μονομερή, και τα υφιστάμενα μοντέλα θα πρέπει να αξιολογηθούν. Οι παραμέτρους για την αξιολόγηση τους είναι οι ενδοατομικές και οι διαμοριακές αποστάσεις, ο αριθμός των *rockets* και των *cavities*, η "προσβάσιμη" "θαμμένη" επιφάνεια του μοντέλου σε σχέση με τον διαλύτη, και τέλος η τιμή του ΔG . Για τα δύο πρώτα επαναλήφθηκε η διαδικασία μέσω WhatIF και δομικής βελτιστοποίησης μέσω Chimera όπως αυτή περιγράφηκε παραπάνω. Τα βελτιστοποιημένα πλέον μοντέλα ονομάζονται όπως το αρχείο εισόδου και στο τέλος χρησιμοποιείται ο όρος *tuned* προκειμένου να υποδηλωθεί η βελτιστοποίηση δομής μέσω του Chimera (Στο παράδειγμά μας `b8gx11-1huehomomodeltuned.pdb`). Για τον υπολογισμό του αριθμού των *rockets* και των *cavities* χρησιμοποιήθηκε το πρόγραμμα Pocasa (<http://altair.sci.hokudai.ac.jp/g6/service/pocasa>), ενώ για το υπολογισμό της "θαμμένης" και "προσβάσιμης" επιφάνειας του μοντέλου από τον διαλύτη και της τιμής ΔG χρησιμοποιήθηκε το πρόγραμμα Pisa (<http://www.ebi.ac.uk/pdbe/pisa/>). Σημειώνεται ότι και τα δύο προγράμματα εκτελούνται διαδικτυακά μέσω διακομιστή (*server*).

Το Pocasa χρησιμοποιεί ως αρχείο εισόδου το αρχείο του παραδείγματος μας, `b8gx11-1huehomomodeltuned.pdb`. Εκεί καθορίζεται πόσες αλυσίδες έχει το μοντέλο και εισάγονται ανάλογα στο πεδίο "ChainID" (Στο παράδειγμά μας "A,B" μιας και πρόκειται για διμερές). Στα υπόλοιπα πεδία (Prode radius, SPF, PDF, Top N, grid size και atom type) αποδεχόμαστε τις Default τιμές τους.

POCASA 1.1

It is recommended to read the [manual](#) before trying the web POCASA.

The screenshot shows the POCASA 1.1 web interface. It features several input fields and radio buttons for configuring the search parameters. The 'Input as' section has two options: 'file' (selected) and 'PDB ID'. The 'file' option is linked to a 'Choose File' button and a text input field containing 'b8gx11-1hue ... I tuned.pdb'. Below this is a note: 'pdb file which contains protein coordinates'. The 'Chain ID' field contains 'A,B' with a note: '"NULL" (means the first chain) or like: "A", "a,b", "1 2" or "ABC"'. The 'Probe radius' field contains '2' with a note: 'The radius (Å) of probe sphere (positive integer)'. The 'SPF' field contains '16' with a note: 'Single Point Flag (nonnegative integer less than 28)'. The 'PDF' field contains '18' with a note: 'Protein Depth Flag (nonnegative integer less than 28)'. The 'Top N' field contains '5' with a note: 'The number of results to show. specify 0 to show all candidates.'. The 'grid size' section has two radio buttons: '1 Å' (selected) and '0.5 Å'. The 'atom type' section has two radio buttons: 'protein' (selected) and 'all'. At the bottom of the form are two buttons: 'Get Pockets and Cavities' and 'Reset'.

[\[Back\]](#) [\[Laboratory of X-ray structural biology\]](#)

Εικόνα 28 Στιγμιότυπο εισαγωγής προγράμματος POCASA. Διακρίνεται το πεδίο εισαγωγής (input as) του αρχείου .pdb καθώς και η επιλογή των αλυσίδων (ChainID) του ομο/ετεροδιμερούς.

Στο αρχείο εξόδου του POCASA έχουμε τον αριθμό των εσοχών (cavities) που βρέθηκαν καθώς και τον υπολογισμό των όγκων τους που είναι προσβάσιμες στον διαλύτη.

Grid size: 1.0 Å

The screenshot displays the POCASA software interface with four main panels:

- Output files:** A list of generated files including `b8gx11-1hue_homomodel_tuned.pdb`, `b8gx11-1hue_homomodel_tuned_Parameters.txt`, `b8gx11-1hue_homomodel_tuned_Pocket_DepthCenters.pdb`, `b8gx11-1hue_homomodel_tuned_TopN_pockets.pdb`, and `b8gx11-1hue_homomodel_tuned_simple.pdb`.
- Program's output:** A text window showing the execution log, including command line arguments, version information, and a detailed list of 5 3D pockets found, such as "Pocket 84's volume is 188, VD value is 476, the average VD is 2.53369".
- Rank order:** A list of the top 5 pockets ranked by volume and VD value: Rank 1 is Pocket 84, Rank 2 is Pocket 70, Rank 3 is Pocket 37, Rank 4 is Pocket 243, and Rank 5 is Pocket 161.
- Check the result immediately with Jmol:** A control panel for the Jmol viewer, featuring radio buttons for "Wireframe", "Cartoon", and "van der Waals Surface", a dropdown menu for "Pockets and cavities" (set to "No. 84"), and "Select All" and "Deselect All" buttons.

Εικόνα 29 Στιγμιότυπο αποτελεσμάτων προγράμματος POCASA. Στο αριστερό μέρος διακρίνονται τα αποτελέσματα της εισαχθείσας πρωτεΐνης υπολογίζοντας τον όγκο των κοιλοτήτων που έχουν υπολογιστεί. Στο δεξί μέρος ταξινομούνται χωρικά τα αποτελέσματα, ενώ στο κάτω μέρος δίνεται η δυνατότητα στον χρήστη να γίνει απεικόνιση μέσω Jmol.

Το Pisa χρησιμοποιεί ως αρχείο εισόδου, το αρχείο .pdb του νέο παραγόμενου μοντέλου, στο παράδειγμα μας το αρχείο b8gx11-1huehomodeltuned.pdb. Στην συνέχεια επιλέγεται η εντολή interfaces ώστε να υπολογιστούν οι αλληλεπιδράσεις της πρωτεΐνης με τον διαλύτη.

Protein Data Bank
in Europe
Bringing Structure to Biology

PDBePISA

PISA Query.

Submission Form Structure Analysis Database Searches

Coordinate file Choose File No file chosen Upload

PDB entry

Analysis: 2 amino acid chains in ASU

Cell parameters: A: not given Alpha: not given
B: not given Beta: not given
C: not given Gamma: not given

Space symmetry group: not given

Crystallographic information not found. You may give the cell parameters and the space symmetry group in the fields above. You may also submit without crystal data, in which case no symmetry mates will be explored.


Interfaces Monomers Assemblies

DBe PISA v1.52 [20/10/2014]

Εικόνα 30 Στιγμιότυπο εισαγωγής προγράμματος PISA. Διακρίνεται το πεδίο εισαγωγής του αρχείου .pdb ή εναλλακτικά της αναζήτησης της δομής μέσω του κωδικού PDB.

Το αρχείο εξόδου του PISA περιλαμβάνει την επιφάνεια του είναι προσβάσιμη στον διαλύτη (7112 \AA^2) την ελεύθερη ενέργεια ΔG (-35,6 kcal/mol) καθώς και την τιμή p -value η οποία όταν έχει τιμή μεγαλύτερη από 0.5 δηλώνει μικρή υδροφοβικότητα ενώ όταν είναι μικρότερη από 0.5 δηλώνει ότι η δομή είναι υδρόφοβη (0.017, στο παράδειγμά μας, αρά μεγάλη υδροφοβικότητα)

PISA Interface List.


Session Map  (id=526-J9-K6B)

Start **Interfaces** Interface Search

Monomers

Assemblies

Interfaces in b8gx11-1hue homomodel tuned.pdb

Interfaces  XML View Details Download Search

#	Structure 1			x	Structure 2			interface	ΔG	ΔG	N_{HB}	N_{SB}	N_{DS}	CSS
	NN	Range	Surface \AA^2		Range	Surface \AA^2	area, \AA^2							
1	A	195 48	7112	◇	B	194 49	7127	1922.6	-35.6	0.017	2	0	0	0.000

View Details Download Search

PDBe PISA v1.52 [20/10/2014]

Εικόνα 31 Στιγμιότυπο αποτελεσμάτων προγράμματος PISA. Διακρίνονται από αριστερά προς τα δεξιά η αλυσίδα, ο αριθμός των ατόμων, ο αριθμός των αμινοξικών καταλοίπων, η προσβάσιμη στον διαλύτη επιφάνεια. Επίσης εμφανίζεται και η τιμή του συνολικού ΔG , καθώς και το P-Value του ΔG .

3.4 Εντοπισμός του σημείου της αλληλουχίας διμερισμού των HU

Όπως αναφέρθηκε στην εισαγωγή, η ύπαρξη της αλληλουχίας του σημείου διμερισμού ή αλλιώς Dimerization Signals (DS) των ομο/ετερο διμερών είναι αδιαμφισβήτητη. Πως μπορούμε όμως να εντοπίσουμε, και ακόμη πως μπορούμε να υπερθέσουμε τα DS του κάθε μοντέλου σε σχέση με το DS της αντίστοιχης δομής-οδηγού του;

Αυτή την απάντηση θα αποπειραθούμε να δώσουμε παρακάτω. Πρωτίστως γνωρίζουμε ότι το DS δεν είναι άλλο παρά μια σειρά 5 συγκεκριμένων αμινοξικών κατάλοιπων που απαντώνται κάθε φορά συνήθως στο κέντρο κάθε αλυσίδας. Δείχνει, επίσης να είναι υπεύθυνο για τον ομοδιμερισμό ή ετεροδιμερισμό των πρωτεϊνών της οικογένειας HU. Η μικρή αυτή λοιπόν η ακολουθία αποτελείται από τα Gly - Phe - Gly - Thr - Phe και υπάρχουν βάσιμες υποψίες ότι η στερεοδιάταξη των πλευρικών αλυσίδων της Phe είναι υπεύθυνη για την αλλαγή της τιμής του ΔG . Συνεπώς αφού το εντοπίσουμε, τροποποιούμε μέσω του Notepad το αντίστοιχο αρχείο .pdb του μοντέλου, απογυμνώνοντας το από τα “περιττά κατάλοιπα”. Το νέο αρχείο .pdb που προκύπτει, έχει το όνομα ds 5res (από τα 5 κατάλοιπα) και ακολουθεί το αρχικό όνομα του μοντέλου (π.χ ds 5resb8gx11.pdb).

SHEET	8	8 1	LYS B 168	PHE B 172	0					
ATOM	701	N	GLY A 47	3.269	-7.526	8.793	1.00	68.43		
ATOM	702	CA	GLY A 47	1.863	-7.714	8.415	1.00	68.43		
ATOM	703	C	GLY A 47	1.447	-7.024	7.108	1.00	68.43		
ATOM	704	O	GLY A 47	0.796	-7.653	6.274	1.00	68.43		
ATOM	705	H	GLY A 47	3.904	-8.280	8.578	1.00	0.00		
ATOM	706	HA2	GLY A 47	1.662	-8.780	8.312	1.00	0.00		
ATOM	707	HA3	GLY A 47	1.230	-7.322	9.213	1.00	0.00		
ATOM	708	N	PHE A 48	1.958	-5.817	6.836	1.00	138.99		
ATOM	709	CA	PHE A 48	1.491	-5.026	5.694	1.00	138.99		
ATOM	710	CB	PHE A 48	1.284	-3.576	6.120	1.00	138.99		
ATOM	711	CG	PHE A 48	0.558	-2.794	5.046	1.00	138.99		
ATOM	712	CD1	PHE A 48	-0.799	-3.073	4.805	1.00	138.99		
ATOM	713	CD2	PHE A 48	1.227	-1.842	4.255	1.00	138.99		
ATOM	714	CE1	PHE A 48	-1.481	-2.401	3.784	1.00	138.99		
ATOM	715	CE2	PHE A 48	0.539	-1.164	3.232	1.00	138.99		
ATOM	716	CZ	PHE A 48	-0.804	-1.474	2.973	1.00	138.99		
ATOM	717	C	PHE A 48	2.376	-5.072	4.449	1.00	138.99		
ATOM	718	O	PHE A 48	1.857	-5.060	3.329	1.00	138.99		
ATOM	719	H	PHE A 48	2.487	-5.357	7.569	1.00	0.00		
ATOM	720	HA	PHE A 48	0.513	-5.404	5.387	1.00	0.00		
ATOM	721	HB2	PHE A 48	0.684	-3.579	7.028	1.00	0.00		
ATOM	722	HB3	PHE A 48	2.244	-3.105	6.343	1.00	0.00		
ATOM	723	HD1	PHE A 48	-1.330	-3.783	5.427	1.00	0.00		
ATOM	724	HD2	PHE A 48	2.267	-1.628	4.436	1.00	0.00		
ATOM	725	HE1	PHE A 48	-2.531	-2.607	3.631	1.00	0.00		
ATOM	726	HE2	PHE A 48	1.045	-0.425	2.631	1.00	0.00		
ATOM	727	HZ	PHE A 48	-1.325	-0.977	2.171	1.00	0.00		
ATOM	728	N	GLY A 49	3.700	-4.943	4.595	1.00	68.65		
ATOM	729	CA	GLY A 49	4.511	-4.678	3.416	1.00	68.65		
ATOM	730	C	GLY A 49	6.016	-4.568	3.568	1.00	68.65		
ATOM	731	O	GLY A 49	6.574	-4.419	4.642	1.00	68.65		
ATOM	732	H	GLY A 49	4.126	-4.947	5.519	1.00	0.00		
ATOM	733	HA2	GLY A 49	4.331	-5.491	2.733	1.00	0.00		
ATOM	734	HA3	GLY A 49	4.148	-3.770	2.940	1.00	0.00		
ATOM	735	N	THR A 50	6.712	-4.543	2.436	1.00	137.78		
ATOM	736	CA	THR A 50	8.148	-4.228	2.403	1.00	137.78		
ATOM	737	CB	THR A 50	9.021	-5.487	2.593	1.00	137.78		
ATOM	738	OG1	THR A 50	10.377	-5.148	2.428	1.00	137.78		
ATOM	739	CG2	THR A 50	8.790	-6.636	1.624	1.00	137.78		
ATOM	740	C	THR A 50	8.467	-3.420	1.154	1.00	137.78		
ATOM	741	O	THR A 50	8.148	-3.840	0.040	1.00	137.78		
ATOM	742	H	THR A 50	6.197	-4.605	1.565	1.00	0.00		
ATOM	743	HA	THR A 50	8.372	-3.575	3.247	1.00	0.00		
ATOM	744	HB	THR A 50	8.868	-5.855	3.610	1.00	0.00		
ATOM	745	HG21	THR A 50	9.503	-7.433	1.822	1.00	0.00		
ATOM	746	HG22	THR A 50	7.787	-7.032	1.754	1.00	0.00		

ext file

lengt

ATOM	2106	HE1	PHE	B	140	8.073	7.091	-0.878	1.00	0.00
ATOM	2107	HE2	PHE	B	140	4.372	5.070	-0.086	1.00	0.00
ATOM	2108	HZ	PHE	B	140	6.824	5.011	-0.321	1.00	0.00
ATOM	2109	N	GLY	B	141	2.203	8.593	-3.552	1.00	55.97
ATOM	2110	CA	GLY	B	141	1.641	7.323	-3.990	1.00	55.97
ATOM	2111	C	GLY	B	141	0.133	7.165	-3.989	1.00	55.97
ATOM	2112	O	GLY	B	141	-0.558	7.583	-3.072	1.00	55.97
ATOM	2113	H	GLY	B	141	1.587	9.327	-3.214	1.00	0.00
ATOM	2114	HA2	GLY	B	141	1.969	7.192	-5.008	1.00	0.00
ATOM	2115	HA3	GLY	B	141	2.066	6.519	-3.391	1.00	0.00
ATOM	2116	N	THR	B	142	-0.347	6.278	-4.854	1.00	59.76
ATOM	2117	CA	THR	B	142	-1.771	5.920	-4.913	1.00	59.76
ATOM	2118	CB	THR	B	142	-2.458	6.525	-6.144	1.00	59.76
ATOM	2119	OG1	THR	B	142	-2.310	7.923	-6.109	1.00	59.76
ATOM	2120	CG2	THR	B	142	-3.962	6.261	-6.155	1.00	59.76
ATOM	2121	C	THR	B	142	-1.917	4.411	-4.863	1.00	59.76
ATOM	2122	O	THR	B	142	-1.471	3.710	-5.774	1.00	59.76
ATOM	2123	H	THR	B	142	0.306	5.806	-5.461	1.00	0.00
ATOM	2124	HA	THR	B	142	-2.287	6.323	-4.043	1.00	0.00
ATOM	2125	HB	THR	B	142	-2.004	6.130	-7.053	1.00	0.00
ATOM	2126	HG1	THR	B	142	-2.404	8.265	-7.002	1.00	0.00
ATOM	2127	HG21	THR	B	142	-4.411	6.724	-7.033	1.00	0.00
ATOM	2128	HG22	THR	B	142	-4.161	5.192	-6.184	1.00	0.00
ATOM	2129	HG23	THR	B	142	-4.420	6.679	-5.258	1.00	0.00
ATOM	2130	N	PHE	B	143	-2.459	3.896	-3.761	1.00	104.32
ATOM	2131	CA	PHE	B	143	-2.849	2.494	-3.694	1.00	104.32
ATOM	2132	CB	PHE	B	143	-2.884	1.959	-2.270	1.00	104.32
ATOM	2133	CG	PHE	B	143	-1.623	2.041	-1.448	1.00	104.32
ATOM	2134	CD1	PHE	B	143	-1.273	3.244	-0.807	1.00	104.32
ATOM	2135	CD2	PHE	B	143	-0.871	0.875	-1.218	1.00	104.32
ATOM	2136	CE1	PHE	B	143	-0.154	3.285	0.041	1.00	104.32
ATOM	2137	CE2	PHE	B	143	0.239	0.914	-0.359	1.00	104.32
ATOM	2138	CZ	PHE	B	143	0.600	2.119	0.268	1.00	104.32
ATOM	2139	C	PHE	B	143	-4.230	2.290	-4.304	1.00	104.32
ATOM	2140	O	PHE	B	143	-5.141	3.082	-4.054	1.00	104.32
ATOM	2141	H	PHE	B	143	-2.830	4.518	-3.061	1.00	0.00
ATOM	2142	HA	PHE	B	143	-2.128	1.914	-4.258	1.00	0.00
ATOM	2143	HB2	PHE	B	143	-3.660	2.492	-1.749	1.00	0.00
ATOM	2144	HB3	PHE	B	143	-3.184	0.914	-2.325	1.00	0.00
ATOM	2145	HD1	PHE	B	143	-1.880	4.129	-0.937	1.00	0.00
ATOM	2146	HD2	PHE	B	143	-1.172	-0.060	-1.668	1.00	0.00
ATOM	2147	HE1	PHE	B	143	0.112	4.205	0.540	1.00	0.00
ATOM	2148	HE2	PHE	B	143	0.805	0.014	-0.163	1.00	0.00
ATOM	2149	HZ	PHE	B	143	1.451	2.147	0.933	1.00	0.00
END										

Εικόνες 32-33 Μετασχηματισμός των αρχείων .pdb των μοντέλων, σε αρχεία 5 αμινοξικών καταλοίπων στο κέντρο κάθε αλυσίδας (Dimerization Signal)

Επαναλαμβάνουμε την ίδια ακριβώς διαδικασία για κάθε αντίστοιχο αρχείο .pdb της δομής οδηγού, μετονομάζοντας το αρχείο σε ds template.pdb (π.χ 4yew-ds template).

SCALE3	0.000000	0.000000	0.015667	0.000000							
ATOM	317	N	GLY	A	46	8.874	-24.901	-11.063	1.00	63.41	N
ATOM	318	CA	GLY	A	46	8.485	-26.299	-11.186	1.00	62.61	C
ATOM	319	C	GLY	A	46	6.992	-26.415	-11.395	1.00	66.32	C
ATOM	320	O	GLY	A	46	6.551	-27.037	-12.365	1.00	67.28	O
ATOM	321	N	PHE	A	47	6.204	-25.775	-10.487	1.00	59.94	N
ATOM	322	CA	PHE	A	47	4.738	-25.716	-10.480	1.00	57.51	C
ATOM	323	C	PHE	A	47	4.185	-24.742	-11.546	1.00	56.92	C
ATOM	324	O	PHE	A	47	3.506	-25.186	-12.469	1.00	55.40	O
ATOM	325	CB	PHE	A	47	4.223	-25.373	-9.071	1.00	58.65	C
ATOM	326	CG	PHE	A	47	2.728	-25.425	-8.905	1.00	59.36	C
ATOM	327	CD1	PHE	A	47	2.099	-26.597	-8.506	1.00	62.13	C
ATOM	328	CD2	PHE	A	47	1.947	-24.286	-9.097	1.00	60.47	C
ATOM	329	CE1	PHE	A	47	0.713	-26.642	-8.334	1.00	62.23	C
ATOM	330	CE2	PHE	A	47	0.557	-24.338	-8.947	1.00	62.31	C
ATOM	331	CZ	PHE	A	47	-0.048	-25.513	-8.556	1.00	60.17	C
ATOM	332	N	GLY	A	48	4.462	-23.447	-11.410	1.00	50.97	N
ATOM	333	CA	GLY	A	48	3.981	-22.454	-12.367	1.00	50.36	C
ATOM	334	C	GLY	A	48	4.027	-21.021	-11.869	1.00	56.01	C
ATOM	335	O	GLY	A	48	4.425	-20.766	-10.730	1.00	55.54	O
ATOM	336	N	THR	A	49	3.596	-20.072	-12.720	1.00	53.27	N
ATOM	337	CA	THR	A	49	3.598	-18.647	-12.397	1.00	53.05	C
ATOM	338	C	THR	A	49	2.208	-17.990	-12.452	1.00	56.33	C
ATOM	339	O	THR	A	49	1.561	-18.007	-13.492	1.00	55.06	O
ATOM	340	CB	THR	A	49	4.586	-17.899	-13.334	1.00	61.51	C
ATOM	341	OG1	THR	A	49	5.837	-18.610	-13.403	1.00	68.68	O
ATOM	342	CG2	THR	A	49	4.809	-16.432	-12.917	1.00	51.21	C
ATOM	343	N	PHE	A	50	1.790	-17.359	-11.354	1.00	54.28	N
ATOM	344	CA	PHE	A	50	0.567	-16.567	-11.299	1.00	54.49	C
ATOM	345	C	PHE	A	50	0.987	-15.132	-11.698	1.00	64.47	C
ATOM	346	O	PHE	A	50	1.736	-14.475	-10.966	1.00	64.74	O
ATOM	347	CB	PHE	A	50	-0.051	-16.585	-9.891	1.00	55.18	C
ATOM	348	CG	PHE	A	50	-0.587	-17.919	-9.434	1.00	56.35	C
ATOM	349	CD1	PHE	A	50	-1.901	-18.299	-9.717	1.00	60.08	C
ATOM	350	CD2	PHE	A	50	0.203	-18.779	-8.685	1.00	57.85	C
ATOM	351	CE1	PHE	A	50	-2.398	-19.535	-9.290	1.00	60.41	C
ATOM	352	CE2	PHE	A	50	-0.297	-20.009	-8.251	1.00	60.80	C
ATOM	353	CZ	PHE	A	50	-1.597	-20.372	-8.547	1.00	59.19	C
TER	508		ASN	A	90						
ATOM	1240	N	GLY	C	46	-13.026	-18.994	-3.708	1.00	59.01	N
ATOM	1241	CA	GLY	C	46	-12.440	-17.748	-3.255	1.00	58.85	C
ATOM	1242	C	GLY	C	46	-11.269	-17.446	-4.167	1.00	64.19	C
ATOM	1243	O	GLY	C	46	-11.252	-16.415	-4.848	1.00	67.94	O
ATOM	1244	N	PHE	C	47	-10.342	-18.400	-4.282	1.00	54.95	N
ATOM	1245	CA	PHE	C	47	-9.179	-18.276	-5.151	1.00	51.68	C
ATOM	1246	C	PHE	C	47	-9.490	-18.688	-6.613	1.00	54.18	C
ATOM	1247	O	PHE	C	47	-9.677	-17.820	-7.456	1.00	55.54	O

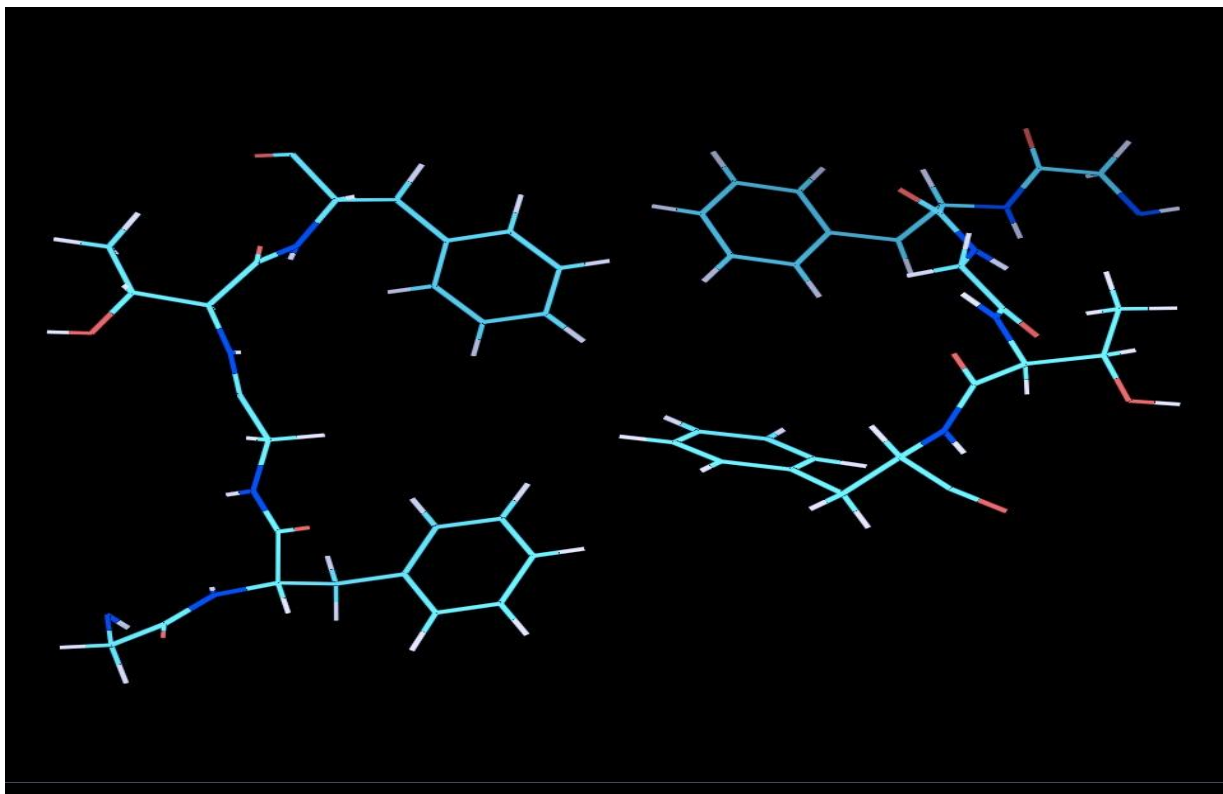
st file

length : 30 886

ATOM	350	CD2	PHE	A	50	0.203	-18.779	-8.685	1.00	57.85		
ATOM	351	CE1	PHE	A	50	-2.398	-19.535	-9.290	1.00	60.41		
ATOM	352	CE2	PHE	A	50	-0.297	-20.009	-8.251	1.00	60.80		
ATOM	353	CZ	PHE	A	50	-1.597	-20.372	-8.547	1.00	59.19		
TER	508		ASN	A	90							
ATOM	1240	N	GLY	C	46	-13.026	-18.994	-3.708	1.00	59.01		
ATOM	1241	CA	GLY	C	46	-12.440	-17.748	-3.255	1.00	58.85		
ATOM	1242	C	GLY	C	46	-11.269	-17.446	-4.167	1.00	64.19		
ATOM	1243	O	GLY	C	46	-11.252	-16.415	-4.848	1.00	67.94		
ATOM	1244	N	PHE	C	47	-10.342	-18.400	-4.282	1.00	54.95		
ATOM	1245	CA	PHE	C	47	-9.179	-18.276	-5.151	1.00	51.68		
ATOM	1246	C	PHE	C	47	-9.490	-18.688	-6.613	1.00	54.18		
ATOM	1247	O	PHE	C	47	-9.677	-17.820	-7.456	1.00	55.54		
ATOM	1248	CB	PHE	C	47	-8.000	-19.057	-4.546	1.00	51.68		
ATOM	1249	CG	PHE	C	47	-6.681	-18.878	-5.230	1.00	50.89		
ATOM	1250	CD1	PHE	C	47	-6.040	-17.643	-5.227	1.00	51.59		
ATOM	1251	CD2	PHE	C	47	-6.023	-19.966	-5.789	1.00	51.17		
ATOM	1252	CE1	PHE	C	47	-4.788	-17.495	-5.817	1.00	51.52		
ATOM	1253	CE2	PHE	C	47	-4.770	-19.818	-6.358	1.00	53.35		
ATOM	1254	CZ	PHE	C	47	-4.171	-18.579	-6.388	1.00	51.12		
ATOM	1255	N	GLY	C	48	-9.537	-19.989	-6.900	1.00	47.90		
ATOM	1256	CA	GLY	C	48	-9.778	-20.518	-8.235	1.00	45.69		
ATOM	1257	C	GLY	C	48	-9.733	-22.031	-8.341	1.00	49.58		
ATOM	1258	O	GLY	C	48	-9.747	-22.745	-7.331	1.00	50.16		
ATOM	1259	N	THR	C	49	-9.726	-22.539	-9.580	1.00	46.17		
ATOM	1260	CA	THR	C	49	-9.694	-23.984	-9.832	1.00	45.77		
ATOM	1261	C	THR	C	49	-8.500	-24.423	-10.657	1.00	47.98		
ATOM	1262	O	THR	C	49	-8.181	-23.802	-11.678	1.00	46.99		
ATOM	1263	CB	THR	C	49	-11.087	-24.506	-10.325	1.00	50.90		
ATOM	1264	OG1	THR	C	49	-11.908	-24.656	-9.170	1.00	51.31		
ATOM	1265	CG2	THR	C	49	-11.022	-25.877	-11.026	1.00	44.59		
ATOM	1266	N	PHE	C	50	-7.853	-25.494	-10.195	1.00	44.62		
ATOM	1267	CA	PHE	C	50	-6.742	-26.159	-10.872	1.00	45.47		
ATOM	1268	C	PHE	C	50	-7.353	-27.393	-11.510	1.00	52.60		
ATOM	1269	O	PHE	C	50	-7.877	-28.273	-10.823	1.00	51.15		
ATOM	1270	CB	PHE	C	50	-5.593	-26.534	-9.900	1.00	46.11		
ATOM	1271	CG	PHE	C	50	-4.894	-25.346	-9.283	1.00	46.57		
ATOM	1272	CD1	PHE	C	50	-3.867	-24.692	-9.958	1.00	47.77		
ATOM	1273	CD2	PHE	C	50	-5.255	-24.888	-8.024	1.00	47.36		
ATOM	1274	CE1	PHE	C	50	-3.214	-23.606	-9.379	1.00	47.75		
ATOM	1275	CE2	PHE	C	50	-4.605	-23.805	-7.447	1.00	49.67		
ATOM	1276	CZ	PHE	C	50	-3.576	-23.181	-8.120	1.00	47.42		
TER	1437		LYS	C	90							
MASTER	328	0	0	6	6	0	0	6	1433	4	0	16
END												

Εικόνες 34-35 Μετασχηματισμός των αρχείων .pdb των δομών-οδηγών (templates), σε αρχεία 5 αμινοξικών καταλοίπων στο κέντρο κάθε αλυσίδας (Dimerization Signal).

Το επόμενο βήμα είναι να ομαδοποιήσουμε τα μοντέλα εκείνα που έχουν κοινά templates. Τέλος εκτελέστηκε υπέρθεση δομών ανάμεσα στα DS των μοντέλων και στο DS του κοινού template, προκειμένου να παρατηρήσουμε και οπτικά τις αλλαγές στις επιμέρους στερεοδιατάξεις. Στο παράδειγμα μας απεικονίζεται η υπέρθεση των DS των b8gx11 και r0can2 που έχουν ως template την 4YEW.



Εικόνα 36 Στιγμιότυπο απεικόνισης του DS της δομής-οδηγός 4YEW_A και του μοντέλου BG8X11 σε υπέρθεση (Παρατηρείται ταύτιση δομών που όπως θα δούμε παρακάτω μαρτυράει κοινή τιμή της ελεύθερης ενέργειας ΔG)

3.5 Υπολογισμός των ενδομοριακών και διαμοριακών αλληλεπιδράσεων στα μοντέλα των ΗΥ

Στην συνέχεια εισάγουμε το βελτιστοποιημένο μοντέλο στον αλγόριθμο PIC (Protein Interaction Calculator) στην ηλεκτρονική διεύθυνση <http://pic.mbu.iisc.ernet.in/job.html> [67] μέσω του οποίου υπολογίζονται οι ενδομοριακές και διαμοριακές αλληλεπιδράσεις. Χαρακτηριστικά αναφέρουμε ότι υπολογίζονται σε ενδομοριακό επίπεδο, οι υδροφοβικές, οι αρωματικές και οι ιοντικές αλληλεπιδράσεις καθώς και οι υδρογονικοί δεσμοί σε κάθε αλυσίδα. Σε διαμοριακό επίπεδο ισχύουν οι μετρήσεις του ενδομοριακού επιπέδου και συμπληρώνονται τόσο με τις αλληλεπιδράσεις αρωματικών δακτυλίων με αρωματικούς δακτυλίους, όσο και με τις αλληλεπιδράσεις αρωματικών δακτυλίων με θείο (S).

P I C : Protein Interactions Calculator
Molecular Biophysics Unit, Indian Institute of Science, Bangalore.

HOME HELP CRITERIA **SUBMIT JOB** CONTACT US LAB PAGE

NOTE:
Upload query structure in the three-dimensional coordinate set format of the protein data bank. The filename should be a four letter code with a .pdb or .ent or .atm or .txt extensions. e.g.,: 1EJG.pdb

INTRAPROTEIN INTERACTIONS
* This option can accept a monomeric or a multichain protein file.

Upload a file in PDB format: no file selected

- Hydrophobic Interactions Enter the interaction cut-off value (Default 5A)
- Disulphide Bridges
- Main Chain-Main Chain Hydrogen Bonds
- Main Chain-Side Chain Hydrogen Bonds
- Side Chain-Side Chain Hydrogen Bonds
- Ionic Interactions Enter the interaction cut-off value (Default 6A)
- Aromatic-Aromatic Interaction : cut-off value to (Default 4.5A to 7A)
- Aromatic-Sulphur Interactions Enter the interaction cut-off value (Default 5.3A)
- Cation-Pi Interactions Enter the interaction cut-off value (Default 6A)

To E-mail the results enter a valid e-mail address: (Optional)

PROTEIN-PROTEIN INTERACTIONS

* This option can accept only a multichain protein file.

Upload a file in PDB format: no file selected

- Hydrophobic Interactions** Enter the interaction cut-off value (Default 5A)
- Disulphide Bridges**
- Main Chain-Main Chain Hydrogen Bonds**
- Main Chain-Side Chain Hydrogen Bonds**
- Side Chain-Side Chain Hydrogen Bonds**
- Ionic Interactions** Enter the interaction cut-off value (Default 6A)
- Aromatic-Aromatic Interactions** : cut-off value to (Default 4.5A to 7A)
- Aromatic-Sulphur Interactions** Enter the interaction cut-off value (Default 5.3A)
- Cation-Pi Interactions** Enter the interaction cut-off value (Default 6A)

To E-mail the results enter a valid e-mail address: (Optional)

[Top](#)

Solvent accessibility is calculated using NACCESS. Residues that has all atoms relative accessibility value greater than 7 are considered buried. Residues that has all atoms relative accessibility value less than or equal to 7 are considered exposed.

SOLVENT ACCESSIBILITY BASED CALCULATIONS

Upload a file in PDB format: no file selected

Calculation of Intraprotein Interactions based on solvent accessibility

* This option can accept a monomeric or a multichain protein file

- Interactions between buried residues**
- Interactions between exposed residues**

Calculation of Protein-Protein Interactions based on solvent accessibility

* This option can accept only a multichain protein file.

Εικόνα 23-24 Αποτύπωση αρχικής σελίδας PIC στο οποίο απεικονίζονται οι επιθυμητές μετρήσεις που θα γίνουν από τον χρήστη είτε σε ενδομοριακό είτε σε διαμοριακό επίπεδο. Αξιοσημείωτο είναι ότι εκτός από τις προκαθορισμένες τιμές της απόστασης ο χρήστης έχει την δυνατότητα να εισάγει και εναλλακτικές τιμές.

4. Αποτελέσματα

4.1 Συλλογή και αξιολόγηση αμινοξικών ακολουθιών στις βάσεις δεδομένων.

Χρησιμοποιήθηκε αρχικά η βάση δεδομένων Genbank. Πραγματοποιήθηκε αναζήτηση (query) στο πεδίο αναζήτησης με την λέξη κλειδί "histone like DNA binding protein" και επιλέγοντας στην επιλογή Protein. Από την αρχική αναζήτηση προέκυψαν συνολικά 161.919 καταχωρήσεις. Στην συνέχεια, αποθηκεύτηκε η αναζήτηση τοπικά στον σκληρό δίσκο του υπολογιστή με την μορφή αρχείου .txt. Ακολούθως, συντάχτηκε κώδικας σε γλώσσα προγραμματισμού Perl. Το αρχείο .txt χρησιμοποιήθηκε ως αρχείο εισόδου προκειμένου να σαρωθεί με τις ανάλογες λέξεις κλειδιά που ορίστηκαν, και στην συνέχεια μέσω pattern matching να γίνει ο επιθυμητός διαχωρισμός της υφιστάμενης πληροφορίας. Το αρχείο κώδικα καθώς και οι λέξεις κλειδιά που χρησιμοποιήθηκαν αναφέρονται χαρακτηριστικά στην αντίστοιχη ενότητα του κεφαλαίου της μεθοδολογίας.

Με την παραπάνω διαδικασία από τις συνολικά 161.919 καταχωρήσεις στην Genbank, βρέθηκαν 120.009 να απαντώνται σε προκαρυωτικούς, 38.399 σε ευκαρυωτικούς οργανισμούς και 391 σε Αρχαία. Στην ανωτέρω βάση δεδομένων από την καταχώρηση δεν ήταν δυνατό να ομαδοποιηθούν τα αποτελέσματα με βάση τον σχολιασμό των καταχωρήσεων, αν δηλαδή πρόκειται για δευτερογενή αρχεία (reviewed) ή για πρωτογενή (unreviewed) και όπως ήταν αναμενόμενο τα εν λόγω αποτελέσματα δεν χρησιμοποιήθηκαν.

Σε αναλογία με την Genbank, ακολουθήσαμε την ίδια διαδικασία και με την βάση δεδομένων της Uniprot. Στην περίπτωση της Uniprot, έγινε αναζήτηση (query) στο πεδίο εισαγωγής με την ίδια λέξη κλειδί "histone like DNA binding protein". Τα αποτελέσματα αυτής της αναζήτησης ήταν 109.441 καταχωρήσεις. Στην συνέχεια αποθηκεύτηκαν τα δεδομένα σε μορφή αρχείου .txt και περαιτέρω σάρωση τους με αρχείο κώδικα γραμμένο σε γλώσσα προγραμματισμού Perl μέσω της διαδικασίας του Pattern Matching με λέξεις κλειδιά, όπως αναλύθηκε εκτενώς στο κεφάλαιο της

μεθοδολογίας. Ως αρχείο εισόδου, σε αναλογία με την προγενέστερη καταχώρηση, χρησιμοποιήθηκε το αρχείο αρχικής αποθήκευσης των καταχωρήσεων .txt.

Πρόεκυψαν 56.972 καταχωρήσεις να απαντώνται σε προκαρυωτικούς οργανισμούς, 50.631 σε ευκαρυωτικούς οργανισμούς, 1.101 στα Αρχαία ενώ 1.838 βρέθηκαν να μην ανήκουν σε καμία από τις παραπάνω κατηγορίες μάλλον διότι μπορεί να προέρχεται η πληροφορία είτε από ελεύθερο DNA είτε ο οργανισμός να μην είναι πιστοποιημένος. Επίσης με βάση τον σχολιασμό των καταχωρήσεων, αν δηλαδή πρόκειται για δευτερογενή αρχεία (reviewed) ή για πρωτογενή (unreviewed) βρέθηκαν 5.015 Reviewed και 104.426 Unreviewed.

Στην συνέχεια το αρχείο εξόδου με τα 5.015 Reviewed αποτελέσματα χρησιμοποιήθηκε ως αρχείο εισόδου σε νέα σάρωση προκειμένου να διαπιστωθεί ποια από τις 5.015 καταχωρήσεις ανήκουν σε προκαρυωτικούς, ποιες σε ευκαρυωτικούς οργανισμούς και ποιές στα Αρχαία. Το αποτέλεσμα ήταν ότι 714 καταχωρήσεις ανήκουν σε προκαρυωτικούς οργανισμούς, 4124 σε ευκαρυωτικούς οργανισμούς και 177 θεωρούνται αχαρακτήριστα.

Πίνακας 1.Κατανομή καταχωρήσεων σε Reviewed και Unreviewed, σε Ευκαρυωτικούς και Προκαρυωτικούς, Αρχαία και Λοιπούς Οργανισμούς από τις δυο βάσεις δεδομένων, Genbank και Uniprot.

Βάσεις	Αρχικές καταχωρήσεις	Reviewed	Unreviewed	Ευκαρυωτικά	Προκαρυωτικά	Αρχαία	Λοιποί Οργανισμοί
Genbank	161.919	-	-	38.399	120.009	391	3.120
Uniprot	109.441	5.015	104.426	50.631	56.972	1.101	690

Στην συνέχεια απομονώσαμε από την πρώτη αναζήτηση μας, τις 714 Reviewed οπού και τις χρησιμοποιήσαμε ως αρχείο εισόδου πλέον, σε νέο αρχείο κώδικα Perl, με τις κατάλληλες λέξεις κλειδιά, ώστε να διαχωρίσουμε ποιές από αυτές είναι HU και ποιές είναι IHF. Το αρχείο εξόδου μας έδειξε ότι 542 ανήκουν στην οικογένεια HU ενώ οι 172 στην οικογένεια IHF. Τέλος, με μία γρήγορη σάρωση, παρατηρούμε άμεσα ότι αρκετές από τις 542 καταχωρήσεις έχουν όμοιο AccessionID, και ως εκ τούτου ο αριθμός των πρωτεϊνών που μας ενδιαφέρουν, περιορίζεται από 542 σε 40.

4.2 Συλλογή δομικών δεδομένων στη βάση PDB

Κάθε πρωτεϊνική ακολουθία των HU ακολουθιών που επιλέχθηκαν από την UNIPROT χρησιμοποιήθηκε στο πρόγραμμα BLAST και μέσω του οποίου αναζητήσαμε, με τις προκαθορισμένες (default) παραμέτρους, στη βάση δεδομένων PDB, όπως περιγράφηκε στη μεθοδολογία. Στόχος ήταν να εντοπίσουμε την πλησιέστερη λυμένη κρυσταλλογραφική δομή που έχει κατατεθεί στην βάση PDB. Οι βασικές παράμετροι που αποτελούν κριτήριο για την επιλογή είναι η χαμηλότερη δυνατή τιμή στο πεδίο του E-Value και η υψηλότερη δυνατή τιμή στο πεδίο Query Coverage. Τα αποτελέσματα συνοψίζονται στον παρακάτω Πίνακα 2.

Πίνακας 2. Συνοπτική παρουσίαση των HU πρωτεϊνών που θα μοντελοποιηθούν (UNIPROT ID, target sequences) με την αντίστοιχη κρυσταλλογραφική δομή που έχει κατατεθεί στην βάση PDB ID που θα χρησιμοποιηθεί ως πρότυπο (template) σε συνδυασμό με την τιμή Ομοιότητας (E-value), του ποσοστού Ταυτότητας (identity) και του ποσοστού Αλληλοεπικάλυψης (query coverage).

No	UNIPROT ID (target sequence)	PDB ID (template)	%Ταυτότητα (Identity)	Ομοιότητα (E-value)	% Αλληλόεπικάλυψη (Query Coverage)
1	B8GX11	4YEW_A	49%	4e-21	95%
2	P0CAV2		95%	4e-21	49%
3	P0ACF1	1MUL_A	100%	3e-60	100%
4	P0ACF2		100%	3e-60	100%
5	P0A1R6	4YEY_A	99%	1e-59	100%
6	P0A1R7		99%	1e-59	100%
7	EOJ6W8		100%	3e-60	100%
8	P0ACF3		100%	3e-60	100%
9	P0A1R8	2O97_B	99%	8e-58	100%
10	P0A1R9		99%	8e-58	100%
11	P0ACF6		100%	2e-58	100%
12	P0ACF7		100%	2e-58	100%
13	P0ACF5		100%	2e-58	100%

14	P64388		100%	2e-37	70%
15	P64389		100%	2e-37	70%
16	Q9JR30		98%	4e-32	59%
17	Q9KHS6		100%	1e-40	73%
18	P0C0H2		100%	9e-54	90%
19	P0C0H3		100%	9e-54	90%
20	P0DB64	5FBM_A	100%	9e-54	90%
21	P0DB65		100%	9e-54	90%
22	P0A3H1		100%	1e-60	100%
23	P0A3H2		100%	1e-60	100%
24	P64386		90%	4e-11	30%
25	P64387		90%	4e-11	30%
26	P68573	1HUE_A	97%	9e-42	71%
27	P68574		97%	9e-42	71%
28	Q5HUP6		90%	5e-30	54%
29	Q9Z8C7		90%	1e-11	31%
30	Q46121		90%	5e-30	54%
31	P0DMK4		96%	7e-28	54%
32	I1WEI8		54%	7e-28	96%
33	P0A3H3	4QJN_A	62%	4e-34	97%
34	P0A3H4		62%	4e-34	97%
35	P0A3H5		50%	5e-20	94%
36	P0A3H6	3RHI_A	50%	5e-20	94%
37	O68451	1IHF_B	36%	5e-09	70%
38	P0A3H7		57%	3e-26	43%
39	P0A3H8	4DKY_A	57%	3e-26	43%
40	Q9XB18		100%	1e-69	46%

Σύμφωνα με τον παραπάνω πίνακα οι πρωτεϊνικές ακολουθίες μας βρέθηκαν να έχουν πολύ καλό ποσοστό τόσο αλληλοεπικάλυψης όσο και ταυτότητας καθώς και ικανοποιητική τιμή ομοιότητας, εκτός από τα μοντέλα P0A3H7, P0A3H8, Q9XB18 με κοινή δομή-οδηγό την 4DKY_A τα οποία και απορρίφθηκαν, διότι παρουσιάζουν πολύ χαμηλό ποσοστό αλληλοεπικάλυψης. Το εν λόγω ποσοστό που προκύπτει καθιστά αδύνατη την δημιουργία in silico μοντέλων των εν λόγω πρωτεϊνικών ακολουθιών.

4.3 Στοιχίση των πρωτεϊνικών αλληλουχιών στόχων με τις αντίστοιχες αλληλουχίες των δομών

Στην συνέχεια εκτελέστηκε η διαδικασία της στοιχίσης μέσω CLUSTAL Omega μεταξύ άγνωστης πρωτεϊνικής ακολουθίας και ακολουθίας δομής-οδηγού, ώστε να εντοπιστεί και να τεκμηριωθεί η ομολογία των δύο πρωτεϊνικών ακολουθιών, μέσω της ομοιότητας που εντοπίζεται μέσω της σύγκρισης. Επιπλέον κατά την στοιχίση εντοπίζονται τα συντηρητικά αμινοξικά κατάλοιπα και η οικογένεια της άγνωστης ακολουθίας, προκειμένου να εισαχθεί και να είναι η δυνατή η κατασκευή in silico μοντέλου μέσω της διεπαφής του Modeller που υπάρχει στο PyMod 2.0. Από την έξοδο του αρχείου στοιχίσης, απεικονίζεται κάτω από κάθε ζεύγος σε μορφή CLUSTAL. Η μορφή CLUSTAL (η οποία προήλθε από το ομώνυμο πρόγραμμα), είναι πιο απλή στην αρχή (δεν περιέχει σε ξεχωριστή ενότητα τα ονόματα των ακολουθιών, και οι ακολουθίες δίνονται συνεχόμενα, χωρίς κενά), αλλά περιέχει μια επιπλέον γραμμή σε κάθε τμήμα το οποίο διαχωρίζει 60 αμινοξικά κατάλοιπα της στοιχίσης. Στον Πίνακα 3 συνοψίζονται όλες οι στοιχίσεις των ακολουθιών στόχων με τις αντίστοιχες πρωτεϊνικές ακολουθίες των δομών-οδηγών όπως εκείνες προέκυψαν από την αναζήτηση BLAST στην PDB που αναφέρθηκε στην μεθοδολογία.

Πίνακας 3 Στοιχισή κατά ζεύγη των ακολουθιών HU της Uniprot με τις αντίστοιχες ακολουθίες των πρωτεϊνών με λυμένη δομή(PDB template) με την χρήση του προγράμματος CLUSTALW. Το «:» και το «.» συμβολίζουν μεγάλη και μικρότερη συντήρηση αντίστοιχα (εξαρτώνται από τον πίνακα ομοιότητας, όχι μόνο από το ποσοστό), ενώ το κενό (« ») συμβολίζει καμία ταύτιση.

```

sp|B8GX11 (target)      MTTKAEVLVTAIAEKAGINKNQAKDALEAFIEAVTDSLKSGQDVRVLVGFCTFKAVTRAAGTARNPRTGETVNRPAKSTARFQVGEGLKSSLNS
sp|P0CAV2 (target)    MTTKAEVLVTAIAEKAGINKNQAKDALEAFIEAVTDSLKSGQDVRVLVGFCTFKAVTRAAGTARNPRTGETVNRPAKSTARFQVGEGLKSSLNS
4YEW:A (template)    -MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
.:.:*: ** *.* * * **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|E0J6W8 (target)    -MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
sp|P0ACF3 (target)    -MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
sp|P0A1R6 (target)    -MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
sp|P0A1R7 (target)    -MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
4YEY:A (template)    AMNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
*****:.* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|I1WEI8 (target)    MNKQELIDAVAAQTGASKAQGTETLDTLLEVIKKAIVSKGDSVQLIGFSGFGSKRAARTGRNPKTGETIKIPAAKTVKFTAGKAFKDAVNKR-----
sp|P0DMK4 (target)    MNKQELIDAVAAQTGASKAQGTETLDTLLEVIKKAIVSKGDSVQLIGFSGFGSKRAARTGRNPKTGETIKIPAAKTVKFTAGKAFKDAVNKR-----
sp|P0A3H3 (target)    MNKNELVSAVAEKAGLTKDAASAVDAVFETVQSELKNGDIRLAGFGSFSVSRREASKGRNPSTGAEVDIPARNVFKFSAGKGLKDAVNS-----
sp|P0A3H4 (target)    MNKTDLINAQADLTKEEAGSVAADFVESIQNSLAKGEKQVLIQFGNFEVREARAARKGRNPQTGKEIDI PASKVPAPFKAGKALKDAVKLEHHHHHH
4QJN:A (template)    MNKTDLINAQADLTKEEAGSVAADFVESIQNSLAKGEKQVLIQFGNFEVREARAARKGRNPQTGKEIDI PASKVPAPFKAGKALKDAVKLEHHHHHH
**:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|O68451 (target)    MITKNYLIDKIHDKLNYLSKEDVKSDVLDLIDVLESKQKRIEIRNFGNFSIRKRFPESEKFNVTV-----Y-----YRMPKNLFKE
1IHF:B (template)    -MTKSELIERLATQQSHIPAKTVEADKEMLEHMASTLAQGERIEIRGFGSFSLSHYRAPRTGRNPKTGDVLEGGVPHFKPKGKELDRDRANIYG-
**:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|P0A3H5 (target)    ---MNRSELVAALADRAEVTRKDAADAVLAFAEVDIVSKGDEKVTIPGFLTERTHRAARTARNPQTGEPQIPAGYSVKVSAGSKLKEAAGK
sp|P0A3H6 (target)    ---MNRSELVAALADRAEVTRKDAADAVLAFAEVDIVSKGDEKVTIPGFLTERTHRAARTARNPQTGEPQIPAGYSVKVSAGSKLKEAAGK
3RHI:A (template)    SNMKNTELKINVAQNAEISQKEATVVVQTVVESITNTLA-AGEKQVLIQFGNFEVREARAARTGRNPQTGEEQIAASKVPAPFKAGKLEKAVK--
**:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|P0ACF1 (target)    MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
sp|P0ACF2 (target)    MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
1MUL:A (template)    MNKTQLIDVIAEKAELSKTQAKAALESTLAAITESLKEGDAVQLVGFCTFKVNHRAERTGRNPQTGKEIKIAAANVPFVSGKALKDAVK
*****:.* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|P64574 (target)    MNNHFGKGLM-----AGLKATHADSAVNVTKFCADYKRGFVLVGYSHRMYEKTGDRQLSAWEAGILTRR-----YGLDKEMVMDFRENNSCSTLRFFMAGYRLEN-----
sp|P64386 (target)    MATMTKKKLISITISQDHIHPNHVVRTVIQ--NPLDKMTDALVQG-----DR-LEFRDFGLVQVVERKPKVGRNPKNA--AVPIHI PARRAVKFTPGKRMKRLIETPTKSS--
sp|P64387 (target)    MATMTKKKLISITISQDHIHPNHVVRTVIQ--NPLDKMTDALVQG-----DR-LEFRDFGLVQVVERKPKVGRNPKNA--AVPIHI PARRAVKFTPGKRMKRLIETPTKSS--
sp|Q9Z8C7 (target)    MATMTKKKLISITISQDHIHPNHVVRTVIQ--NPLDKMTDALVQG-----DR-LEFRDFGLVQVVERKPKVGRNPKNA--AVPIHI PARRAVKFTPGKRMKRLIETPTKSS--
sp|P0A3H1 (target)    ---MNKTELEINAQADLTKEEAGSVAADFVESIQNSLAKGEKQVLIQFGNFEVREARAARKGRNPQTG--EEMEIPASKVPAPFKPGKALKDAVK-----
sp|P0A3H2 (target)    ---MNKTELEINAQADLTKEEAGSVAADFVESIQNSLAKGEKQVLIQFGNFEVREARAARKGRNPQTG--EEMEIPASKVPAPFKPGKALKDAVK-----
sp|Q5HUP6 (target)    ---MTKADFLISLVAQTAGLTKKDATTATD--AVISTITDVLAKG-----DS-ISFIGFGTFTSTQERAAREARVPSTG--KTIKVPATRVAKFKVGNLKEAVAKASGKSKK
sp|Q46121 (target)    ---MTKADFLISLVAQTAGLTKKDATTATD--AVISTITDVLAKG-----DS-ISFIGFGTFTSTQERAAREARVPSTG--KTIKVPATRVAKFKVGNLKEAVAKASGKSKK
1HUE:A (template)    ---MNKTELEINAQADLTKEEAGSVAADFVESIQNSLAKGEKQVLIQFGNFEVREARAARKGRNPQTG--EEMEIPASKVPAPFKPGKALKDAVK-----
.:.:*: ** *.* * * **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|Q9JR30 (target)    MNKSELIQAIADAELS KRAAAEFVNAFVSVVTVQELKGDNDVTLVGFCTFHAAQSAERQGRNPKTGEPLTIAAKT PKFRSGKALKDAVNG
sp|P64388 (target)    MNKSELIQAIQAEDISKAAQKALDATTNAVITALKQGDVTVLVGFCTFYVGERAERQGRNPKTGEPLTIAAAKTPKFRAGKALKDAL--
sp|P64389 (target)    MNKSELIQAIQAEDISKAAQKALDATTNAVITALKQGDVTVLVGFCTFYVGERAERQGRNPKTGEPLTIAAAKTPKFRAGKALKDAL--
sp|Q9KHS6 (target)    MNKSELIDAIQAASADLPKAAAGRALDAVIESVTGALKAGDSVVLVGFCTFVSDRPAIRGRNPQTGKLEIAAAKPKGFKAGKALKEAVN-
sp|P0A1R8 (target)    MNKSQLIEKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
sp|P0A1R9 (target)    MNKSQLIEKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
sp|P0ACF5 (target)    MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
sp|P0ACF6 (target)    MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
sp|P0ACF7 (target)    MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
2097:B (template)    MNKSQLIDKIAAGADISKAAAGRALDAIIASVTESLKEGDDVALVGFCTFAVKERAARTGRNPQTGKEITIAAAKVPSPFRAGKALKDAVN-
*****:.* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

sp|P0C0H2 (target)    MANKQDLIAKVAEATELTKKDSAAAVDAVFSTIEAF LAEGEKVQLIGFNGFEVREARAARKGRNPQTGAEIEIAASKVPAPFKAGKALKDAVK-----
sp|P0C0H3 (target)    MANKQDLIAKVAEATELTKKDSAAAVDAVFSTIEAF LAEGEKVQLIGFNGFEVREARAARKGRNPQTGAEIEIAASKVPAPFKAGKALKDAVK-----
sp|P0DB65 (target)    MANKQDLIAKVAEATELTKKDSAAAVDAVFSTIEAF LAEGEKVQLIGFNGFEVREARAARKGRNPQTGAEIEIAASKVPAPFKAGKALKDAVK-----
sp|P0DB64 (target)    MANKQDLIAKVAEATELTKKDSAAAVDAVFSTIEAF LAEGEKVQLIGFNGFEVREARAARKGRNPQTGAEIEIAASKVPAPFKAGKALKDAVK-----
5F6M:A (template)    MANKQDLIAKVAEATELTKKDSAAAVDAVFSAVSYLAKGEKQVLIQFGNFEVREARAARKGRNPQTGEEIKIKASKVPAPFKAGKALKDAVKHHHHHHHH
*****:.* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:* **:*

```

4.4 Κατασκευή των μοντέλων HU με την τεχνική του Homology Modeling

Εύλογα καταλήγουμε στο συμπέρασμα ότι με την μέθοδο του Homology Modeling προσπαθούμε να προσδιορίσουμε θεωρητικά την άγνωστη δομή μιας πρωτεΐνης, εφόσον υπάρχει σημαντική ομοιότητα με μια πρωτεΐνη γνωστής δομής. Σε γενικές γραμμές, πρωτεΐνες με ομοιότητα στην ακολουθία άνω του 30%, η οποία εκτείνεται σε μήκος μεγαλύτερο των 80 καταλοίπων, θεωρείται πως διπλώνουν στο χώρο με παρόμοιο τρόπο (Ref). Τουλάχιστον αυτό ισχύει για τις περιοχές συγκεκριμένης (μη τυχαίας) δευτεροταγούς δομής (Ref). Η γνωστή δομή λέγεται οδηγός (template), ενώ η άγνωστη στόχος (target).

Η αναλυτική διαδικασία του Homology Modeling έχει ως εξής:

1. Εντοπισμός της ακολουθίας του μοντέλου στις βάσεις δεδομένων.
2. Εύρεση μιας κατάλληλης πρωτεΐνης που να ικανοποιεί τις απαιτήσεις ομολογίας με την ακολουθία του μοντέλου.
3. Εντοπισμός της πειραματικά προσδιορισμένης δομής-οδηγού στις βάσεις δεδομένων.
4. Στοιχίση των δύο ακολουθιών.
5. Δημιουργία του μοντέλου, με ειδικό λογισμικό το οποίο δέχεται ως δεδομένα τη δομή-οδηγό και τη στοιχίση των ακολουθιών.
6. Οπτική παρατήρηση του μοντέλου, και έλεγχος για πιθανές ανεπίτρεπτες αποστάσεις μεταξύ ατόμων.
7. Ενδεχόμενη βελτίωση της στοιχίσης των ακολουθιών, με σκοπό τη βελτίωση του μοντέλου.
8. Δημιουργία του νέου μοντέλου.
9. Έλεγχος του ενεργειακά ελαχιστοποιημένου μοντέλου.

10. Αξιολόγηση του μοντέλου, ταιριάζοντάς το στο χώρο με κάποια πειραματικά προσδιορισμένη δομή (τη δομή - οδηγό ή κάποια άλλη).

Ακολουθώντας την διαδικασία που περιγράψαμε περιληπτικά παραπάνω και εκτενέστερα στο αντίστοιχο κομμάτι της μεθοδολογίας κατασκευάστηκαν 37 πρωτεϊνικά μοντέλα HU.

Για την αξιολόγηση των μοντέλων έγιναν μια σειρά από αξιολογήσεις (models validation) όπως:

1. Χρήση του διαδικτυακού εξυπηρετητή WhatIF το οποίο και πραγματοποιεί μια σειρά από ελέγχους (tests) όπως μήκος δεσμών και γωνίες δεσμών.
2. Ενεργειακή ελαχιστοποίηση των υφιστάμενων μοντέλων μέσω του Chimera.
3. Χρήση του διαδικτυακού εξυπηρετητή WhatIF στα βελτιστοποιημένα μοντέλα. Με βάση τον παραπάνω αλγόριθμο εφόσον η τιμή της αξιολόγησης των διαμοριακών και ενδομοριακών αποστάσεων, το μήκος και τις γωνίες των δεσμών είναι μικρότερη από 1.00 τότε θεωρείται ένα δομικά αποδεκτό μοντέλο.
4. Υπέρθωση με το template και υπολογισμός για κάθε ζευγάρι του RMS (Root Mean Square) με το πρόγραμμα Rymol.
5. Εκτέλεση και αξιολόγηση των μοντέλων μέσω του PIC ώστε να καταμετρηθούν οι υδρογονικοί και οι ιοντικοί δεσμοί, να εντοπιστούν τα υδρόφοβα τμήματα τόσο σε ενδομοριακό όσο και διαμοριακό επίπεδο.

4.5 Αξιολόγηση μονομερών μοντέλων HU μέσω του WhatIF

Τα υπολογισμένα μοντέλα αξιολογήθηκαν (models validation) με τα παρακάτω βήματα:

1. Χρήση του διαδικτυακού προγράμματος WhatIF το οποίο και πραγματοποιεί μια σειρά από ελέγχους:

a. μήκος δεσμών,

b. γωνίες δεσμών,

2. Χρήση του διαδικτυακού προγράμματος WhatIF στα ενεργειακά ελαχιστοποιημένα μοντέλα. Με βάση τον παραπάνω αλγόριθμο εφόσον η τιμή της αξιολόγησης των διαμοριακών και ενδομοριακών αποστάσεων, το μήκος και τις γωνίες των δεσμών είναι μικρότερη από 1.00 τότε θεωρείται ένα δομικά αποδεκτό μοντέλο.

3. Υπέρθεση με το template και υπολογισμός για κάθε ζευγάρι του rms με το πρόγραμμα PyomoL.

Ακολουθώντας την διαδικασία που περιγράψαμε στη μεθοδολογία υπολογίστηκαν και αξιολογήθηκαν 37 πρωτεϊνικά μοντέλα HU. Ο Πίνακας 4 παρουσιάζει το σύνολο των πρωτεϊνικών μοντέλων των HU που επιλέξαμε και μελετάμε.

Πίνακας 4. Συνοπτική παρουσίαση των HU πρωτεϊνών που μοντελοποιήθηκαν (UNIPROT ID) με τη χρήση του αντίστοιχου PDB ID template. Κάθε μοντέλο της HU ονομάστηκε όπως φαίνεται στη στήλη MODEL στη στήλη WhatIF εμφανίζεται το αρχικό σκορ και στην στήλη RMS παρατίθεται το σκορ της υπέρθεσης μεταξύ της δομής στόχου και της δομής-οδηγού.

No	UNIPROTID	PDB ID	WhatIF	RMS(PyMol)	MODEL
1	B8GX11	4YEW_A	7.939	1.973	B8GX11-4YEW_A
2	EOJ6W8	4YEY_A	9.896	1.720	EOJ6W8-4YEY_A
3	I1WEI8	4QJN_A	3.225	2.274	I1WEI8-4QJN_A
4	O68451	1IHF_B	3.187	2.029	O68451-1IHF_B
5	P0A1R6	4YEY_A	10.811	0.863	P0A1R6-4YEY_A
6	P0A1R7	4YEY_A	10.811	0.863	P0A1R7-4YEY_A
7	P0A1R8	2O97_B	3.377	0.508	P0A1R8-2O97_B
8	P0A1R9	2O97_B	3.377	0.508	P0A1R9-2O97_B
9	P0A3H1	1HUE_A	4.790	1.127	P0A3H1-1HUE_A
10	P0A3H2	1HUE_A	4.790	1.127	P0A3H2-1HUE_A
11	P0A3H3	4QJN_A	2.744	1.210	P0A3H3-4QJN_A
12	P0A3H4	4QJN_A	2.762	0.898	P0A3H4-4QJN_A
13	P0A3H5	3RHI_A	5.685	4.267	P0A3H5-3RHI_A
14	P0A3H6	3RHI_A	6.409	4.416	P0A3H6-3RHI_A
15	P0ACF1	1MUL_A	4.211	0.791	P0ACF1-1MUL_A
16	P0ACF2	1MUL_A	4.211	0.791	P0ACF2-1MUL_A
17	P0ACF3	4YEY_A	9.896	1.720	P0ACF3-4YEY_A
18	P0ACF5	2O97_B	3.541	1.269	P0ACF5-2O97_B
19	P0ACF6	2O97_B	3.541	1.269	P0ACF6-2O97_B
20	P0ACF7	2O97_B	3.541	1.267	P0ACF7-2O97_B
21	P0C0H2	5FBM_A	2.644	0.763	P0C0H2-5FBM_A
22	P0C0H3	5FBM_A	2.644	0.763	P0C0H3-5FBM_A
23	P0CAV2	4YEW_A	7.939	1.973	P0CAV2-4YEW_A
24	P0DB64	5FBM_A	2.644	0.763	P0DB64-5FBM_A
25	P0DB65	5FBM_A	2.644	0.762	P0DB65-5FBM_A
26	P0DMK4	4QJN_A	3.225	2.274	P0DMK4-4QJN_A
27	P64386	1HUE_A	8.400	1.479	P64386-1HUE_A
28	P64387	1HUE_A	8.451	2.334	P64387-1HUE_A
29	P64388	2O97_B	3.465	1.348	P64388-2O97_B
30	P64389	2O97_B	4.156	0.522	P64389-2O97_B
31	P68573	1HUE_A	5.797	0.962	P68573-1HUE_A
32	P68574	1HUE_A	5.797	0.962	P68574-1HUE_A
33	Q5HUP6	1HUE_A	5.566	1.048	Q5HUP6-1HUE_A
34	Q9JR30	2O97_B	5.736	1.492	Q9JR30-2O97_B
35	Q9KHS6	2O97_B	2.981	0.822	Q9KHS6-2O97_B
36	Q9Z8C7	1HUE_A	8.372	1.628	Q9Z8C7-1HUE_A
37	Q46121	1HUE_A	5.566	1.048	Q46121-1HUE_A

4.6 Ενεργειακή ελαχιστοποίηση των HU μοντέλων μέσω του Chimera.

Το επόμενο βήμα για να ελεγχθεί ένα νεοσύστατο μοντέλο, είναι να ελαχιστοποιηθεί ενεργειακά. Αυτό επιτυγχάνεται μέσω της Μοριακής Μηχανικής/Δυναμικής για την ελαχιστοποίηση της ενέργειας του συστήματος (μήκη/γωνίες δεσμών, στερεοχημικές αλληλεπιδράσεις, δεσμοί-H, κ.λ.π.). Η ελαχιστοποίηση της ενέργειας είναι μια προσπάθεια η οποία γίνεται προκειμένου να μειωθεί (ή αν είναι εφικτό να εκμηδενιστεί) η τυπική απόκλιση μεταξύ των θέσεων των ατόμων της κύριας αλυσίδας ως συνάρτηση της αμινοξικής ταυτότητας του νεοσύστατου μοντέλου.

Ένα αξιόπιστο λογισμικό που μας προσφέρει μια τέτοια δυνατότητα βρίσκεται στην ιστοσελίδα (<https://www.cgl.ucsf.edu/chimera>) εκτελείται τοπικά στον υπολογιστή και ονομάζεται Chimera [66]. Το Chimera δεν είναι μόνο ένα ισχυρό πρόγραμμα απεικόνισης μορίων γραμμένο σε γλώσσα προγραμματισμού Python. Είναι εμπλουτισμένο με αρκετές διεπαφές συμπεριλαμβανομένου και της δυνατότητας να υπάρχει ανάλυση Μοριακής Δυναμικής αρά και ενεργειακής ελαχιστοποίησης μοντέλων που έχουν κατασκευαστεί *in silico*.

Συνεπώς εισάγουμε κάθε νεοσύστατο μοντέλο στο Chimera και εκτελούμε ενεργειακή ελαχιστοποίηση. Στην συνέχεια το εν λόγω μοντέλο επαναξιολογείται μέσω του WhatIF. Το σκορ μικρότερο του 1,00 θεωρείται ότι το μοντέλο δομικά είναι αποδεκτό. Για όλα τα νεοσύστατα πρωτεϊνικά μοντέλα έχει εκτελεστεί η προαναφερθείσα διαδικασία και τα αποτελέσματα συνοψίζονται στον Πίνακα 5.

Πίνακας 5 Σύγκριση των HU μοντέλων με τις αντίστοιχες δομές templates της PDB. Στη στήλη BEFORE CHIMERA καταγράφεται ο συντελεστής $\chi\chi\chi$ που προκύπτει από το WhatIF πριν τη δομική βελτιστοποίηση μέσω Chimera και ο συντελεστής $\chi\chi\chi$ μετά τη δομική βελτιστοποίηση μέσω Chimera. Τα τελικά μοντέλα συγκρίθηκαν μέσω υπέρθεση με τη χρήση του PyMol και η στήλη RMS είναι ενδεικτική της ποιότητας των μοντέλων σε σχέση με τη δομή πρότυπο

A/A	UNIPROT ID	PDB ID	MODEL	WhatIF		Validation
				BEFORE CHIMERA	AFTER CHIMERA	RMS
1	B8GX11	4YEW_A	B8GX11-4YEW_A	7.939	0.329	1.973
2	P0CAV2		P0CAV2-4YEW_A	7.939	0.329	1.973
3	P0ACF1	1MUL_A	P0ACF1-1MUL_A	4.211	----	0.791
4	P0ACF2		P0ACF2-1MUL_A	4.211	----	0.791
5	P0A1R6	4YEY_A	P0A1R6-4YEY_A	10.811	1.111	0.863
6	P0A1R7		P0A1R7-4YEY_A	10.811	1.107	0.863
7	EOJ6W8		EOJ6W8-4YEY_A	9.896	0.062	1.720
8	P0ACF3		P0ACF3-4YEY_A	9.896	0.062	1.720
9	P0ACF5	2O97_B	P0ACF5-2O97_B	3.541	0.125	1.269
10	P0ACF6		P0ACF6-2O97_B	3.541	0.125	1.269
11	P0ACF7		P0ACF7-2O97_B	3.541	----	1.267
12	P0A1R8		P0A1R8-2O97_B	3.377	----	0.508
13	P0A1R9		P0A1R9-2O97_B	3.377	----	0.508
14	P64388		P64388-2O97_B	3.465	0.067	1.348
15	P64389		P64389-2O97_B	4.156	0.065	0.522
16	Q9JR30	Q9JR30-2O97_B	5.736	0.174	1.492	
17	Q9KHS6	Q9KHS6-2O97_B	2.981	0.124	0.822	
18	P0DMK4	4QJN_A	P0DMK4-4QJN_A	3.225	----	2.274
19	I1WEI8		I1WEI8-4QJN_A	3.225	----	2.274
20	P0A3H3		P0A3H3-4QJN_A	2.744	0.552	1.210
21	P0A3H4	P0A3H4-4QJN_A	2.762	0.201	0.898	
22	P0A3H5	3RHI_A	P0A3H5-3RHI_A	5.685	----	4.267
23	P0A3H6		P0A3H6-3RHI_A	6.409	0.107	4.416
24	P0DB64		P0DB64-5FBM_A	2.644	0.432	0.763
25	P0DB65	5FBM_A	P0DB65-5FBM_A	2.644	0.431	0.762
26	P0C0H2		P0C0H2-5FBM_A	2.644	0.431	0.763
27	P0C0H3		P0C0H3-5FBM_A	2.644	0.430	0.763
28	P64386	1HUE_A	P64386-1HUE_A	8.400	0.283	1.479
29	P64387		P64387-1HUE_A	8.451	0.271	2.334
30	P0A3H1		P0A3H1-1HUE_A	4.790	0.370	1.127
31	P0A3H2		P0A3H2-1HUE_A	4.790	0.370	1.127
32	P68573		P68573-1HUE_A	5.797	0.274	0.962
33	P68574		P68574-1HUE_A	5.797	0.274	0.962
34	Q5HUP6		Q5HUP6-1HUE_A	5.566	----	1.048
35	Q9Z8C7		Q9Z8C7-1HUE_A	8.372	0.552	1.628
36	Q46121		Q46121-1HUE_A	5.566	----	1.048
37	O68451		1IHF_B	O68451-1IHF_B	3.187	0.755

4.7 Κατασκευή ομοδιμερών & ετεροδιμερών μοντέλων HU.

Δεδομένου ότι η μελέτη μας, αφορά HU πρωτεΐνες, θα πρέπει να ληφθεί υπόψη ότι οι παραπάνω ακολουθίες είναι υπομονάδες κάποιων ομοδιμερών ή ετεροδιμερών. Συνεπώς με βάση τον οργανισμό στο οποίο απαντώνται θα πρέπει να χαρακτηριστούν αν ανήκουν σε κάποιο ομοδιμερές ή ετεροδιμερές. Βασικό κριτήριο, για αυτήν την κατηγοριοποίηση, όπως αναφέρθηκε και στην αντίστοιχη παράγραφο της μεθοδολογίας, είναι αν δύο υπομονάδες (A και B) της HU απαντώνται στον ίδιο οργανισμό. Αν το τελευταίο ισχύει τότε προκύπτει ετεροδιμερές άνω σε άλλη περίπτωση (αν βρεθεί μια υπομονάδα A ή B σε ένα οργανισμό) προκύπτει ομοδιμερές. Με βάση τον παραπάνω συλλογισμό, και με αναζήτηση στην καταχώρηση της βάσης δεδομένων Uniprot ως προς τον οργανισμό αλλά και το είδος υπομονάδας HU, για τις υφιστάμενες πρωτεϊνικές ακολουθίες, προέκυψαν τα παρακάτω αποτελέσματα, τα οποία παρατίθενται συγκεντρωτικά στον παρακάτω πίνακα.

Πίνακας 6 Καταχωρήσεις της Uniprot με τις αντίστοιχους οργανισμούς και την δημιουργία ομοδιμερών ή ετεροδιμερών αντίστοιχα. Παρατηρείστε ότι με υπομονάδες A και B του ίδιου οργανισμού καταλήγουμε σε δημιουργία ετεροδιμερών.

A/A	Uniprot ID	HU Subunit	Organism	Homomodel	Heteromodel
01	BG8X11	DNA-binding HU-alpha	protein <i>Caulobacter crescentus</i> (strain NA1000 / CB15N)	NAI	
02	E0J6W8	DNA-binding HU-alpha	protein <i>Escherichia coli</i> (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B-766 / W)	NAI	
03	I1WEI8	DNA-binding HU-alpha	protein <i>Burkholderia pseudomallei</i> (strain 1026b)	NAI	
04	O68451	DNA-binding HU-alpha	protein <i>Rickettsia rickettsii</i> (strain Sheila Smith)	NAI	
05	P0A1R6	DNA-binding HU-alpha	protein <i>Salmonella typhimurium</i> (strain LT2 / SGSC1412 / ATCC 700720)		NAI
06	P0A1R7	DNA-binding HU-alpha	protein <i>Salmonella typhi</i>		NAI
07	P0A1R8	DNA-binding HU-beta	protein <i>Salmonella typhimurium</i> (strain LT2 / SGSC1412 / ATCC 700720)		NAI
08	P0A1R9	DNA-binding HU-beta	protein <i>Salmonella typhi</i>		NAI
09	P0A3H1	DNA-binding HU-alpha	protein <i>Bacillus caldolyticus</i>	NAI	
10	P0A3H2	DNA-binding HU-alpha	protein <i>Bacillus caldotenax</i>	NAI	
11	P0A3H3	DNA-binding HRL18	protein <i>Rhizobium leguminosarum</i>	NAI	
12	P0A3H4	DNA-binding HRL18	protein <i>Rhizobium radiobacter</i> (<i>Agrobacterium tumefaciens</i>) (<i>Agrobacterium radiobacter</i>)	NAI	
13	P0A3H5	DNA-binding	protein <i>Streptomyces coelicolor</i> (strain ATCC	NAI	

14	P0A3H6	HU 1 DNA-binding	protein	<i>BAA-471 / A3(2) / M145</i> <i>Streptomyces lividans</i>	NAI
15	P0ACF1	HU 1 DNA-binding	protein	<i>Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC)</i>	NAI
16	P0ACF2	HU-alpha DNA-binding	protein	<i>Escherichia coli O157:H7</i>	NAI
17	P0ACF3	HU-alpha DNA-binding	protein	<i>Shigella flexneri</i>	NAI
18	P0ACF5	HU-alpha DNA-binding	protein	<i>Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC)</i>	NAI
19	P0ACF6	HU-beta DNA-binding	protein	<i>Escherichia coli O157:H7</i>	NAI
20	P0ACF7	HU-beta DNA-binding	protein	<i>Shigella flexneri</i>	NAI
21	P0C0H2	HU-alpha DNA-binding	protein	<i>Streptococcus pyogenes</i>	NAI
22	P0C0H3	HU-alpha DNA-binding	protein	<i>Streptococcus pyogenes serotype M1</i>	NAI
23	P0CAV2	HU-alpha DNA-binding	protein	<i>Caulobacter crescentus (strain ATCC 19089 / CB15)</i>	NAI
24	P0DB64	HU-alpha DNA-binding	protein	<i>Streptococcus pyogenes serotype M3 (strain ATCC BAA-595 / MGAS315)</i>	NAI
25	P0DB65	HU-alpha DNA-binding	protein	<i>Streptococcus pyogenes serotype M3 (strain SSI-1)</i>	NAI
26	P0DMK4	HU-alpha DNA-binding	protein	<i>Burkholderia pseudomallei (strain K96243)</i>	NAI
27	P64386	HU-beta DNA-binding	protein	<i>Chlamydia trachomatis (strain D/UW-3/Cx)</i>	NAI
28	P64387	HU-beta DNA-binding	protein	<i>Chlamydia muridarum (strain MoPn / Nigg)</i>	NAI
29	P64388	HU-beta DNA-binding	protein	<i>Neisseria meningitidis serogroup A / serotype 4A (strain Z2491)</i>	NAI
30	P64389	HU-beta DNA-binding	protein	<i>Neisseria meningitidis serogroup B (strain MC58)</i>	NAI
31	P68573	SPBc2 prophage-derived DNA-binding	protein	<i>Bacillus subtilis (strain 168)</i>	NAI
32	P68574	HU 2 DNA-binding	protein	<i>Bacillus phage SPbeta (Bacillus phage SPBc2) (Bacteriophage SP-beta)</i>	NAI
33	Q5HUP6	HU-alpha DNA-binding	protein	<i>Campylobacter jejuni (strain RM1221)</i>	NAI
34	Q9JR30	HU-beta 2 DNA-binding	protein	<i>Neisseria meningitidis serogroup A / serotype 4A (strain Z2491)</i>	NAI
35	Q9KHS6	HU-beta DNA-binding	protein	<i>Pseudomonas fluorescens (strain ATCC BAA-477 / NRRL B-23932 / Pf-5)</i>	NAI
36	Q9Z8C7	HU-alpha DNA-binding	protein	<i>Chlamydia pneumoniae (Chlamydophila pneumoniae)</i>	NAI
37	Q46121	HU-alpha DNA-binding	protein	<i>Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11168)</i>	NAI

Έτσι καταλήξαμε ότι έκτος από τις πρωτεΐνες P0A1R6-P0A1R8, P0A1R7-P0A1R9, P0ACF1-P0ACF5, P0ACF2-P0ACF6, P0ACF3-P0ACF7 οι οποίες απαρτίζουν υπομονάδα A και B που ανήκουν στον ίδιο οργανισμό και συνεπώς ετεροδιμερίζονται, οι υπόλοιπες θα ομοδιμεριστούν.

Το επόμενο βήμα είναι η κατασκευή των ομο/έτερο διμερών μοντέλων. Για να επιτευχθεί κάτι τέτοιο θα πρέπει, όπως αναφέρθηκε στην αντίστοιχη ενότητα της μεθοδολογίας, πρωτίστως πρέπει να οριστούν οι δομές οδηγών (templates) και ακολούθως να εισαχθούν οι αντίστοιχες ακολουθίες (στόχου και οδηγού) σε υπορουτίνες του γνωστού πλέον Modeller. Κατά την διαδικασία όμως προέκυψε ένα πρόβλημα. Οι δομές οδηγών που βρέθηκαν αρχικά μέσω του BLAST έναντι της PDB, δεν ήταν δυνατό να χρησιμοποιηθούν όλες για την κατασκευή των ομο/έτεροδιμερών, διότι αρκετά templates είναι ομοδιμερή και εμείς θέλουμε να κατασκευάσουμε ετεροδιμερή μοντέλα ή και το αντίστροφο. Η λύση που δόθηκε ήταν να εκτελέσουμε εκ νέου BLAST έναντι της PDB και να επιλέξουμε την πλησιέστερη δομή, για δομή οδηγό, η οποία αφενός θα πληρεί το ελάχιστο δυνατό E-Value και το μέγιστο δυνατό Query Coverage, αλλά και αφετέρου και ως διμερές θα συμφωνεί με το είδος του διμερούς που επιθυμούμε να κατασκευάσουμε.

Πίνακας 7 Καταχωρήσεις της Uniprot με τις αντίστοιχες των δομών-οδηγών(templates) της PDB, και το νέο template που έχει επιλεγεί με σημείο αναφοράς αν πρόκειται για κατασκευή διμερούς

No	Uniprot ID	Template	New Template
01	BG8X11	4YEW_A	1HUE
02	E0J6W8	4YFY_A	4YFH
03	I1WEI8	4QJN_A	1HUE
04	O68451	1IHF_B	2NP2
04	P0A1R6- P0A1R8	1MUL_A-2O97_B	4YFY_A-2O97_B
05	P0A1R7- P0A1R9	4YFY_A-2O97_B	4YFY_A-2O97_B
06	P0A3H1	1HUE_A	1HUE
07	P0A3H2	1HUE_A	1HUE
08	P0A3H3	4QJN_A	1HUE
09	P0A3H4	4QJN_A	1HUE
10	P0A3H5	3RHI_A	1WTU
11	P0A3H6	3RHI_A	1WTU
12	P0ACF1- P0ACF5	1MUL_A-2O97_B	4YFY_A-2O97_B
13	P0ACF2- P0ACF6	1MUL_A-2O97_B	4YFY_A-2O97_B
14	P0ACF3- P0ACF7	4YFY_A-2O97_B	1MUL_A-2O97_B
15	P0C0H2	5FBM_A	1HUE
16	P0C0H3	5FBM_A	1HUE
17	P0CAV2	4YEW_A	1HUE
18	P0DB64	5FBM_A	1HUE
19	P0DB65	5FBM_A	1HUE

20	P0DMK4	4QJN_A	1HUE
21	P64386	1HUE_A	1HUE
22	P64387	1HUE_A	1HUE
23	P64388	2O97_B	4YFY
24	P64389	2O97_B	4YFY
25	P68573	1HUE_A	1HUE
26	P68574	1HUE_A	1HUE
27	Q5HUP6	1HUE_A	1HUE
28	Q9JR30	2O97_B	4YFY
29	Q9KHS6	2O97_B	1HUE
30	Q9Z8C7	1HUE_A	1HUE
31	Q46121	1HUE_A	1HUE

4.8 Αξιολόγηση των ομο-ετερο διμερών μοντέλων HU

Στην συνέχεια, εκτελέστηκε η γνωστή από το κεφαλαίο της μεθοδολογίας διαδικασία, αξιολόγησης των προς κατασκευή ομο/ετερο διμερών μοντέλων. Όπως κάθε νεοσύστατο *in silico* μοντέλο έτσι και τα παραπάνω θα πρέπει να αξιολογηθούν δομικά. Σε αναλογία με την κατασκευή των αρχικών μοντέλων, και τα ομο/ετερο διμερή αξιολογήθηκαν α) Μέσω WhatIF ως προς τους ενδομοριακούς και διαμοριακούς δεσμούς τους, β) Εισήχθησαν στο Chimera για την ενεργειακή τους ελαχιστοποίηση γ) Επαναλήφθηκε η διαδικασία μέσω του WhatIF δ) Μέσω του προγράμματος PISA υπολογίστηκε η τιμή ΔG και η επιφάνεια που είναι εκτεθειμένη στον διαλύτη ε) Υπολογίστηκε το RMS μέσω δομικής υπέρθεσης μέσω του προγράμματος PyMol. Τα αποτελέσματα που προέκυψαν συνοψίζονται στον παρακάτω πίνακα.

Πίνακας 8 Καταχωρήσεις των ομο/ετεροδιμερών μοντέλων με τα αντίστοιχα σκορ του WhatIF πριν και μετά την δομική βελτιστοποίηση μέσω Chimera, το αντίστοιχο ΔG και η επιφάνεια που είναι εκτεθειμένη στον διαλύτη όπως αυτά υπολογίστηκαν μέσω του PISA, το αντίστοιχο RMS όπως αυτό προκύπτει από το PyMol και η ονομασία του αντίστοιχού *in silico* διμερούς που κατασκευάστηκε.

A/A	Homo/Hetero model	WhatIF Validation				
		Before Chimera	After Chimera	ΔG of PISA(-) in kcal/mol	Buried Surface Area Dimer (\AA^2)	RMS
01	B8GX11-1HUE homomodel	16.207	0.464	35.6	1922,6	0.316
02	E0J6W8-4YFH homomodel	4.081	0.470	33.9	1852	0.354
03	I1WEI8-1HUE homomodel	16.550	0.811	35.8	1979,5	0.376
04	O68451-2NP2 homomodel	15.834	1.464	44.5	2136,4	3.517
05	P0A1R6-P0A1R8 with 4YFY_A-2O97_B heteromodel	4.575	0.484	32.2	1911,2	(*)
06	P0A1R7-P0A1R9 with 4YFY_A-2O97_B heteromodel	4.575	0.475	32.2	1910,3	(*)
07	P0A3H1-1HUE homomodel	9.790	0.885	34.8	1795,8	0.375
08	P0A3H2-1HUE homomodel	9.790	0.884	34.8	1795,6	0.375
09	P0A3H3-1HUE homomodel	15.345	0.800	30.3	1986,3	0.361
10	P0A3H4-1HUE homomodel	15.345	0.800	30.3	1985,6	0.361
11	P0A3H5-1WTU homomodel	60.304	3.182	38.4	2649,8	4.215
12	P0A3H6-1WTU homomodel	26.184	1.789	34.5	2220,5	0.527
13	P0ACF1-P0ACF5with 4YFY_A-2O97_B heteromodel	5.405	0.590	33.5	1908,9	(*)
14	P0ACF2- P0ACF6 with4YFY_A-2O97_Bheteromodel	6.292	1.124	34.0	1908,3	(*)
15	P0ACF3- P0ACF7 With1MUL_A-2O97_Bheteromodel	9.897	0.267	0.9	167,1	(*)
16	P0C0H2-1HUE homomodel	14.018	0.906	37.3	1844,1	0.362
17	P0C0H3-1HUE homomodel	14.018	0.931	37.3	1845,3	0.361
18	P0CAV2-1HUE homomodel	16.207	0.463	35.6	1922	0.324
19	P0DB64-1HUE homomodel	14.018	0.929	37.3	1845,3	0.363
20	P0DB65-1HUE homomodel	14.018	0.932	37.3	1845,1	0.363
21	P0DMK4-1HUE homomodel	16.550	0.799	35.8	1979,4	0.375
22	P64386-1HUE homomodel	28.128	3.384	39.0	2267,1	0.429

23	P64387-1HUE homomodel	28.128	3.389	39.0	2266,6	0.429
24	P64388-4YEY homomodel	6.497	0.678	37.1	1891,3	0.190
25	P64389-4YEY homomodel	6.497	0.678	37.1	1891,4	0.190
26	P68573-1HUE homomodel	20.757	1.277	33.5	2022,5	0.424
27	P68574-1HUE homomodel	20.757	1.277	33.5	2022,5	0.424
28	Q5HUP6-1HUE homomodel	21.134	1.001	35.2	1962,3	0.395
29	Q9JR30-4YFH homomodel	11.176	0.981	40.8	1952,4	0.362
30	Q9KHS6-1HUE homomodel	15.059	0.564	37.3	1754,7	0.362
31	Q9Z8C7-1HUE homomodel	20.302	1.886	40.1	2315,3	0.456
32	Q46121-1HUE homomodel	21.134	1.001	35.2	1962,2	0.395

(*) *Μη δυνατή μέτρηση RMS μιας και πρόκειται για ετεροδιμέρη.*

Οι καταχωρήσεις με το όνομα P0A3H5 και O68451 δεν περιλαμβάνονται στην συνέχεια της μελέτης αφού οι τιμές του RMS σε συνδυασμό με το σκορ του WhatIF μετά την ενεργειακή ελαχιστοποίηση που προκύπτουν, είναι πολύ πάνω από το κατώφλι, και ως εκ τούτου κρίνονται απαράδεκτα.

4.9 Μελέτη της επίδρασης της αλληλουχίας διμερισμού στη σταθερότητα των HU

Αφού ολοκληρώθηκε η κατασκευή των *in silico* των 32 μοντέλων (27 μονομερή και 5 ετεροδιμερή) και βελτιστοποιήθηκε η δομή τους, με την διαδικασία που αναφέρθηκε στην αντίστοιχη παράγραφο της ενότητας της μεθοδολογίας, στην συνέχεια θα μελετήσουμε την αλληλουχία διμερισμού προέκυψαν (dimerizations signals) όχι μόνο των πρωτεϊνών που κατασκευάστηκαν, αλλά και των αντίστοιχων δομών-οδηγών. Για την καλύτερη οπτικοποίηση των αποτελεσμάτων παρατίθενται οι παρακάτω πίνακες.

Ο πρώτος πίνακας απεικονίζει την δημιουργία αρχείων δομών DS με κατάληξη .pbd και ομαδοποίηση των νεοσύστατων DS με το αντίστοιχο DS της κοινής δομής-οδηγού (template).

Πίνακας 9 Κατανομή DS των νεοσύστατων μοντέλων ανά DS των δομών-οδηγών.

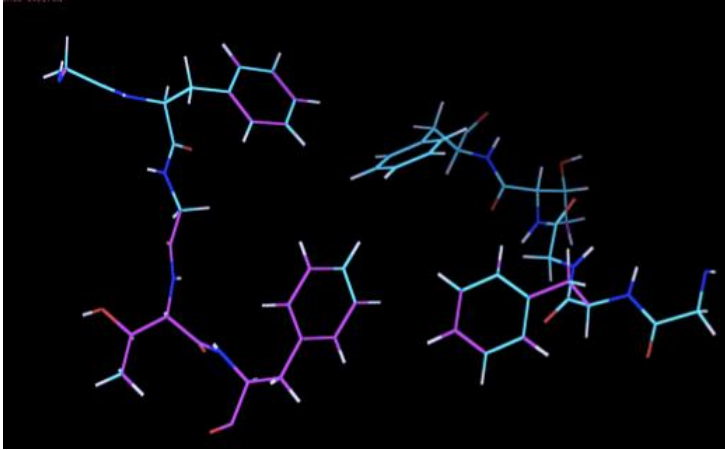
DS of TEMPLATES	4YEW	4YEY	5FBM	1MUL	2O97	1HUE	3RHI	1IHF
DS of MODELS	BG8X11	E0J6W8	P0C0H2	P0A1R6- P0A1R8	P0ACF2- P0ACF6 P0ACF3- P0ACF7 P64388	P0A3H1 P0A3H2 P68573	P0A3H6	P64386 P64387 Q9Z8C7
	P0CAV2	P0A1R7- P0A1R9 P0ACF1- P0ACF5 P0DB65	P0C0H3 P0DB64		P64389 Q9JR30 Q9KHS6	P68574 Q5HUP6 Q46121 I1WEI8 P0A3H3 P0A3H4 P0DMK4		

Το επόμενο βήμα είναι η απεικόνιση μέσω PyMol της κάθε ομάδας πλέον (template με αντίστοιχα μοντέλα). Ακολουθεί υπέρθεση δομών και οπτική παρατήρηση των δακτυλίων της πλευρικής αλυσίδας των Phe του κάθε μοντέλου με τους αντίστοιχους δακτυλίους του template. Από αυτή την παρατήρηση προκύπτει ότι μοντέλα που έχουν την ίδια τιμή ΔG , οι δομές τους ταυτίζονται. Το παραπάνω αποτέλεσμα μας

επιβεβαιώνει τις αρχικές μας ενδείξεις ότι το DS διαδραματίζει καθοριστικό ρόλο στην διαμόρφωση της τιμής ΔG και κατά συνέπεια στην δημιουργία ή μη ομο/ετερο διμερούς. Στην συνέχεια ακολουθεί υπέρθεση των DS των templates με τα αντίστοιχα DS των θεωρητικών μοντέλων, μέσω του PyMol. Κάτω από κάθε απεικόνιση συνοδεύεται από πίνακα στον οποίο αναγράφεται το ΔG του αντίστοιχου μοντέλου. Με κόκκινο χρώμα χρωματίζονται ίδιες τιμές ΔG ώστε να τονιστεί η δομική ταύτιση που παρατηρείται.

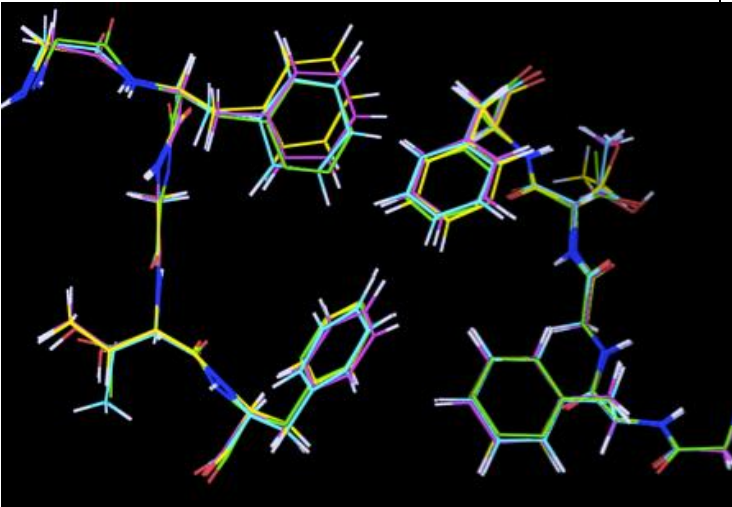
4.10 Συγκριτική ανάλυση των διαφόρων DS και συσχέτιση σταθερότητας

4YEW	
<p>Πίνακας 10. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.</p>	
DS Πρωτεΐνης	ΔG (kcal/mol)
BG8X11	-35.6
P0CAV2	-35.6



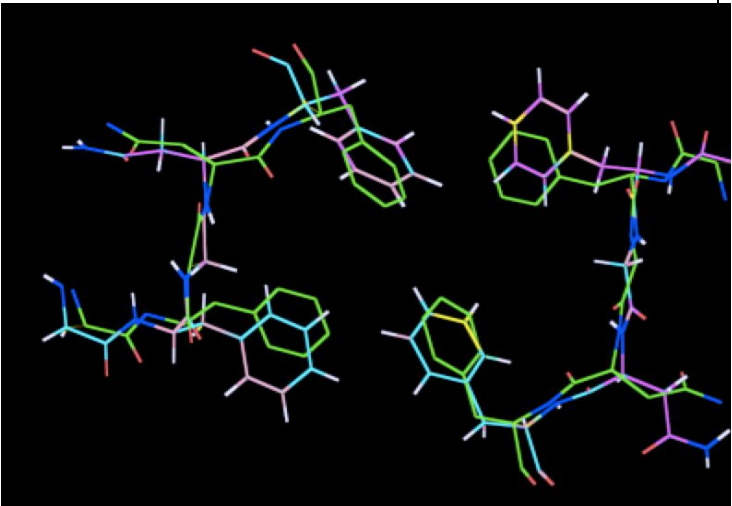
Εικόνα 35 Απεικόνιση του DS της δομής-οδηγός 4YEW και του μοντέλου BG8X11 και P0CAV2 σε υπέρθεση (Παρατηρείται ταύτιση δομών η οποία αναφέρεται σε κοινή τιμή της ελεύθερης ενέργειας ΔG)

4YEY	
<p>Πίνακας 11. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.</p>	
DS Πρωτεΐνης	ΔG (kcal/mol)
E0J6W8	-33.9
P0A1R7- P0A1R9	-32.2
P0ACF1- P0ACF5	-33.5



Εικόνα 36 Απεικόνιση του DS της δομής - οδηγός 4YEY και των αντίστοιχων μοντέλων με κοινή δομή - οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

5FBM	
<p>Πίνακας 12. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.</p>	
DS Πρωτεΐνης	ΔG (kcal/mol)
P0C0H2	-37.3
P0C0H3	-37.3
P0DB64	-37.3
P0DB65	-37.3

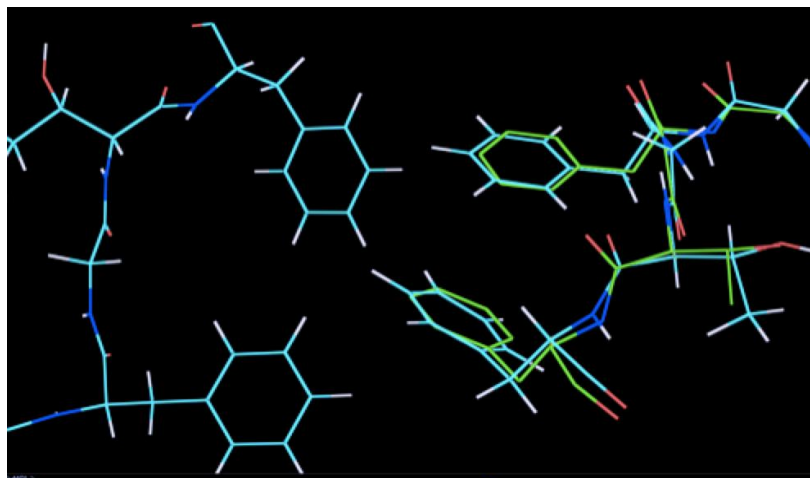


Εικόνα 37 Απεικόνιση του DS της δομής-οδηγός 5FBM και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

1MUL

Πίνακας 13. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

DS Πρωτεΐνης	ΔG (kcal/mol)
P0A1R6- P0A1R8	-32.2

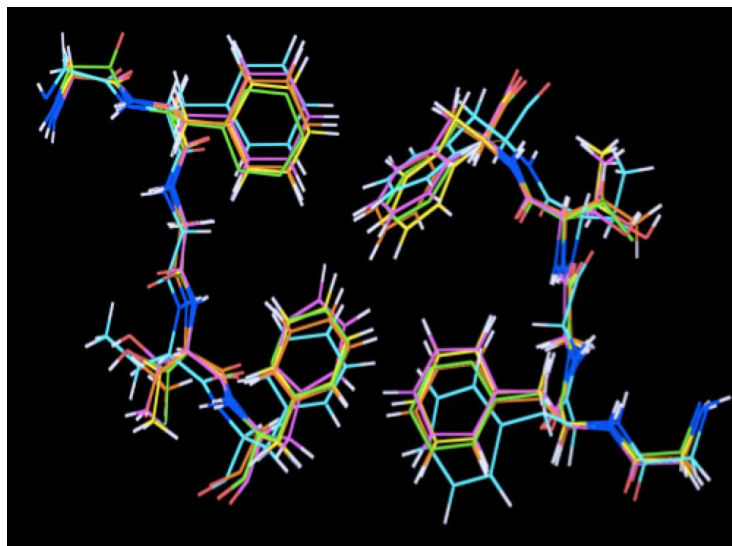


Εικόνα 38 Στιγμιότυπο απεικόνισης του DS της δομής-οδηγός 1MUL και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

2O97

Πίνακας 14. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

DS Πρωτεΐνης	ΔG (kcal/mol)
P0ACF2- P0ACF6	-34.0
P0ACF3- P0ACF7	0.9
P64388	-37.1
P64389	-37.1
Q9JR30	-40.8
Q9KHS6	-37.3

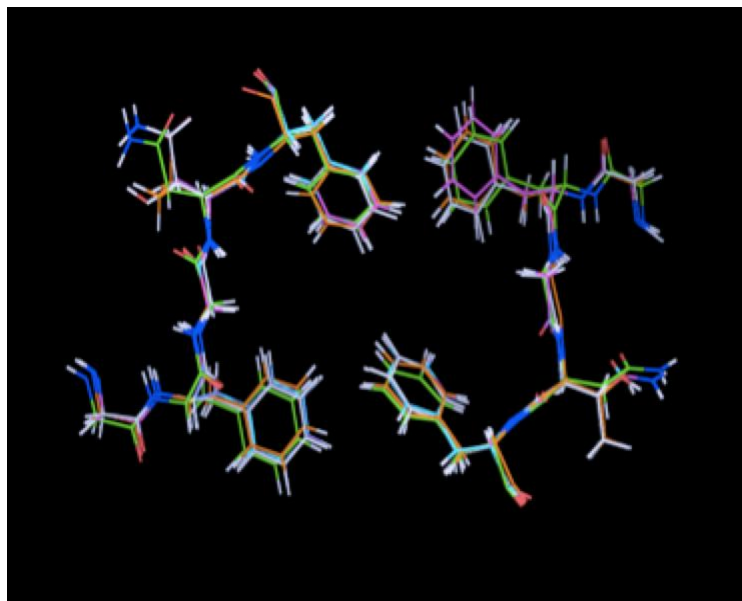


Εικόνα 39 Στιγμιότυπο απεικόνισης του DS της δομής-οδηγός 2O97 και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

1HUE A PART

Πίνακας 15. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

P0A3H1	-34.8
P0A3H2	-37.3
P68573	-33.5
P68574	-33.5
Q5HUP6	-35.2

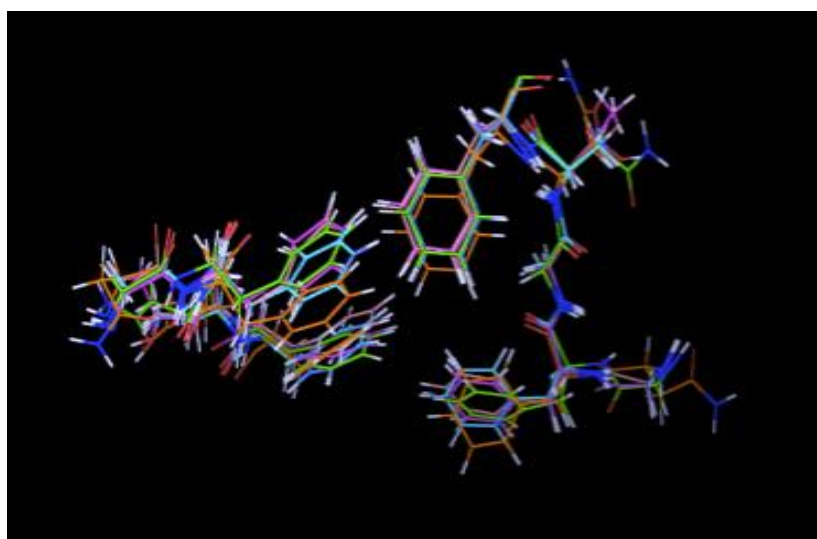


Εικόνα 40 Απεικόνιση του DS της δομής-οδηγός 1HUE και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

1HUE B PART

Πίνακας 16. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

Q46121	-35.2
I1WEI8	-35.8
P0A3H3	-30.3
P0A3H4	-30.3
P0DMK4	-35.8

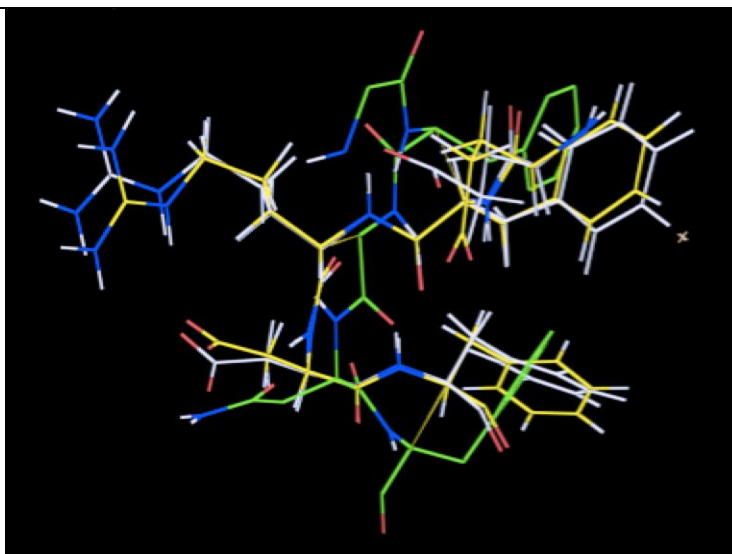


Εικόνα 41. Απεικόνιση του DS της δομής-οδηγός 1HUE και των αντίστοιχων μοντέλων με κοινή δομή - οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

1IHF

Πίνακας 17. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

DS Πρωτεΐνης	ΔG (kcal/mol)
P64386	-39.0
P64387	-39.0
Q9Z8C7	-40.1

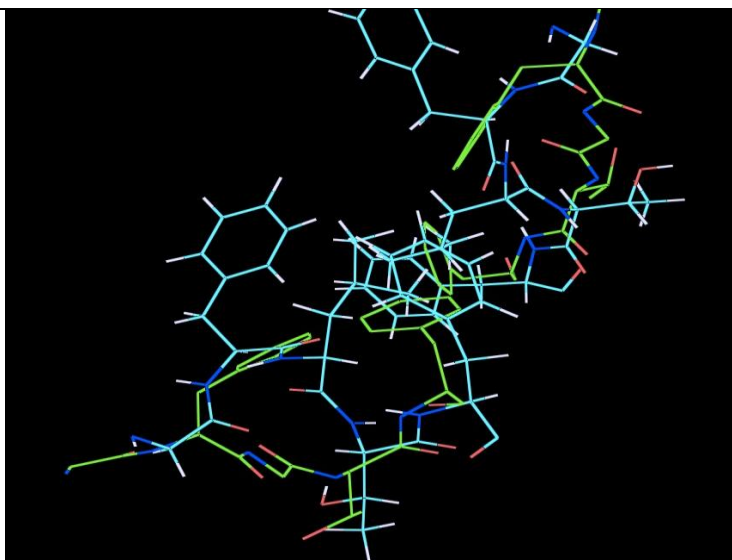


Εικόνα 42 Απεικόνιση του DS της δομής-οδηγός 1IHF και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

3RHI

Πίνακας 18. Συσχέτιση DS των νεοσύστατων μοντέλων με τα αντίστοιχα ΔG τους, όπως εκείνα έχουν προκύψει από το PISA.

DS Πρωτεΐνης	ΔG (kcal/mol)
P0A3H6	-34.5



Εικόνα 43 Απεικόνιση του DS της δομής-οδηγός 1IHF και των αντίστοιχων μοντέλων με κοινή δομή-οδηγό σε υπέρθεση. (Παρατηρείται ταύτιση δομών σε όλα τα θεωρητικά μοντέλα στα οποία αποδίδεται κοινή τιμή της ελεύθερης ενέργειας ΔG)

4.11 Συγκριτική ανάλυση των ασθενών αλληλεπιδράσεων στις HU

Όπως αναφέρθηκε στο σχετικό κεφάλαιο της μεθοδολογίας μέσω του Protein Interactions Calculator (P.I.C.) αναλύθηκαν οι ασθενείς αλληλεπιδράσεις όχι μόνο των νεοσύστατων μοντέλων αλλά και των λυμένων κρυσταλλογραφικά δομών που χρησιμοποιήθηκαν ως δομές - οδηγοί. Εκτελέστηκαν και αξιολογήθηκαν και καταμετρήθηκαν οι υδρογονικοί και οι ιοντικοί δεσμοί, οι αλληλεπιδράσεις αρωματικών δακτυλίων τόσο ανάμεσα τους , όσο και με άτομα θείου. Επίσης εντοπίστηκαν τα υδρόφοβά τμήματα. Όλα τα παραπάνω πραγματοποιήθηκαν τόσο σε ενδομοριακό όσο και διαμοριακό επίπεδο. Οι ακόλουθες μετρήσεις εκτελέστηκαν τόσο στα νεοσύστατα μοντέλα όσο και στις κρυσταλλογραφικά λυμένες δομές που χρησιμοποιήθηκαν ως δομές - οδηγοί (templates). Παρακάτω παρατίθεται συγκριτικός πίνακας μεταξύ των δομών - οδηγών.

Comparison of the molecular interactions and geometry between the templates of HU	1HUE	1WTU	2NP2	4DKY	4QJN	4YEY	4YFH
Intra - subunit interactions							
Intra - subunit hydrophobic interactions within 5A	65	60	50	50	54	34	42
Intra - subunit main chain- main chain H-bonds	109	133	143	130	254	110	114
Intra - subunit main chain- side chain H-bonds	25	17	35	41	68	25	24
Intra - subunit side chain- side chain H-bonds	14	2	14	20	31	30	20
Intra - subunit ionic interactions within 6A	10	2	18	5	9	5	7
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	2	2	4	3	7	4	3
Intra - subunit cation-pi interactions	0	0	2	0	1	0	0
Intra - subunit aromatic -sulphur interactions	0	0	0	0	0	0	0
Inter - subunit interactions							
Inter - monomers hydrophobic interactions within 5A	56	48	60	0	48	44	44
Inter - subunit main chain- main chain H-bonds	10	3	8	0	11	5	3
Inter - subunit main chain- side chain H-bonds	1	15	5	0	4	4	2
Inter - subunit side chain- side chain H-bonds	1	4	17	0	7	0	2
Inter - monomers ionic interactions	0	7	5	0	2	1	1
Inter - monomers aromatic interactions	4	0	6	0	6	3	3
Inter - monomers cation-pi interactions	0	0	1	0	0	0	0
Inter - monomers aromatic -sulphur interactions	0	0	0	0	0	0	0

4.12 Συσχετισμός της θερμοκρασίας ανάπτυξης των βακτηρίων με τις ασθενείς αλληλεπιδράσεις των ΗΥ

Στην εισαγωγή της παρούσας διπλωματικής εργασίας, αναφέρθηκε το GOT (Growth Organism Temperature) και το T_m των οργανισμών, στην συνέχεια βρέθηκαν πρωτεϊνικές ακολουθίες και κατασκευαστήκαν θεωρητικά πρωτεϊνικά μοντέλα. Πόσο όμως συσχετίζονται οι τιμές της θερμοκρασίας του T_m των θεωρητικών πρωτεϊνών με τις αντίστοιχες τιμές της θερμοκρασίας GOT των αντίστοιχων οργανισμών που έχουν μετρηθεί και υπολογιστεί πειραματικά; Ποια η σχέση του βαθμού υδροφοβικότητας μιας πρωτεΐνης με το αντίστοιχο GOT του οργανισμού που έχει μελετηθεί;

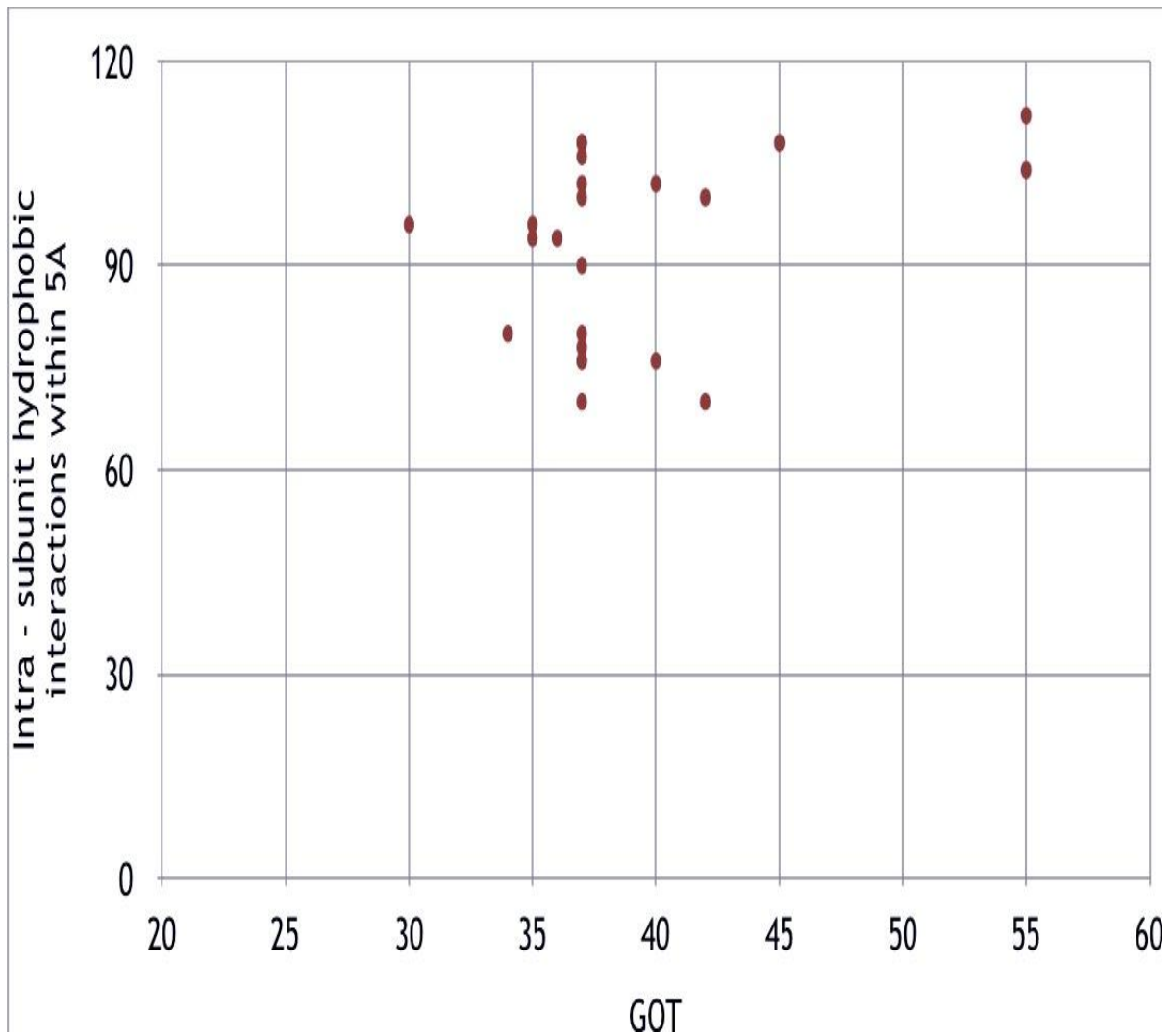
Για να απαντήσουμε στα παραπάνω ερωτήματα, θα πρέπει να συγκεντρώσουμε τα δεδομένα που υπάρχουν σε πίνακα και στην συνέχεια να αποπειραθούμε μέσω γραφημάτων να καταλήξουμε σε ένα συμπέρασμα. Ως εκ τούτου παρακάτω παρατίθενται το AccessionID της Uniprot των πρωτεϊνών, ο αντίστοιχος οργανισμός στο οποίο έχει μελετηθεί, με την αντίστοιχη τιμή GOT .

Πίνακας 19. Καταχωρήσεις των θεωρητικών μοντέλων με τους οργανισμούς στους οποίους απαντώνται κατά Uniprot, σε σχέση με το GOT των τελευταίων.

A/A	ΘΕΩΡΗΤΙΚΟ ΜΟΝΤΕΛΟ	ΟΡΓΑΝΙΣΜΟΣ	GOT ΟΡΓΑΝΙΣΜΟΥ (°C)
1	BG8X11	<i>Caulobacter crescentus</i> (strain NA1000 / CB15N)	28
2	E0J6W8	<i>Escherichia coli</i> (strain ATCC 9637 / CCM 2024 / DSM 1116 / NCIMB 8666 / NRRL B-766 / W)	37
3	I1WEI8	<i>Burkholderia pseudomallei</i> (strain 1026b)	40
4	O68451	<i>Rickettsia rickettsii</i> (strain Sheila Smith)	34
5	P0A1R6-P0A1R8	<i>Salmonella typhimurium</i> (strain LT2 / SGSC1412 / ATCC 700720)	37
6	P0A1R7-P0A1R9	<i>Salmonella typhi</i>	37
7	P0A3H1	<i>Bacillus caldolyticus</i>	55
8	P0A3H2	<i>Bacillus caldotenax</i>	55
9	P0A3H3	<i>Rhizobium leguminosarum</i>	26
10	P0A3H4	<i>Rhizobium radiobacter</i> (<i>Agrobacterium tumefaciens</i>) (<i>Agrobacterium radiobacter</i>)	30
11	P0A3H5	<i>Streptomyces coelicolor</i> (strain ATCC BAA-471 / A3(2) / M145)	29
12	P0A3H6	<i>Streptomyces lividans</i>	29
13	P0ACF1-P0ACF5	<i>Escherichia coli</i> O6:H1 (strain CFT073 / ATCC 700928 / UPEC)	37
14	P0ACF2-P0ACF6	<i>Escherichia coli</i> O157:H7	45
15	P0ACF3-P0ACF7	<i>Shigella flexneri</i>	37

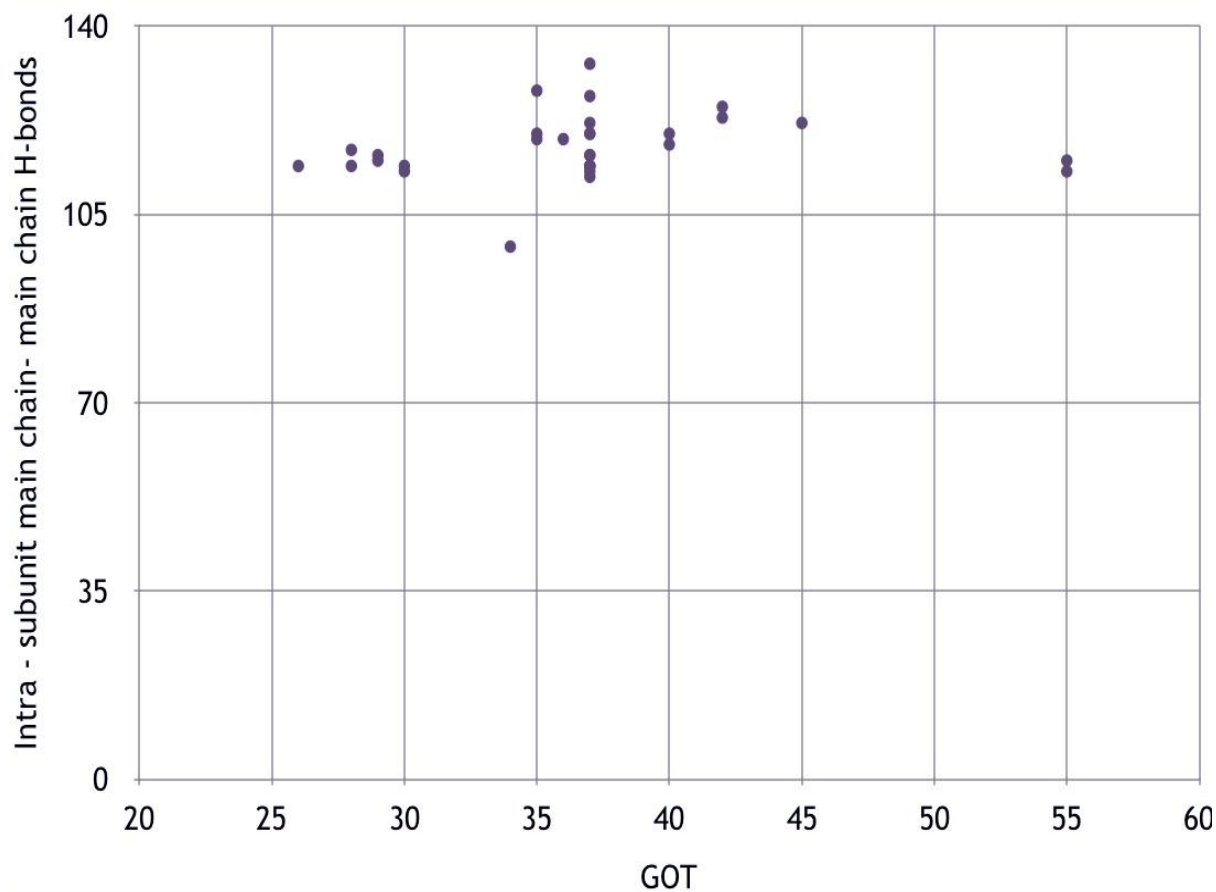
16	P0C0H2	<i>Streptococcus pyogenes</i>	37
17	P0C0H3	<i>Streptococcus pyogenes</i> serotype M1	37
18	P0CAV2	<i>Caulobacter crescentus</i> (strain ATCC 19089 / CB15)	30
19	P0DB64	<i>Streptococcus pyogenes</i> serotype M3 (strain ATCC BAA-595 / MGAS315)	37
20	P0DB65	<i>Streptococcus pyogenes</i> serotype M3 (strain SSI-1)	37
21	P0DMK4	<i>Burkholderia pseudomallei</i> (strain K96243)	40
22	P64386	<i>Chlamydia trachomatis</i> (strain D/UW-3/Cx)	37
23	P64387	<i>Chlamydia muridarum</i> (strain MoPn / Nigg)	36
24	P64388	<i>Neisseria meningitidis</i> serogroup A / serotype 4A (strain Z2491)	35
25	P64389	<i>Neisseria meningitidis</i> serogroup B (strain MC58)	35
26	P68573	<i>Bacillus subtilis</i> (strain 168)	30
27	P68574	<i>Bacillus phage SPbeta</i> (<i>Bacillus phage SPBc2</i>) (<i>Bacteriophage SP-beta</i>)	37
28	Q5HUP6	<i>Campylobacter jejuni</i> (strain RM1221)	42
29	Q9JR30	<i>Neisseria meningitidis</i> serogroup A / serotype 4A (strain Z2491)	35
30	Q9KHS6	<i>Pseudomonas fluorescens</i> (strain ATCC BAA-477 / NRRL B-23932 / Pf-5)	28
31	Q9Z8C7	<i>Chlamydia pneumoniae</i> (<i>Chlamydophila pneumoniae</i>)	37
32	Q46121	<i>Campylobacter jejuni</i> subsp. <i>jejuni</i> serotype O:2 (strain ATCC 700819 / NCTC 11168)	42

Ακολουθώντας, συνδυάζοντας τα παραπάνω δεδομένα μέσω του P.I.C των μοντέλων και των δομών - οδηγών με την τιμή GOT του κάθε οργανισμού, μπορούμε να έχουμε ένα γράφημα διασποράς για κάθε μία μέτρηση. Στο άξονα του X εμπεριέχεται η πληροφορία της θερμοκρασίας για την βέλτιστη ανάπτυξη του οργανισμού και φυσικά μετριέται σε βαθμούς Κελσίου (°C), ενώ τον άξονα Y τα αποτελέσματα της μέτρησης της κάθε διαμοριακής ή ενδομοριακής αλληλεπίδρασης .



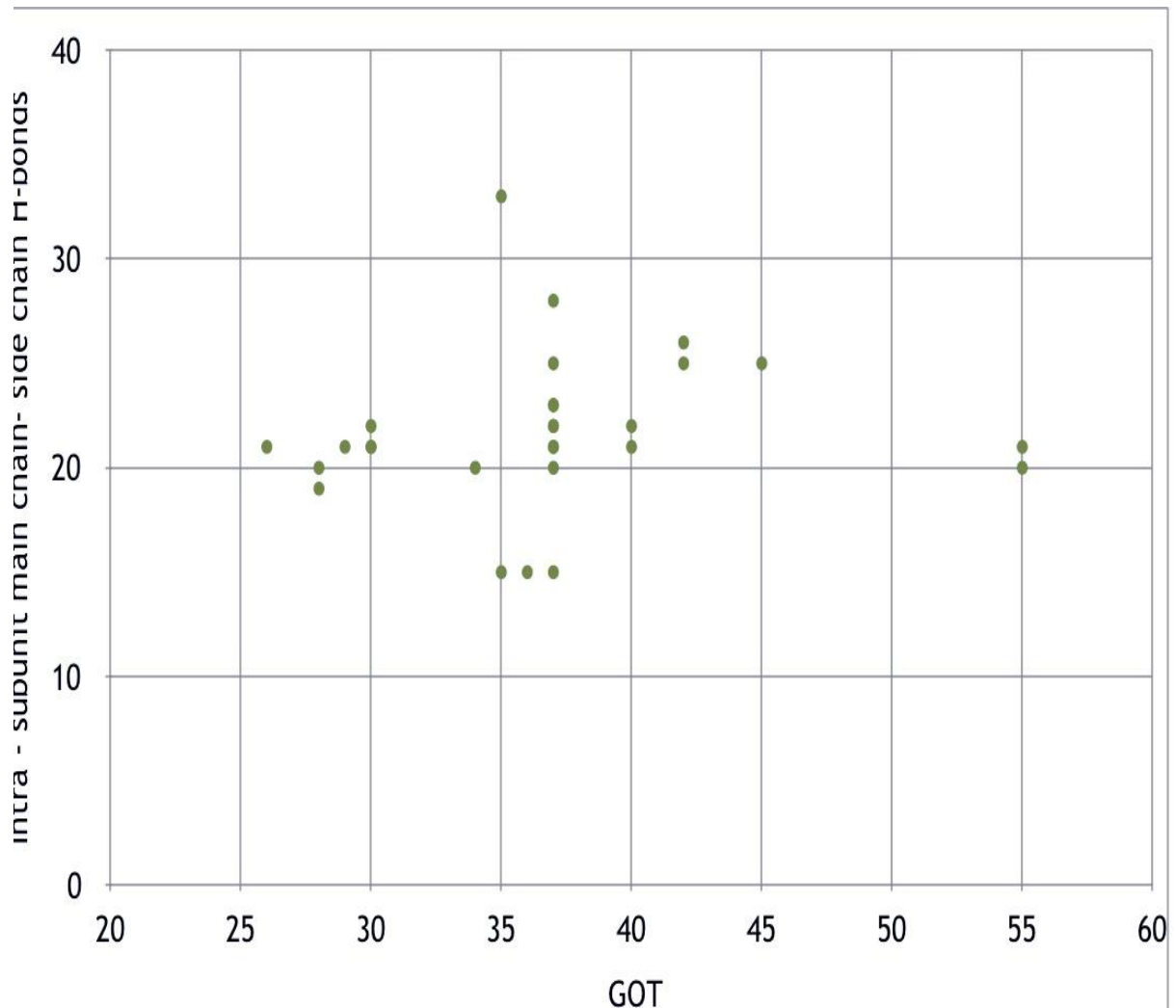
Εικόνα 44 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των διαμοριακών υδροφοβικών αλληλεπιδράσεων, οι οποίες απεικονίζονται στον άξονα των Y, στην απόσταση των 5Å.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των διαμοριακών υδροφοβικών αλληλεπιδράσεων, σε απόσταση 5Å και της θερμοκρασίας ανάπτυξης των οργανισμών.



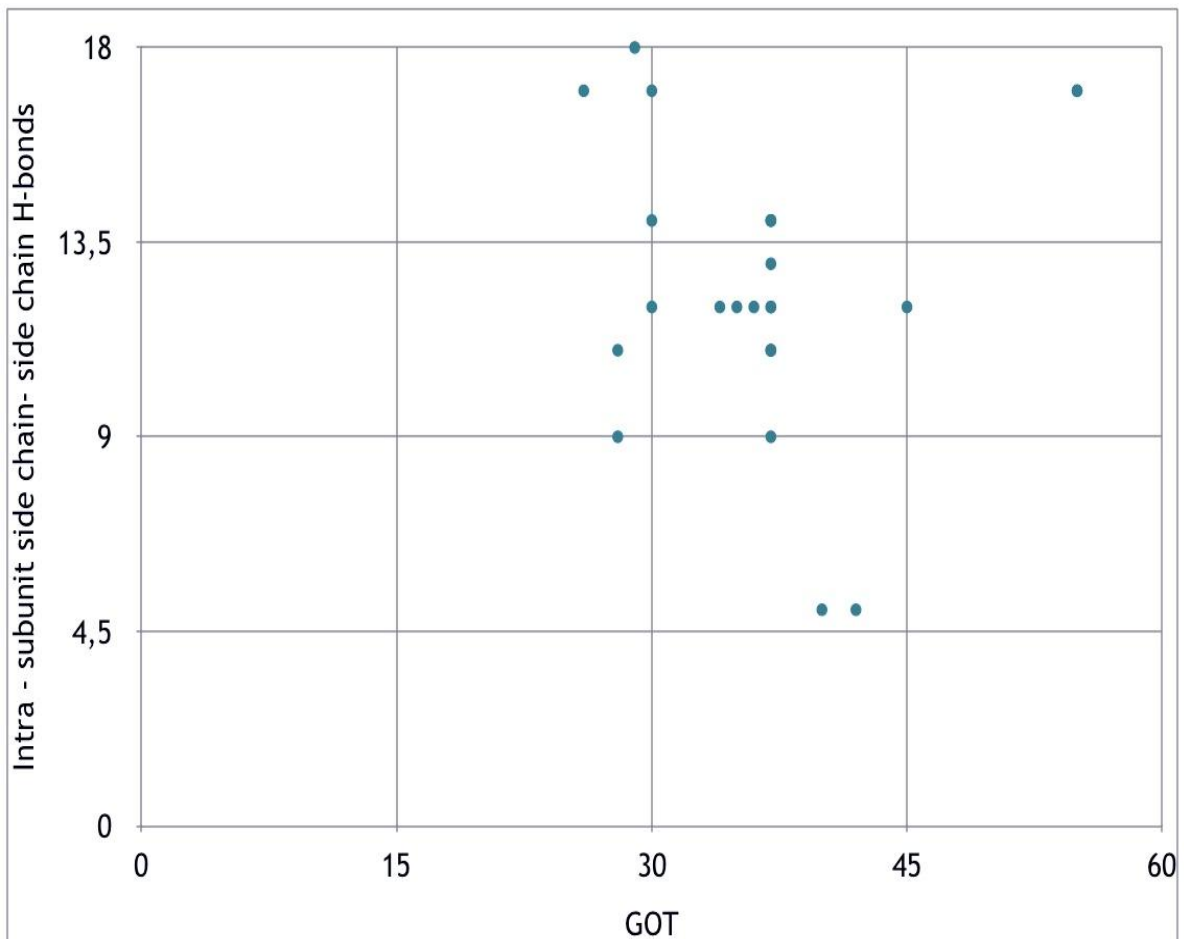
Εικόνα 45 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριων αλυσίδων, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριων αλυσίδων και της θερμοκρασίας ανάπτυξης των οργανισμών.



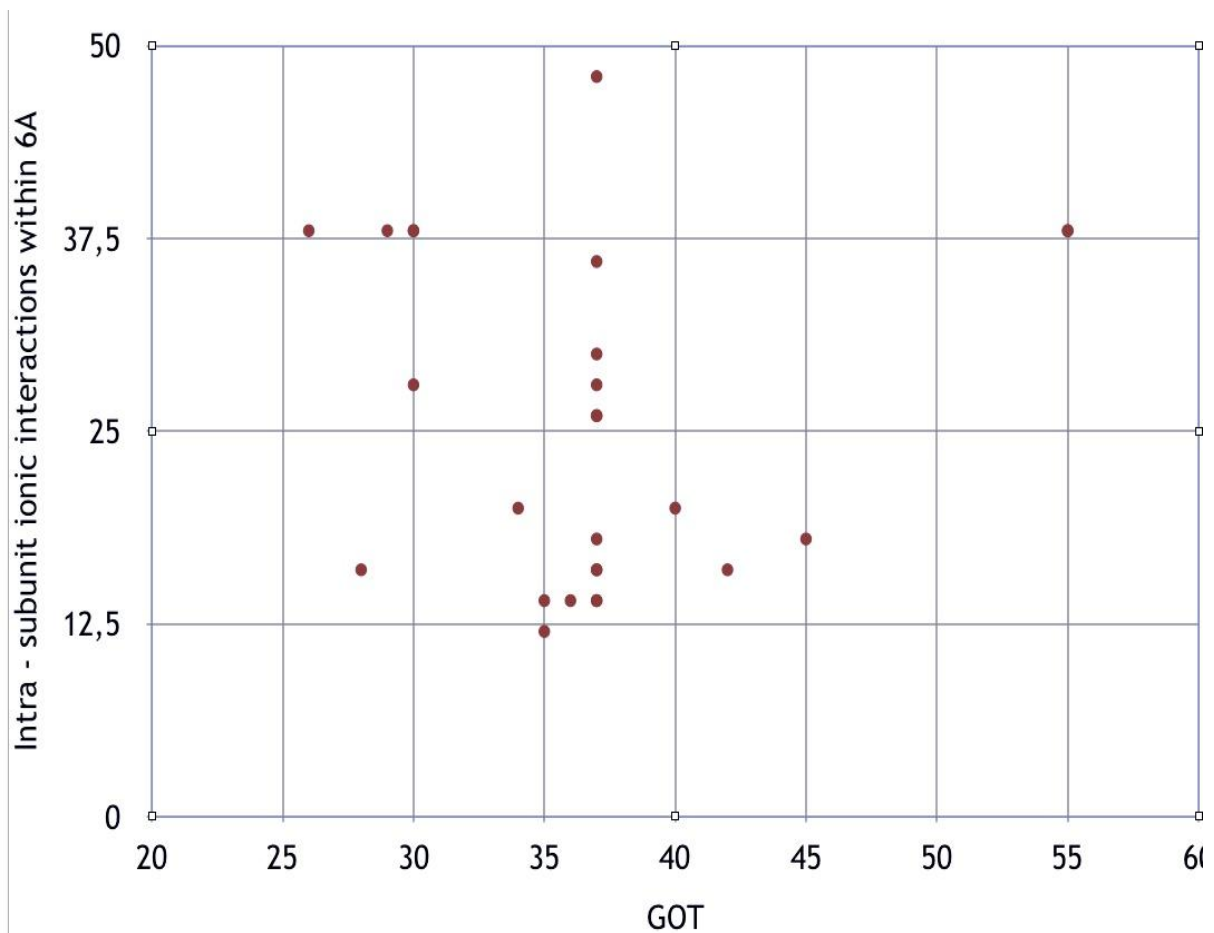
Εικόνα 46 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριας και πλευρικής αλυσίδας, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριας και πλευρικής αλυσίδας και της θερμοκρασίας ανάπτυξης των οργανισμών.



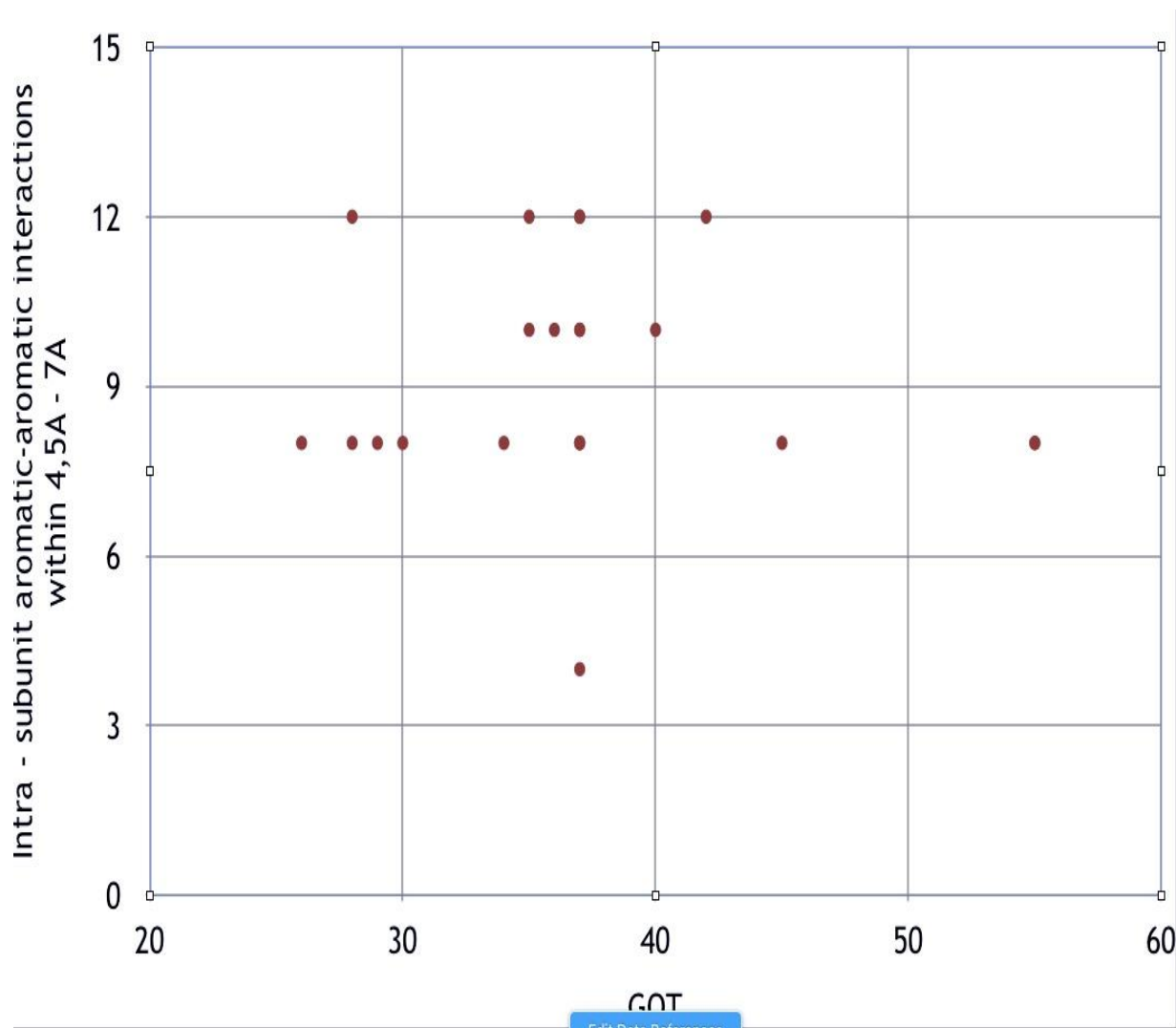
Εικόνα 47 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ πλευρικών αλυσίδων, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι προκύπτει κάποια τάση, αλλά όχι γραμμική συσχέτιση μεταξύ των διαμοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ πλευρικών αλυσίδων και της θερμοκρασίας ανάπτυξης των οργανισμών.



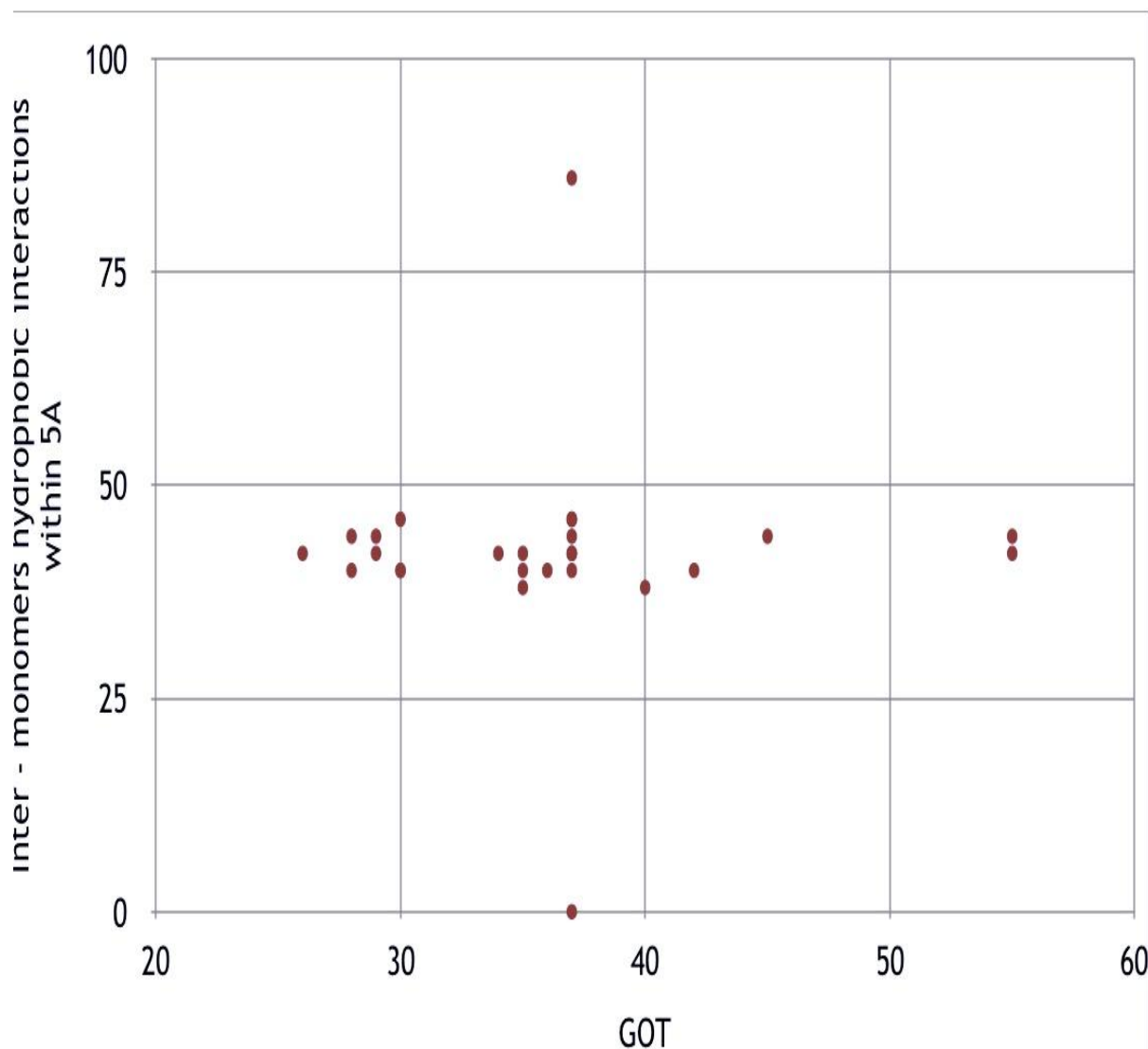
Εικόνα 48 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των Χ συναρτήσει των διαμοριακών ιοντικών αλληλεπιδράσεων στην απόσταση των 6Å , οι οποίες απεικονίζονται στον άξονα των Υ.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια τάση γραμμική συσχέτιση μεταξύ των διαμοριακών ιοντικών αλληλεπιδράσεων στην απόσταση των 6Å και της θερμοκρασίας ανάπτυξης των οργανισμών.



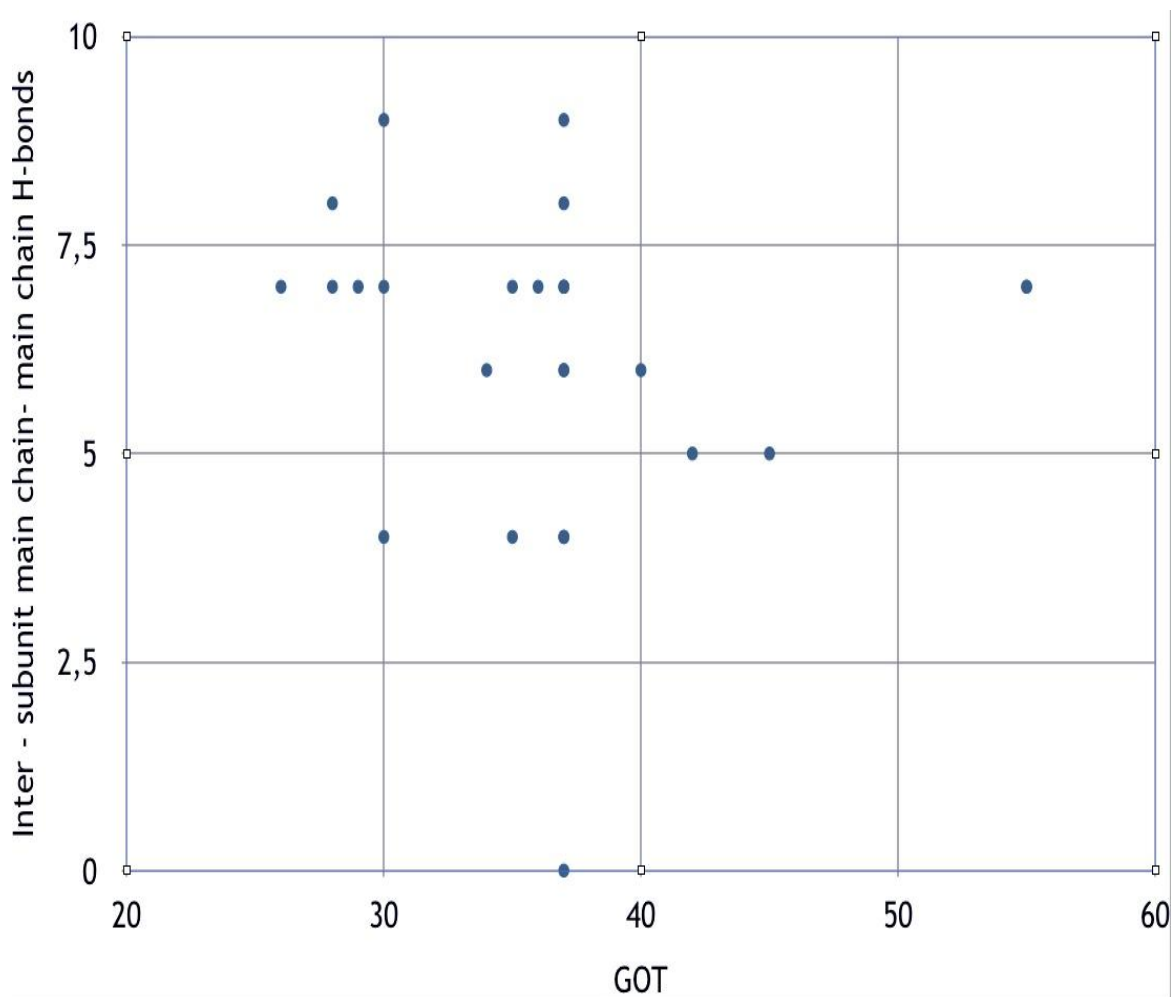
Εικόνα 49 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των διαμοριακών αρωματικών αλληλεπιδράσεων στην απόσταση από 4,5Å - 7Å, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια τάση γραμμική συσχέτιση μεταξύ των διαμοριακών ιοντικών αλληλεπιδράσεων στην απόσταση των 4,5 Å - 7Å και της θερμοκρασίας ανάπτυξης των οργανισμών.



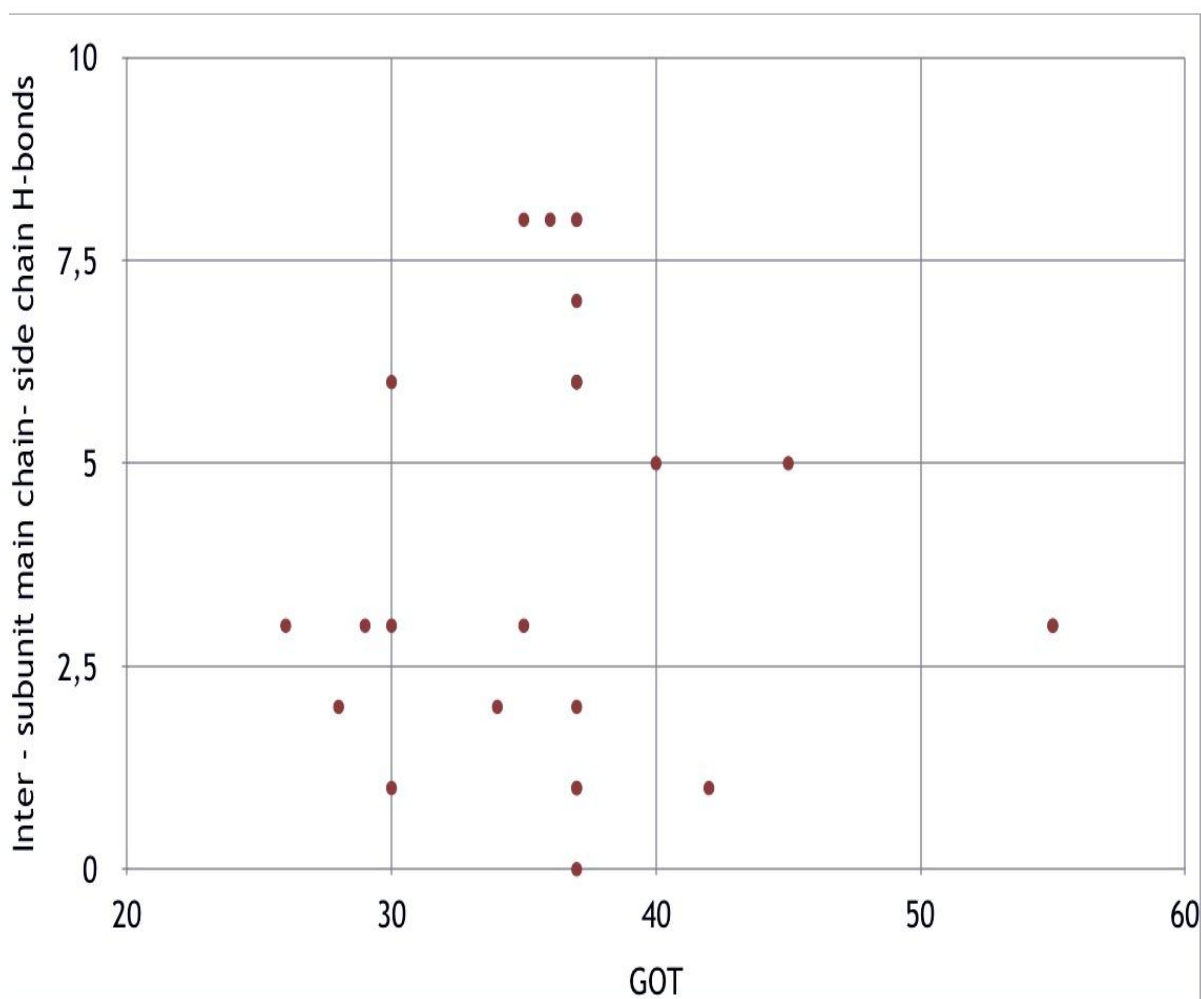
Εικόνα 50 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών υδροφοβικών αλληλεπιδράσεων, οι οποίες απεικονίζονται στον άξονα των Y, στην απόσταση των 5Å.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών υδροφοβικών αλληλεπιδράσεων, σε απόσταση 5Å, και της θερμοκρασίας ανάπτυξης των οργανισμών.



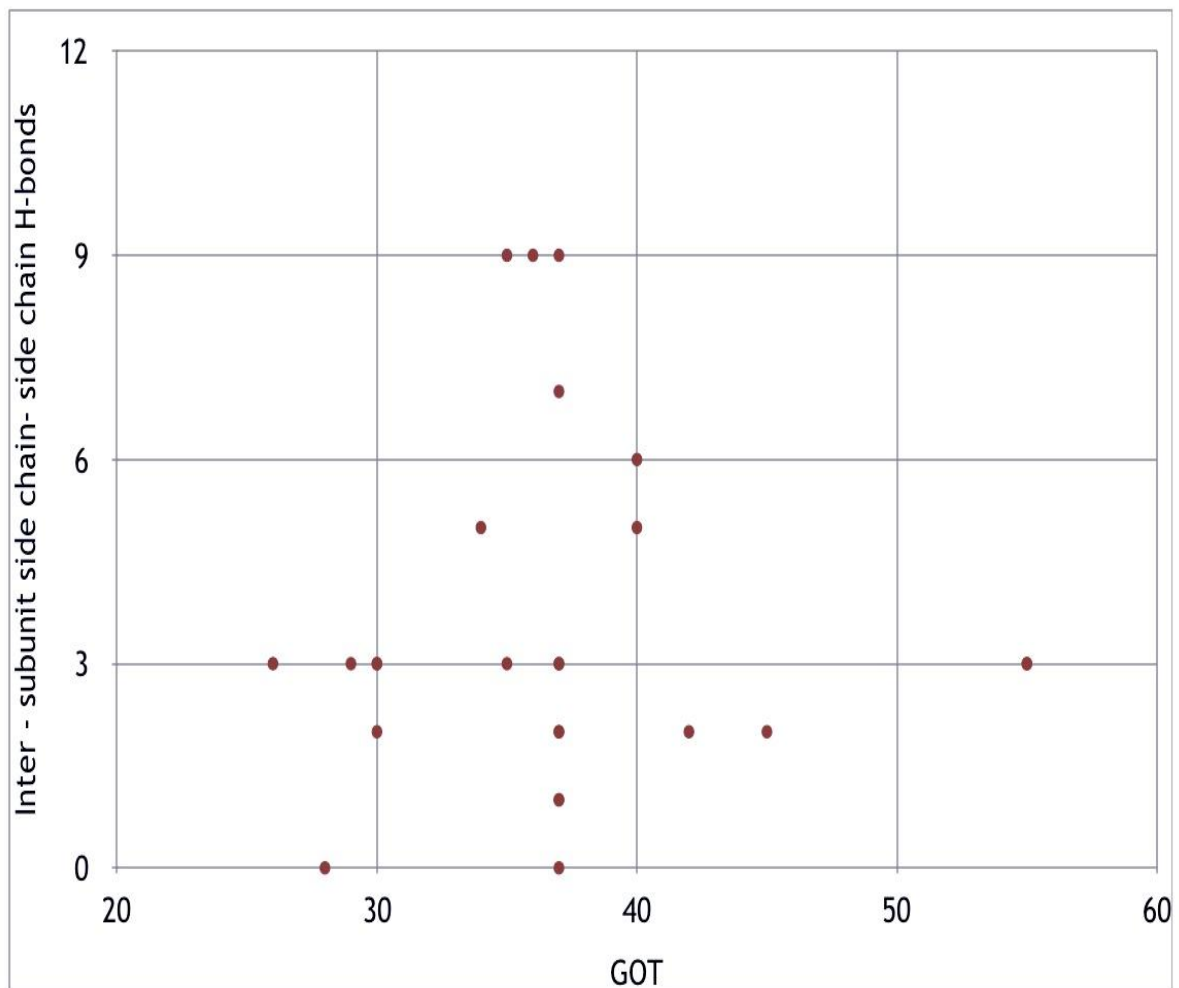
Εικόνα 51 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριων αλυσίδων, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριων αλυσίδων και της θερμοκρασίας ανάπτυξης των οργανισμών.



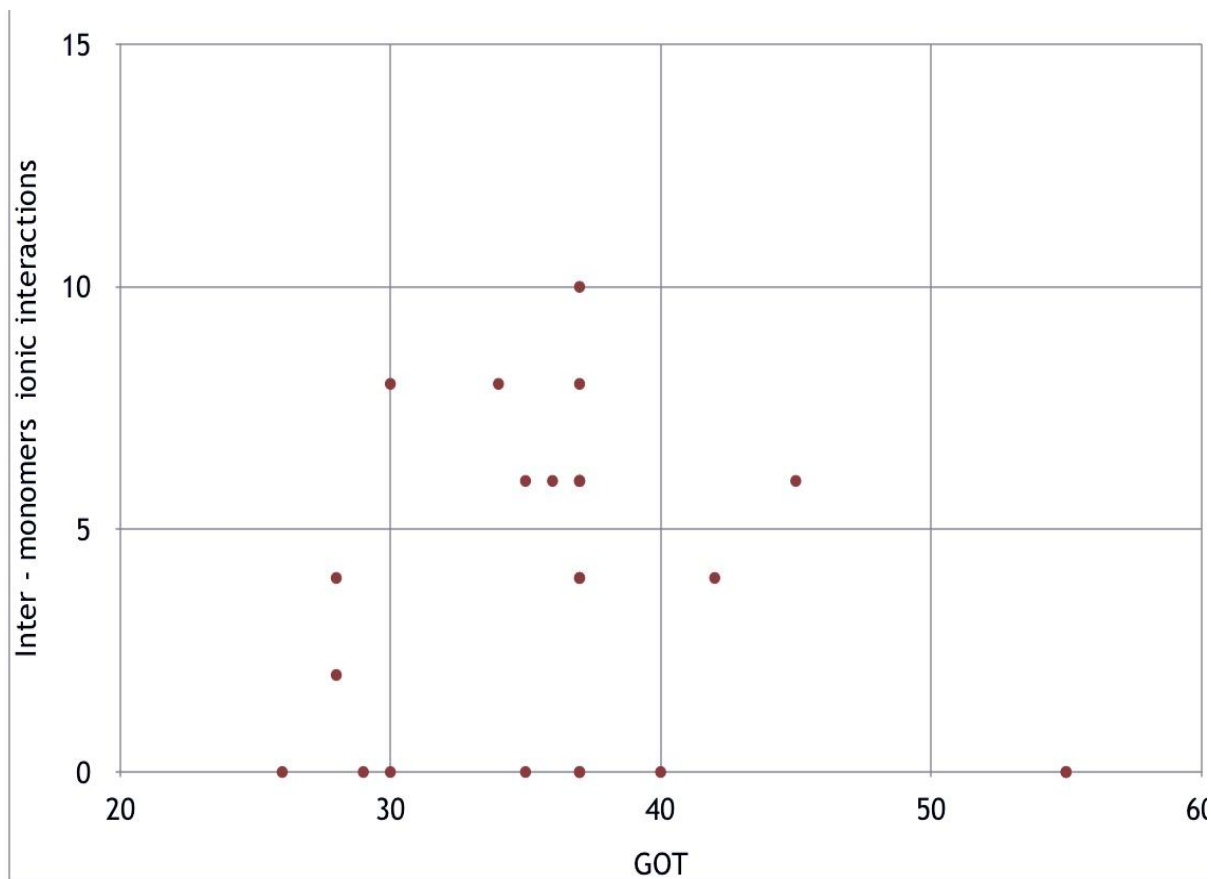
Εικόνα 52 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριας και πλευρικής αλυσίδας, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ κύριας και πλευρικής αλυσίδας και της θερμοκρασίας ανάπτυξης των οργανισμών.



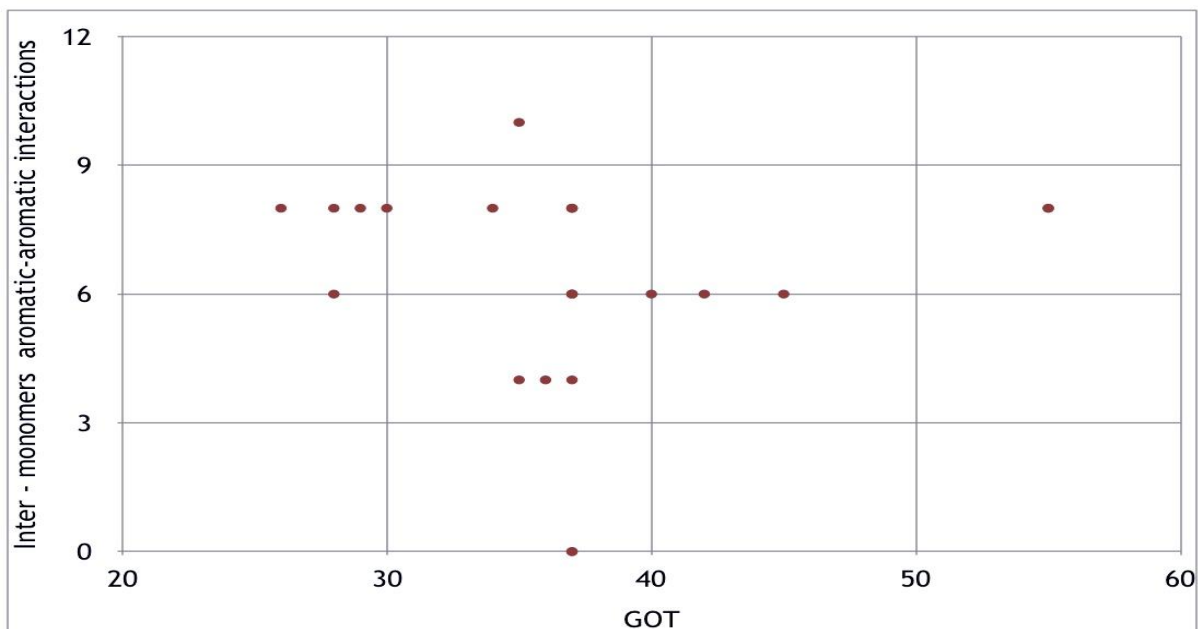
Εικόνα 53 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ πλευρικών αλυσίδων, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών αλληλεπιδράσεων υδρογονικών δεσμών μεταξύ πλευρικών αλυσίδων και της θερμοκρασίας ανάπτυξης των οργανισμών.



Εικόνα 54 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών ιοντικών αλληλεπιδράσεων, οι οποίες απεικονίζονται στον άξονα των Y.

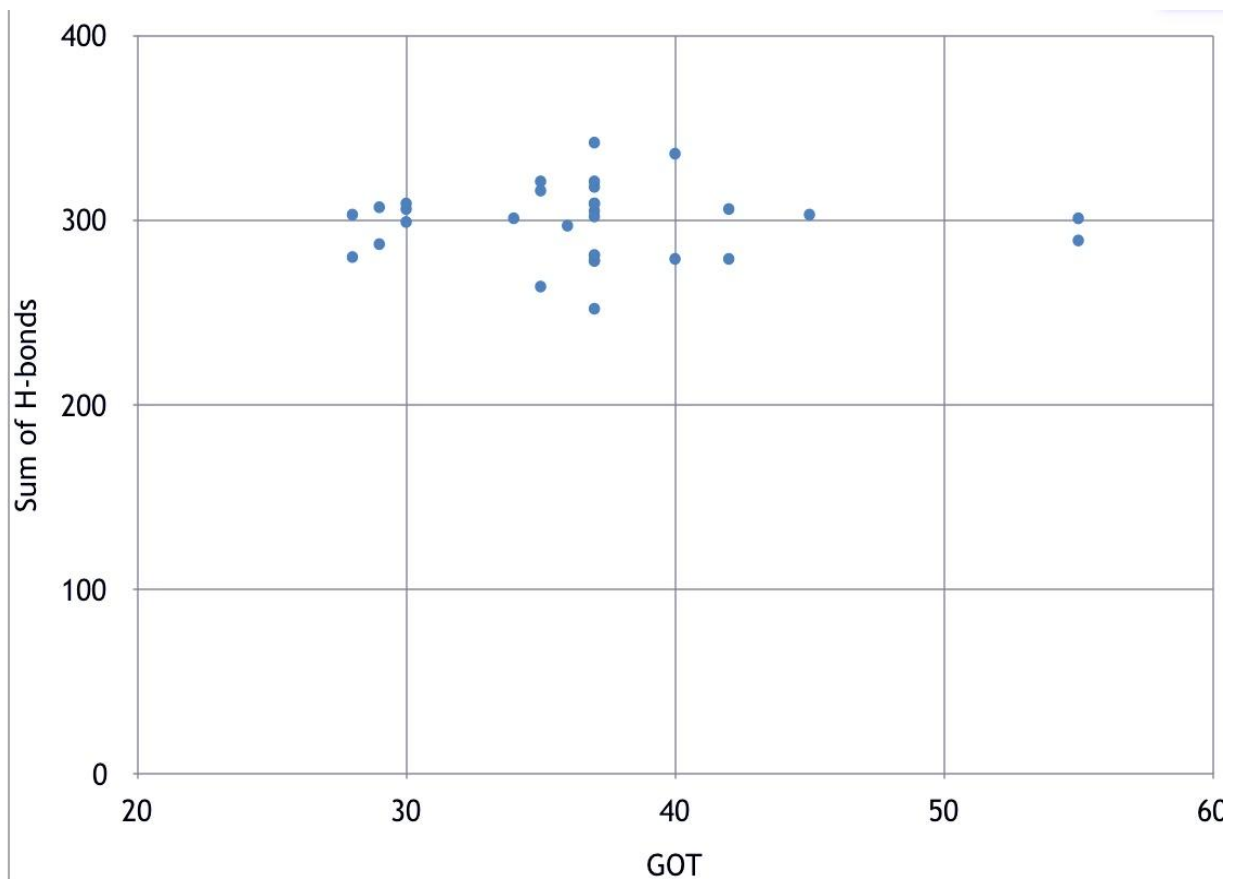
Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών ιοντικών αλληλεπιδράσεων και της θερμοκρασίας ανάπτυξης των οργανισμών.



Εικόνα 55 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει των ενδομοριακών αρωματικών αλληλεπιδράσεων, οι οποίες απεικονίζονται στον άξονα των Y.

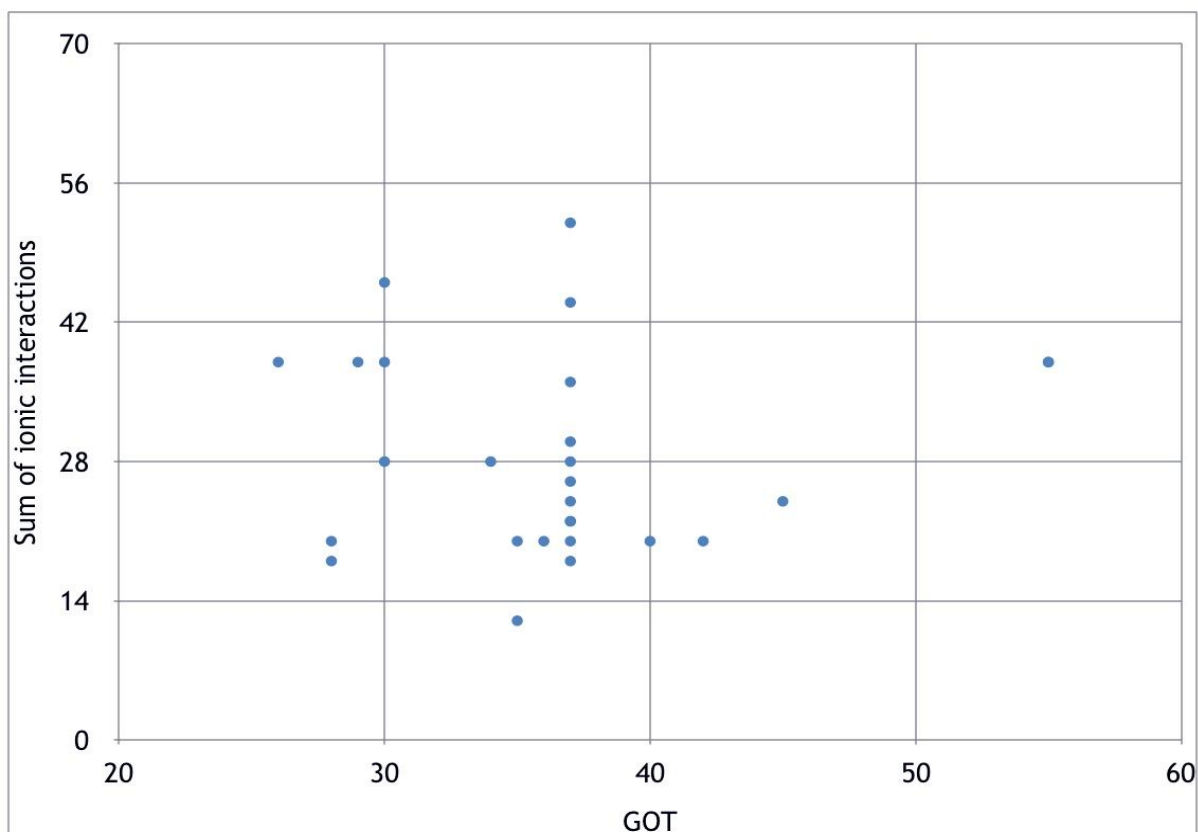
Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ των ενδομοριακών αρωματικών αλληλεπιδράσεων μεταξύ κύριας και πλευρικής αλυσίδας και της θερμοκρασίας ανάπτυξης των οργανισμών.

Ακολουθώντας για να επαληθεύσουμε βιοστατιστικά τα εν λόγω αποτελέσματα, αθροίσαμε όλες τις υδρογονικές καθώς και τις ιοντικές αλληλεπιδράσεις των μοντέλων τόσο σε διαμοριακό όσο και σε ενδομοριακό επίπεδο. Παρακάτω παρατίθενται τα αντίστοιχα διαγράμματα.



Εικόνα 56 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει του αθροίσματος των υδρογονικών αλληλεπιδράσεων τόσο σε διαμοριακό, όσο και σε ενδομοριακό επίπεδο, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ του αθροίσματος των υδρογονικών αλληλεπιδράσεων και της θερμοκρασίας ανάπτυξης των οργανισμών.



Εικόνα 57 Διάγραμμα διασποράς της θερμοκρασίας ανάπτυξης των οργανισμών, οι οποίες απεικονίζονται στον άξονα των X συναρτήσει του αθροίσματος των ιοντικών αλληλεπιδράσεων τόσο σε διαμοριακό, όσο και σε ενδομοριακό επίπεδο, οι οποίες απεικονίζονται στον άξονα των Y.

Από το παραπάνω διάγραμμα παρατηρούμε ότι δεν προκύπτει κάποια γραμμική συσχέτιση μεταξύ του αθροίσματος των ιοντικών αλληλεπιδράσεων και της θερμοκρασίας ανάπτυξης των οργανισμών.

Εν κατακλείδι, από τα παραπάνω διαγράμματα, μπορούμε να συμπεράνουμε ότι ο αριθμός των μη ομοιοπολικών αλληλεπιδράσεων εντός του εκάστοτε μονομερούς καθώς επίσης και των αλληλεπιδράσεων μεταξύ των μονομερών τόσο σε επίπεδο κύριας αλυσίδας, όσο και σε επίπεδο πλευρικών ομάδων, δεν εξαρτάται γραμμικά από την θερμοκρασία βέλτιστης ανάπτυξης του οργανισμού στον οποίο έχουν εκφραστεί. Συνεπώς δεν είναι δυνατό να εντοπιστεί κάποιο χαρακτηριστικό γνώρισμα ή έστω κάποια τιμή κατωφλίου στις ανωτέρω τιμές των αλληλεπιδράσεων η οποία θα μπορούσε να διαχωρίσει αν πρόκειται για μεσόφιλη ή ακραιόφιλη ΗU.

5. Συζήτηση

Ο βαθμός σταθερότητας στην δομή των πρωτεϊνών, η οποία είναι άμεσα συνυφασμένη με την λειτουργία της, αποτελεί ερευνητικό πόλο έλξης εδώ και 60 χρόνια για την επιστημονική κοινότητα. Το βιολογικό πρόβλημα που τίθεται είναι ποιοι είναι οι μηχανισμοί και πως μπορούν να μελετηθούν προκειμένου να αποκρυπτογραφηθούν οι σχέσεις και οι αλληλεπιδράσεις των δευτεροταγών και τριτοταγών δομών μιας πρωτεΐνης και πώς αυτές μπορούν να συμβάλλουν στην φυσικοχημική σταθερότητα των πρωτεϊνικών δομών στον χώρο.

Μία από τις πολλές προσεγγίσεις που έχουν πραγματοποιηθεί, είναι με την βοήθεια της μοριακής δυναμικής μέσω υπολογισμού της ελεύθερης ενεργείας Gibbs των αλληλεπιδράσεων των ατόμων που τις απαρτίζουν. Από τις μεταβολές της τιμής, της ενεργείας αυτής, μπορεί να προβλεφθεί ποιο είναι το ενεργειακό όριο ώστε μια πρωτεϊνική δομή να διατηρεί την δομική της οντότητα σε διαφορετικές περιβαλλοντολογικές, χημικές ή άλλες βιολογικές παραμέτρους (συγκέντρωση του διαλύτη, αντικαταστάσεις σε μεμονωμένα αμινοξέα κ.τ.λ). Εκτός από την προσέγγιση της μοριακής δυναμικής, η ανάπτυξη της κρυσταλλογραφίας έχει συνδράμει αρκετά σε αυτή την προσπάθεια καθώς μέσω των λυμένων κρυσταλλογραφικά δομών μπορούν να μελετηθούν αναλυτικά σε τρισδιάστατο επίπεδο, τόσο η διάταξη των ατόμων της κύριας πρωτεϊνικής αλυσίδας αλλά και η διάταξη, αλληλεπίδραση, και ο σχηματισμός - κατανομή μη ομοιοπολικών αλληλεπιδράσεων μεταξύ των πλευρικών αλυσίδων μιας πρωτεΐνης ή και ακόμη μιας οικογένειας πρωτεϊνών.

Τα τελευταία χρόνια έχουν απομονωθεί και μελετηθεί οργανισμοί από περιβαλλοντολογικές συνθήκες που εδώ και μισό αιώνα αρχίζουν και μελετώνται πιο συστηματικά λόγω της πιθανής εφαρμογής τους σε βιοτεχνολογικές διαδικασίες όπου πολύ συχνά απαιτούνται ακραίες συνθήκες PH και θερμοκρασίας.

Η ανακάλυψη ύπαρξης άλλα και επιβίωσης οργανισμών σε τόσο αφιλόξενες συνθήκες μας έθεσε ένα ακόμη βιολογικό ερώτημα, ποιο εξειδικευμένο από εκείνο της σταθερότητας των πρωτεϊνών. Με την ύπαρξη ζωής δεχόμαστε a priori την ύπαρξη και την λειτουργία πρωτεϊνών στις ακραίες αυτές συνθήκες. Ποιοι είναι όμως

εκείνοι οι μηχανισμοί και ποία η σύσταση των πρωτεϊνών αυτών ώστε να παρουσιάζουν μια προσαρμοσμένη δομική και χημική σύσταση σε τέτοιες συνθήκες ; Η αλλιώς πως μπορεί να μελετηθεί η θερμοσταθερότητα των πρωτεϊνών στους ακραιόφιλους οργανισμούς;

Η δυσκολία στην μελέτη των μηχανισμών θερμοσταθερότητας που αναφέρθηκαν παραπάνω εστιάζεται στο γεγονός ότι η θερμοσταθερότητα των πρωτεϊνών αντανακλά έναν αριθμό από ελάχιστες αλληλεπιδράσεις που δρουν αθροιστικά και είναι συχνά χαρακτηριστικές για κάθε πρωτεΐνη [37]. Η επικρατούσα άποψη είναι ότι διαφορετικοί παράγοντες σταθεροποιούν διαφορετικές πρωτεΐνες [42]. Για τον λόγο αυτό ενδείκνυται η μελέτη των μηχανισμών θερμοσταθερότητας να γίνεται σε μια συγκεκριμένη πρωτεΐνη κάθε φορά. Αυτό μπορεί να επιτευχθεί μέσω πολύ λεπτομερούς σύγκρισης π.χ. μιας θερμοσταθερής πρωτεΐνης με την ομόλογη μεσόφιλη ή ψυχρόφιλη της [42]. Στην αποτελεσματικότητα αυτής της κατεύθυνσης συμβάλει ο γρήγορα αυξανόμενος αριθμός κατατεθειμένων επιλυμένων τρισδιάστατων δομών (με χρήση κρυσταλλογραφίας ή NMR) [54, 56].

Ανάμεσα στις επιλυμένες τρισδιάστατες δομές είναι και η πρωτεΐνη HU η όποια έχει λυθεί από πολλούς οργανισμούς που αναπτύσσονται σε διαφορετικές θερμοκρασίες. Επίσης έχει παρατηρηθεί ότι η θερμοσταθερότητα που παρουσιάζει η παραπάνω πρωτεΐνη είναι ανάλογη της θερμοκρασίας ανάπτυξης του οργανισμού που έχει απομονωθεί. Ακόμη για τις HU έχει παρατηρηθεί ένα μεγάλο εύρος θερμοδυναμικής αντιστρεψιμότητας το όποιο μας επιτρέπει πειράματα και μελέτη στην διερεύνηση της θερμοσταθερότητας. Επιπλέον όλες οι HU πρωτεΐνες είναι σε διμερή μορφή άρα μπορούμε να μελετήσουμε τους μηχανισμούς διμερισμού και σταθεροποίησης των διμερών. Τέλος, παρατηρείται μεγάλο ποσοστό ομολογίας από 30% μέχρι 85% που μας επιτρέπει με ασφάλεια να μελετήσουμε και τις μη λυμένες δομές, χρησιμοποιώντας δομική μοντελοποίηση βασισμένη στην ομολογία.

Εφαρμογές όπως η δομική μοντελοποίηση, η στοίχιση ακολουθιών, ο υπολογισμός μοριακών αλληλεπιδράσεων και η αποτύπωση τρισδιάστατων απεικονίσεων βιομορίων είναι μερικά από τα εργαλεία του κλάδου της βιοπληροφορικής που συνδράμουν με την σειρά τους στην έρευνα και στην μελέτη των πρωτεϊνών. Η

εκθετική ανάπτυξη της τεχνολογίας έχει επιφέρει μεγάλες αλλαγές ως προς την ισχύ, την ταχύτητα αλλά και την αύξηση των ικανοτήτων των ηλεκτρονικών υπολογιστών. Η ανάπτυξη αυτή σε συνδυασμό με την εξέλιξη στους αλγορίθμους αλλά και στις γλώσσες προγραμματισμού ανοίγουν νέους ορίζοντες στο πεδίο της βιολογίας. Στο πεδίο της δομικής βιολογίας και της βιοχημείας υπάρχει πλέον η επεξεργασία μεγαλύτερου όγκου δεδομένων σε μικρότερο χρόνο επεξεργασίας, η δυνατότητα στοίχισης ακολουθιών σε μερικά δευτερόλεπτα, ύπαρξη μεγάλων βάσεων δεδομένων όπου υπάρχουν λυμένες κρυσταλλογραφικά δομές. Σημαντική επίσης είναι και η ανάπτυξη αλγορίθμων που επιτρέπουν την κατασκευή πρωτεϊνικών μοντέλων τα οποία δεν υπάρχει η αντίστοιχη κρυσταλλογραφική δομή. Όλα τα παραπάνω μειώνουν τον χρόνο επεξεργασίας και έρευνας με αποτέλεσμα να υπάρχει εξοικονόμηση τόσο οικονομική όσο και ανθρωπίνων πόρων. Επειδή όμως η τεχνολογία των αλγορίθμων δεν είναι ακόμα αρκετά ώριμη, τα λάθη δεν αποφεύγονται και για το λόγο αυτό οι απαντήσεις που δίνουν οι πρώτοι πρέπει πάντα να εξετάζονται με κριτική-βιολογική σκέψη.

Στην παρούσα διπλωματική εργασία, πραγματοποιήθηκε βιοπληροφορική προσέγγιση των πρωτεϊνών της οικογένειας HU. Πρωτίστως, απομονώθηκαν πρωτεΐνες, της εν λόγω οικογένειας, οι οποίες δεν έχουν λυμένη κρυσταλλογραφικά τρισδιάστατη δομή, από τις δευτερογενείς καταχωρήσεις των πρωτεϊνικών βάσεων δεδομένων. Στην συνέχεια αναζητήθηκαν δομικά πρότυπα, στοιχήθηκαν οι μεταξύ τους ακολουθίες, και κατασκευαστήκαν θεωρητικά μοντέλα. Τα νεοσύστατα θεωρητικά μοντέλα αφού κατασκευάστηκαν, στη συνέχεια αξιολογήθηκαν. Το επόμενο βήμα ήταν η κατασκευή ομοδιμερών ή ετεροδιμερών μοντέλων αναλόγως με την φύση του στελέχους (A ή B) αλλά και με τον οργανισμό που ανήκαν οι αρχικές καταχωρήσεις. Επιπροσθέτως, εντοπίστηκε η περιοχή σήματος ομοδιμερισμού, απομονώθηκε εκτελέστηκε δομική υπέρθεση με την αντίστοιχη περιοχή σήματος ομοδιμερισμού της δομής- οδηγού. Τέλος, πραγματοποιήθηκε ανάλυση των μη ομοιοπολικών αλληλεπιδράσεων εντός του εκάστοτε μονομερούς καθώς επίσης και των αλληλεπιδράσεων μεταξύ των μονομερών τόσο σε επίπεδο κύριας αλυσίδας, όσο και σε επίπεδο πλευρικών ομάδων. Ακολούθησε διερεύνηση των πιθανών συσχετισμών όλων αυτών των παραμέτρων που αναφέραμε με την θερμοκρασία ανάπτυξης του εκάστοτε οργανισμού από τον

οποίο απομονώθηκε κάθε HU. Τα αποτελέσματα σαφώς δεν μας επιτρέπουν να βρούμε στατιστικά σημαντική αλληλεξάρτηση μεταξύ των αριθμού του κάθε είδους μη ομοιοπολικών αλληλεπιδράσεων σε σχέση με την θερμοκρασία ανάπτυξης του εκάστοτε βακτηρίου. Παρά το γεγονός ότι η οικογένεια των HU παρουσιάζει συντηρημένη τριτοταγή δομή, το είδος και ο αριθμός των μη ομοιοπολικών αλληλεπιδράσεων είναι ανεξάρτητα από την θερμοκρασία ανάπτυξης του οργανισμού από τον οποίο προέρχονται και οι HU που μελετήσαμε.

Δυστυχώς έχουμε περιορισμένα πειραματικά δεδομένα που να αφορούν τον τρόπο σταθεροποίησης των διαφόρων μελών της οικογένειας HU τόσο σε επίπεδο δομικό όσο και σε επίπεδο βιοφυσικό και θερμοδυναμικό. Ως εκ τούτου αδυνατούμε προς το παρόν να εκφράσουμε έστω και ασαφείς κανόνες για τις στρατηγικές. Που χρησιμοποιούν οι πρωτεΐνες HU στην προσπάθειά τους να προσαρμοστούν σε διάφορα θερμοκρασιακά περιβάλλοντα. Το μόνο σαφές δεδομένο είναι ότι μεταξύ δυο ειδών HU του ίδιου γένους (*Bacillus stearothermophilus* και *Bacillus subtilis*) τρία συγκεκριμένα αμινοξέα που βρίσκονται στρατηγικά τοποθετημένα στην δομή των δύο πρωτεϊνών, έχει αποδειχτεί ότι είναι υπεύθυνα για τον μηχανισμό σταθεροποίησης των δυο αυτών μορίων. Το παράδειγμα αυτό βέβαια δεν μας δίνει την δυνατότητα γενίκευσης των αποτελεσμάτων άλλα υποστηρίζει για μια ακόμη φορά ότι κάθε μια πρωτεΐνη έχει βρει το δικό της μηχανισμό στα πλαίσια που της επιτρέπουν, η πρωτοταγή της και δευτεροταγή της δομή, για την σταθεροποίησή της. Είναι σημαντικό να αναφέρουμε ότι οι HU πρωτεΐνες σταθεροποιούνται ενδογενώς, δηλαδή η σταθερότητα τους προέρχεται αμιγώς από την πρωτοταγή τους δομή και δεν απαιτείται η παρουσία μη πρωτεϊνικών παραγόντων όπως π.χ. δισθενών μετάλλων ή άλλων μεταβολιτών. Παρόλο αυτά οι HU πρωτεΐνες είναι ένα ελκυστικό μοντέλο διότι είναι πρωτεΐνες που μπορούν να παραχθούν και να απομονωθούν σχετικά εύκολα σε μεγάλες ποσότητες, είναι κρυσταλλώσιμες και αποδιατάσσονται και επαναδιατάσσονται με πλήρη αντιστρεψιμότητα. Το τελευταίο είναι ένα σημαντικό πλεονέκτημα για την πλήρη θερμοδυναμική μελέτη των μορίων αυτών.

Βιβλιογραφία

1. Neidhardt, F.C., *Escherichia coli and Salmonella typhimurium : cellular and molecular biology*. 1997, Washington, D.C.: American Society for Microbiology.
2. Drlica, K. and J. Rouviere-Yaniv, *Histonelike proteins of bacteria*. Microbiol Rev, 1987. **51**(3): p. 301-19.
3. Rouviere-Yaniv, J., M. Yaniv, and J.E. Germond, *E. coli DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA*. Cell, 1979. **17**(2): p. 265-74.
4. Oberto, J. and J. Rouviere-Yaniv, *Serratia marcescens contains a heterodimeric HU protein like Escherichia coli and Salmonella typhimurium*. J Bacteriol, 1996. **178**(1): p. 293-7.
5. Kano, Y., et al., *Molecular cloning and nucleotide sequence of the HU-1 gene of Escherichia coli*. Mol Gen Genet, 1985. **201**(2): p. 360-2.
6. Kano, Y., et al., *Cloning and sequencing of the HU-2 gene of Escherichia coli*. Mol Gen Genet, 1987. **209**(2): p. 408-10.
7. Pettijohn, D.E., *Histone-like proteins and bacterial chromosome structure*. J Biol Chem, 1988. **263**(26): p. 12793-6.
8. Almiron, M., et al., *A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli*. Genes Dev, 1992. **6**(12B): p. 2646-54.
9. Rouviere-Yaniv, J. and F. Gros, *Characterization of a novel, low-molecular-weight DNA-binding protein from Escherichia coli*. Proc Natl Acad Sci U S A, 1975. **72**(9): p. 3428-32.
10. Briat, J.F. and M.J. Chamberlin, *Identification and characterization of a new transcriptional termination factor from Escherichia coli*. Proc Natl Acad Sci U S A, 1984. **81**(23): p. 7373-7.

11. Neilan, J.G., et al., *An African swine fever virus gene with similarity to bacterial DNA binding proteins, bacterial integration host factors, and the Bacillus phage SP01 transcription factor, TF1*. Nucleic Acids Res, 1993. **21**(6): p. 1496.
12. Greene, J.R., et al., *Sequence of the bacteriophage SP01 gene coding for transcription factor 1, a viral homologue of the bacterial type II DNA-binding proteins*. Proc Natl Acad Sci U S A, 1984. **81**(22): p. 7031-5.
13. Claret, L. and J. Rouviere-Yaniv, *Regulation of HU alpha and HU beta by CRP and FIS in Escherichia coli*. J Mol Biol, 1996. **263**(2): p. 126-39.
14. Bahloul, A., F. Boubrik, and J. Rouviere-Yaniv, *Roles of Escherichia coli histone-like protein HU in DNA replication: HU-beta suppresses the thermosensitivity of dnaA46ts*. Biochimie, 2001. **83**(2): p. 219-29.
15. Pinson, V., M. Takahashi, and J. Rouviere-Yaniv, *Differential binding of the Escherichia coli HU, homodimeric forms and heterodimeric form to linear, gapped and cruciform DNA*. J Mol Biol, 1999. **287**(3): p. 485-97.
16. Searcy, D.G., *Histone-like protein in the prokaryote Thermoplasma acidophilum*. Biochim Biophys Acta, 1975. **395**(4): p. 535-47.
17. Searcy, D.G. and R.J. Delange, *Thermoplasma acidophilum histone-like protein. Partial amino acid sequence suggestive of homology to eukaryotic histones*. Biochim Biophys Acta, 1980. **609**(1): p. 197-200.
18. Broyles, S.S. and D.E. Pettijohn, *Interaction of the Escherichia coli HU protein with DNA. Evidence for formation of nucleosome-like structures with altered DNA helical pitch*. J Mol Biol, 1986. **187**(1): p. 47-60.
19. Shindo, H., et al., *Preferential binding of E.coli histone-like protein HU alpha to negatively supercoiled DNA*. Nucleic Acids Res, 1992. **20**(7): p. 1553-8.

20. Tanaka, H., et al., *Role of HU proteins in forming and constraining supercoils of chromosomal DNA in Escherichia coli*. Mol Gen Genet, 1995. **248**(5): p. 518-26.
21. Kobryn, K., B.D. Lavoie, and G. Chaconas, *Supercoiling-dependent site-specific binding of HU to naked Mu DNA*. J Mol Biol, 1999. **289**(4): p. 777-84.
22. Aki, T. and S. Adhya, *Repressor induced site-specific binding of HU for transcriptional regulation*. EMBO J, 1997. **16**(12): p. 3666-74.
23. Boubrik, F. and J. Rouviere-Yaniv, *Increased sensitivity to gamma irradiation in bacteria lacking protein HU*. Proc Natl Acad Sci U S A, 1995. **92**(9): p. 3958-62.
24. Li, S. and R. Waters, *Escherichia coli strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination*. J Bacteriol, 1998. **180**(15): p. 3750-6.
25. Swinger, K.K., et al., *Flexible DNA bending in HU-DNA cocrystal structures*. EMBO J, 2003. **22**(14): p. 3749-60.
26. Pilch, D.S., et al., *Characterizing the DNA binding modes of a topoisomerase I-poisoning terbenzimidazole: evidence for both intercalative and minor groove binding properties*. Drug Des Discov, 1996. **13**(3-4): p. 115-33.
27. Balandina, A., D. Kamashev, and J. Rouviere-Yaniv, *The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids*. J Biol Chem, 2002. **277**(31): p. 27622-8.
28. Paci, M., et al., *Proteins from the prokaryotic nucleoid. High-resolution 1H NMR spectroscopic study of Escherichia coli DNA-binding proteins NS1 and NS2*. Eur J Biochem, 1984. **138**(1): p. 193-200.
29. Wang, S.L. and X.Q. Liu, *The plastid genome of Cryptomonas phi encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein*.

- Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(23): p. 10783-10787.
30. Kimura, M. and K.S. Wilson, *On the DNA binding protein II from Bacillus stearothermophilus. II. The amino acid sequence and its relation to those of homologous proteins from other prokaryotes.* J Biol Chem, 1983. **258**(6): p. 4007-11.
 31. Dijk, J., et al., *On the DNA binding protein II from Bacillus stearothermophilus. I. Purification, studies in solution, and crystallization.* J Biol Chem, 1983. **258**(6): p. 4003-6.
 32. Tanaka, I., et al., *3-A resolution structure of a protein with histone-like properties in prokaryotes.* Nature, 1984. **310**(5976): p. 376-81.
 33. White, S.W., et al., *A protein structural motif that bends DNA.* Proteins, 1989. **5**(4): p. 281-8.
 34. Padas, P.M., K.S. Wilson, and C.E. Vorgias, *The DNA-binding protein HU from mesophilic and thermophilic bacilli: gene cloning, overproduction and purification.* Gene, 1992. **117**(1): p. 39-44.
 35. Vis, H., et al., *Solution structure of the HU protein from Bacillus stearothermophilus.* J Mol Biol, 1995. **254**(4): p. 692-703.
 36. Saitoh, F., et al., *Arginine-55 in the beta-arm is essential for the activity of DNA-binding protein HU from Bacillus stearothermophilus.* Biosci Biotechnol Biochem, 1999. **63**(12): p. 2232-5.
 37. Christodoulou, E. and C.E. Vorgias, *The thermostability of DNA-binding protein HU from mesophilic, thermophilic, and extreme thermophilic bacteria.* Extremophiles, 2002. **6**(1): p. 21-31.
 38. Rouvière-Yaniv, J. and F. Gros, *Characterization of a novel, low-molecular-weight DNA-binding protein from Escherichia coli.* Proceedings of the

- National Academy of Sciences of the United States of America, 1975. **72**(9): p. 3428-3432.
39. Wintrode, P.L. and F.H. Arnold, *Temperature adaptation of enzymes: lessons from laboratory evolution*. Adv Protein Chem, 2000. **55**: p. 161-225.
 40. Jaenicke, R., *Protein stability and molecular adaptation to extreme conditions*. Eur J Biochem, 1991. **202**(3): p. 715-28.
 41. Jaenicke, R. and G. Bohm, *The stability of proteins in extreme environments*. Curr Opin Struct Biol, 1998. **8**(6): p. 738-48.
 42. Fu, H., et al., *Increasing protein stability: importance of $\Delta C(p)$ and the denatured state*. Protein Sci, 2010. **19**(5): p. 1044-52.
 43. Taylor, T.J. and Vaisman, II, *Discrimination of thermophilic and mesophilic proteins*. BMC Struct Biol, 2010. **10 Suppl 1**: p. S5.
 44. Zhou, X.X., et al., *Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins*. Amino Acids, 2008. **34**(1): p. 25-33.
 45. Xu, Y., et al., *Moritella cold-active dihydrofolate reductase: are there natural limits to optimization of catalytic efficiency at low temperature?* J Bacteriol, 2003. **185**(18): p. 5519-26.
 46. Bennett, A.F. and R.E. Lenski, *Phenotypic and evolutionary adaptation of a model bacterial system to stressful thermal environments*. EXS, 1997. **83**: p. 135-54.
 47. Suhre, K. and J.M. Claverie, *Genomic correlates of hyperthermostability, an update*. J Biol Chem, 2003. **278**(19): p. 17198-202.
 48. Tattersall, G.J., et al., *Coping with thermal challenges: physiological adaptations to environmental temperatures*. Compr Physiol, 2012. **2**(3): p. 2151-202.

49. Dams, T. and R. Jaenicke, *Stability and folding of dihydrofolate reductase from the hyperthermophilic bacterium Thermotoga maritima*. *Biochemistry*, 1999. **38**(28): p. 9169-78.
50. Daniel, R.M., M. Dines, and H.H. Petach, *The denaturation and degradation of stable enzymes at high temperatures*. *Biochem J*, 1996. **317** (Pt 1): p. 1-11.
51. Szilagyi, A. and P. Zavodszky, *Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey*. *Structure*, 2000. **8**(5): p. 493-504.
52. Leone, M., et al., *Solution structure and backbone dynamics of the K18G/R82E Alicyclobacillus acidocaldarius thioredoxin mutant: a molecular analysis of its reduced thermal stability*. *Biochemistry*, 2004. **43**(20): p. 6043-58.
53. Kumwenda, B., et al., *Analysis of protein thermostability enhancing factors in industrially important thermus bacteria species*. *Evol Bioinform Online*, 2013. **9**: p. 327-42.
54. Pechkova, E., V. Sivozhelezov, and C. Nicolini, *Protein thermal stability: the role of protein structure and aqueous environment*. *Arch Biochem Biophys*, 2007. **466**(1): p. 40-8.
55. Maes, D., et al., *The crystal structure of triosephosphate isomerase (TIM) from Thermotoga maritima: a comparative thermostability structural analysis of ten different TIM structures*. *Proteins*, 1999. **37**(3): p. 441-53.
56. Dams, T., et al., *The crystal structure of dihydrofolate reductase from Thermotoga maritima: molecular features of thermostability*. *J Mol Biol*, 2000. **297**(3): p. 659-72.
57. Benson, D.A., et al., *GenBank*. *Nucleic Acids Res*, 2013. **41**(Database issue): p. D36-42.
58. Coordinators, N.R., *Database Resources of the National Center for Biotechnology Information*. *Nucleic Acids Res*, 2017. **45**(D1): p. D12-D17.

59. Coordinators, N.R., *Database resources of the National Center for Biotechnology Information*. *Nucleic Acids Res*, 2018. **46**(D1): p. D8-D13.
60. The UniProt, C., *The Universal Protein Resource (UniProt)*. *Nucleic Acids Research*, 2007. **35**(suppl_1): p. D193-D197.
61. The UniProt Consortium, *UniProt: the universal protein knowledgebase*. *Nucleic Acids Research*, 2017. **45**(D1): p. D158-D169.
62. Bairoch, A. and R. Apweiler, *The SWISS-PROT protein sequence database: its relevance to human molecular medical research*. *J Mol Med (Berl)*, 1997. **75**(5): p. 312-6.
63. Boeckmann, B., et al., *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003*. *Nucleic Acids Res*, 2003. **31**(1): p. 365-70.
64. Wu, C.H., et al., *The Protein Information Resource*. *Nucleic Acids Res*, 2003. **31**(1): p. 345-7.
65. Shen, M.Y. and A. Sali, *Statistical potential for assessment and prediction of protein structures*. *Protein Sci*, 2006. **15**(11): p. 2507-24.
66. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. *J Comput Chem*, 2004. **25**(13): p. 1605-12.
67. Tina, K.G., R. Bhadra, and N. Srinivasan, *PIC: Protein Interactions Calculator*. *Nucleic Acids Res*, 2007. **35**(Web Server issue): p. W473-6.

Appendix

Πίνακας 20. Υπολογισμός διαμοριακών και ενδομοριακών αλληλεπιδράσεων μέσω P.I.C των θεωρητικών μοντέλων που κατασκευάστηκαν.

Comparison of the molecular interactions and geometry between the models of HU	B8GX11	E0J6W8	I1WEI8	O68451	P0A1R6- P0A1R8	P0A1R7- P0A1R9	P0A3H1
Intra - subunit interactions							
Intra - subunit hydrophobic interactions within 5A	80	70	80	78	76	76	96
Intra - subunit main chain- main chain H-bonds	117	133	118	99	116	116	113
Intra - subunit main chain- side chain H-bonds	19	28	21	20	22	22	20
Intra - subunit side chain- side chain H-bonds	11	13	5	12	11	11	17
Intra - subunit ionic interactions within 6A	16	48	20	20	16	16	38
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	12	12	10	8	8	8	8
Intra - subunit cation-pi interactions	0	0	0	6	0	0	0
Intra - subunit aromatic -sulphur interactions	0	0	0	0	0	0	0
Inter - subunit interactions							
Inter - monomers hydrophobic interactions within 5A	44	86	38	42	42	42	42
Inter - subunit main chain- main chain H-bonds	7	4	6	6	6	6	7
Inter - subunit main chain- side chain H-bonds	2	7	5	2	6	6	3
Inter - subunit side chain- side chain H-bonds	0	1	6	5	2	2	3
Inter - monomers ionic interactions	4	4	0	8	6	6	0
Inter - monomers aromatic interactions	8	0	6	8	6	6	8
Inter - monomers cation-pi interactions	0	0	1	1	0	0	0
Inter - monomers aromatic -sulphur interactions	0	0	0	0	0	0	0

Πίνακας 21. Υπολογισμός διαμοριακών και ενδομοριακών αλληλεπιδράσεων μέσω P.I.C των θεωρητικών μοντέλων που κατασκευάστηκαν.

Comparison of the molecular interactions and geometry between the models of HU	P0A3H2	P0A3H3	P0A3H4	P0A3H5	P0A3H6	P0ACF1- P0ACF5	P0ACF2- P0ACF6
Intra - subunit interactions							
Intra - subunit hydrophobic interactions within 5A	96	96	94	96	96	70	74
Intra - subunit main chain- main chain H-bonds	115	114	114	115	116	120	122
Intra - subunit main chain- side chain H-bonds	21	21	21	21	21	21	25
Intra - subunit side chain- side chain H-bonds	17	17	17	18	18	12	12
Intra - subunit ionic interactions within 6A	38	38	38	38	38	18	18
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	8	8	8	8	8	8	8
Intra - subunit cation-pi interactions	0	0	0	0	0	0	0
Intra - subunit aromatic -sulphur interactions	0	0	0	0	0	0	0
Inter - subunit interactions							
Inter - monomers hydrophobic interactions within 5A	44	42	40	44	42	42	44
Inter - subunit main chain- main chain H-bonds	7	7	7	7	7	4	5
Inter - subunit main chain- side chain H-bonds	3	3	3	3	3	6	5
Inter - subunit side chain- side chain H-bonds	3	3	3	3	3	3	2
Inter - monomers ionic interactions	0	0	0	0	0	6	6
Inter - monomers aromatic interactions	8	8	8	8	8	6	6
Inter - monomers cation-pi interactions	0	0	0	0	0	0	0
Inter - monomers aromatic -sulphur interactions	0	0	0	0	0	0	0

Πίνακας 22. Υπολογισμός διαμοριακών και ενδομοριακών αλληλεπιδράσεων μέσω P.I.C των θεωρητικών μοντέλων που κατασκευάστηκαν.

Comparison of the molecular interactions and geometry between the models of HU	P0ACF3- P0ACF7	P0CAV2	P0COH2	P0COH3	P0DB64	P0DB65	P0DMK4
Intra - subunit interactions							
Intra - subunit hydrophobic interactions within 5A	90	80	100	102	102	100	78
Intra - subunit main chain- main chain H-bonds	127	114	113	114	113	114	120
Intra - subunit main chain- side chain H-bonds	21	20	23	22	22	23	22
Intra - subunit side chain- side chain H-bonds	14	11	14	14	14	14	5
Intra - subunit ionic interactions within 6A	18	14	26	30	28	28	20
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	4	12	8	8	8	8	10
Intra - subunit cation- π interactions	0	0	0	0	0	0	0
Intra - subunit aromatic -sulphur interactions	0	0	0	0	0	0	0
Inter - subunit interactions							
Inter - monomers hydrophobic interactions within 5A	0	44	46	46	46	46	38
Inter - subunit main chain- main chain H-bonds	0	7	7	8	9	9	6
Inter - subunit main chain- side chain H-bonds	0	2	1	1	1	1	5
Inter - subunit side chain- side chain H-bonds	0	0	1	2	2	2	5
Inter - monomers ionic interactions	6	4	0	0	0	0	0
Inter - monomers aromatic interactions	0	8	8	8	8	8	6
Inter - monomers cation- π interactions	0	0	1	1	1	1	1
Inter - monomers aromatic -sulphur interactions	0	0	0	0	0	0	0

Πίνακας 23. Υπολογισμός διαμοριακών και ενδομοριακών αλληλεπιδράσεων μέσω P.I.C των θεωρητικών μοντέλων που κατασκευάστηκαν.

Comparison of the molecular interactions and geometry between the models of HU	P64386	P64387	P64388	P64389	P68573	P68574	Q5HUP6
Intra - subunit interactions							
Intra - subunit hydrophobic interactions within 5A	108	108	108	106	102	104	102
Intra - subunit main chain- main chain H-bonds	120	119	120	119	113	112	125
Intra - subunit main chain- side chain H-bonds	15	15	15	15	21	21	26
Intra - subunit side chain- side chain H-bonds	12	12	12	12	12	12	5
Intra - subunit ionic interactions within 6A	14	14	14	14	38	36	16
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	10	10	10	10	8	8	12
Intra - subunit cation- π interactions	0	0	0	0	0	0	0
Intra - subunit aromatic -sulphur interactions	3	3	3	3	0	0	2
Inter - subunit interactions							
Inter - monomers hydrophobic interactions within 5A	42	40	42	40	40	40	40
Inter - subunit main chain- main chain H-bonds	7	7	7	7	4	4	5
Inter - subunit main chain- side chain H-bonds	8	8	8	8	6	6	1
Inter - subunit side chain- side chain H-bonds	9	9	9	9	3	3	2
Inter - monomers ionic interactions	6	6	6	6	8	8	4
Inter - monomers aromatic interactions	4	4	4	4	8	8	6
Inter - monomers cation- π interactions	0	0	0	0	0	0	0
Inter - monomers aromatic -sulphur interactions	2	2	2	2	0	0	0

Πίνακας 24. Υπολογισμός διαμοριακών και ενδομοριακών αλληλεπιδράσεων μέσω P.I.C των θεωρητικών μοντέλων που κατασκευάστηκαν.

Comparison of the molecular interactions and geometry between the models of HU	Q9JR30	Q9KHS6	Q9Z8C7	Q46121
Intra - subunit interactions				
Intra - subunit hydrophobic interactions within 5A	76	94	112	102
Intra - subunit main chain- main chain H-bonds	128	114	122	123
Intra - subunit main chain- side chain H-bonds	33	20	25	25
Intra - subunit side chain- side chain H-bonds	12	9	9	5
Intra - subunit ionic interactions within 6A	12	16	26	16
Intra - subunit aromatic-aromatic interactions within 4,5A - 7A	12	8	10	12
Intra - subunit cation-pi interactions	0	0	0	0
Intra - subunit aromatic -sulphur interactions	0	0	5	2
Inter - subunit interactions				
Inter - monomers hydrophobic interactions within 5A	38	40	46	40
Inter - subunit main chain- main chain H-bonds	4	8	7	5
Inter - subunit main chain- side chain H-bonds	3	2	8	1
Inter - subunit side chain- side chain H-bonds	3	0	7	2
Inter - monomers ionic interactions	0	2	10	4
Inter - monomers aromatic interactions	10	6	6	6
Inter - monomers cation-pi interactions	0	0	0	0
Inter - monomers aromatic -sulphur interactions	0	0	1	0

