

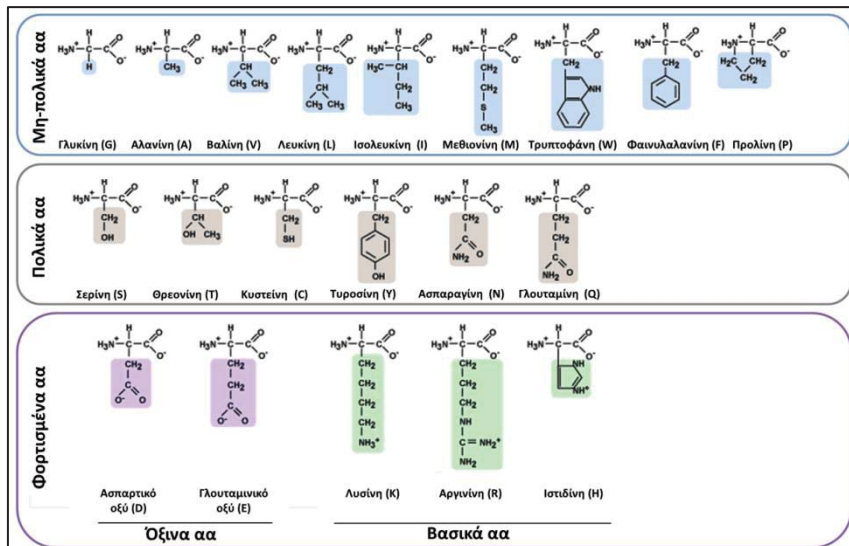
Εισαγωγή

Η κατανόηση των μοριακών μηχανισμών που διέπουν την ζωή απαιτεί την αποκωδικοποίηση των λειτουργιών και της δομής των πρωτεϊνών σε έναν οργανισμό. Δεκάδες χιλιάδες πρωτεΐνες έχουν μελετηθεί τα τελευταία χρόνια με στόχο την αποκρυπτογράφηση της τριτοταγούς δομής τους, που στην ουσία είναι αυτή που καθορίζει και τη λειτουργία τους. Οι πειραματικές μέθοδοι που χρησιμοποιούνται για τη μελέτη της δομής και δράσης των πρωτεϊνών είναι ιδιαίτερες δύσκολες, χρονοβόρες και με μεγάλο οικονομικό κόστος. Σήμερα, στον αιώνα της πληροφορικής, επιχειρείται μέσω αυτής, να μειωθεί ο χρόνος και το κόστος της πρόβλεψης της τρισδιάστατης δομής μίας πρωτεΐνης όταν η γραμμική ακολουθία των αμινοξέων που την αποτελούν, είναι γνωστή. Το ρόλο αυτό έχει αναλάβει το διεπιστημονικό πεδίο της Βιοπληροφορικής. Η Βιοπληροφορική είναι το επιστημονικό πεδίο της σύμπραξης, κυρίως των επιστημών της βιολογίας και της πληροφορικής αλλά και άλλων επιστημών όπως η στατιστική, η χημεία, η βιοχημεία, τα θεωρητικά μαθηματικά και η φυσική.

Στην άσκηση αυτή θα αναφερθούμε σε υπολογιστικές μεθόδους μέσω των οποίων μπορούμε να προβλέψουμε τη τρισδιάστατη δομή μίας πρωτεΐνης. Αρχικά, όμως, θα αναφερθούμε στα βασικά χαρακτηριστικά της δομής των πρωτεϊνών.

1. Δομή πρωτεϊνών

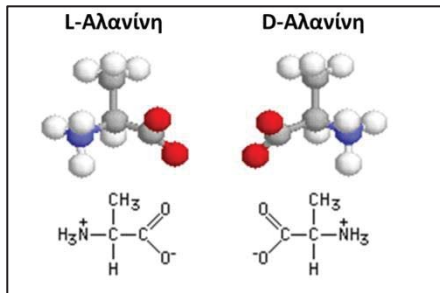
Σύμφωνα με το κεντρικό δόγμα της Βιολογίας το DNA αυτοδιπλασιάζεται και μεταγράφεται σε RNA. Το RNA στη συνέχεια μεταφράζεται σε πρωτεΐνη. Οι πρωτεΐνες είναι μακρομόρια που προκύπτουν από την ένωση α-αμινοξέων. Τα α-αμινοξέα είναι οι οργανικές ενώσεις που έχουν μία αμινομάδα (NH₂) και μία καρβοξυλομάδα (COOH) συνδεδεμένες στον ίδιο άνθρακα. Στον ίδιο άνθρακα είναι συνδεδεμένη και η πλευρική ομάδα, που είναι αυτή που καθορίζει τη μοναδικότητα του κάθε αμινοξέος. Όλες οι πρωτεΐνες των οργανισμών σχηματίζονται από το συνδυασμό 20 α-αμινοξέων που συνδέονται μεταξύ τους με έναν πεπτιδικό δεσμό. Στην Εικόνα 1 παρουσιάζεται η γενική δομή και των 20 αμινοξέων.



Εικόνα 1: Τα 20 αμινοξέα (αα) με το πλήρες όνομά τους και το αρχικό τους. Με χρώμα σκιαγραφούνται οι πλευρικές ομάδες. Με βάση την πολικότητα των πλευρικών ομάδων τα αμινοξέα ταξινομούνται σε μη-πολικά, πολικά και φορτισμένα (όξινα και βασικά).

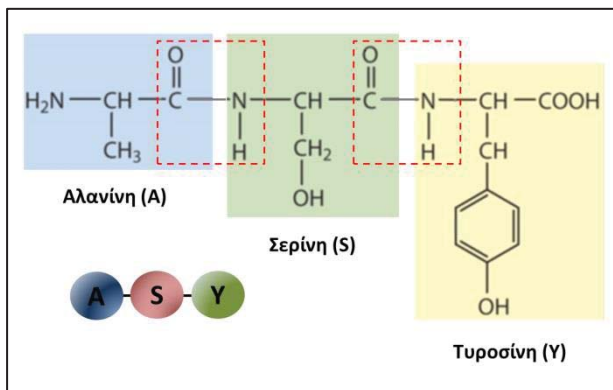
Όλα τα αμινοξέα διαθέτουν ασύμμετρο άτομο άνθρακα, δηλαδή άνθρακα που έχει τέσσερις διαφορετικούς υποκαταστάτες (εκτός από τη γλυκίνη). Σε αυτόν τον άνθρακα (α-άνθρακας), συνδέεται η αμινομάδα, το καρβοξύλιο και η πλευρική ομάδα. Έτσι αναγνωρίζονται δύο

στερεοϊσομερή για τα αμινοξέα: τα D και τα L. Τα D ισομερή έχουν την αμινομάδα δεξιά, ενώ τα L αριστερά (Εικόνα 2).



Εικόνα 2: Τα δύο ισομερή της αλανίνης. Με την αμινομάδα δεξιά (κόκκινο) η D-Αλανίνη και με την αμινομάδα αριστερά η L-Αλανίνη.

Όταν δύο αμινοξέα συνδεθούν με πεπτιδικό δεσμό, τότε σχηματίζεται ένα διπεπτίδιο. Αναλόγως, αν συνδεθεί και ένα τρίτο αμινοξύ σχηματίζεται ένα τριπεπτίδιο κ.ο.κ. (Εικόνα 3). Στο ένα άκρο του πεπτιδίου υπάρχει μία ελεύθερη αμινομάδα (NH₂) και στο άλλο μία ελεύθερη καρβοξυλομάδα (COOH). Η αλληλουχία των αμινοξέων στην πεπτιδική αλυσίδα, καθορίζει την **πρωτοταγή δομή** μιας πρωτεΐνης. Η ειδική και μοναδική σειρά με την οποία βρίσκονται οι πλευρικές αλυσίδες των αμινοξέων, οι οποίες έχουν διαφορετικές ιδιότητες μεταξύ τους, σε μια δεδομένη πρωτεΐνη, προσδιορίζει τους τρόπους που η πολυπεπτιδική αλυσίδα μπορεί να περιστραφεί και να αναδιπλωθεί. Με αυτές της περιστροφές και τις αναδιπλώσεις, κάθε πρωτεΐνη αποκτά μια καθορισμένη και χαρακτηριστική δομή η οποία την διαχωρίζει από κάθε άλλη πρωτεΐνη. Η πρωτοταγής δομή των πρωτεϊνών είναι αυτή που καθορίζει τα ανώτερα δομικά επίπεδα αυτών.



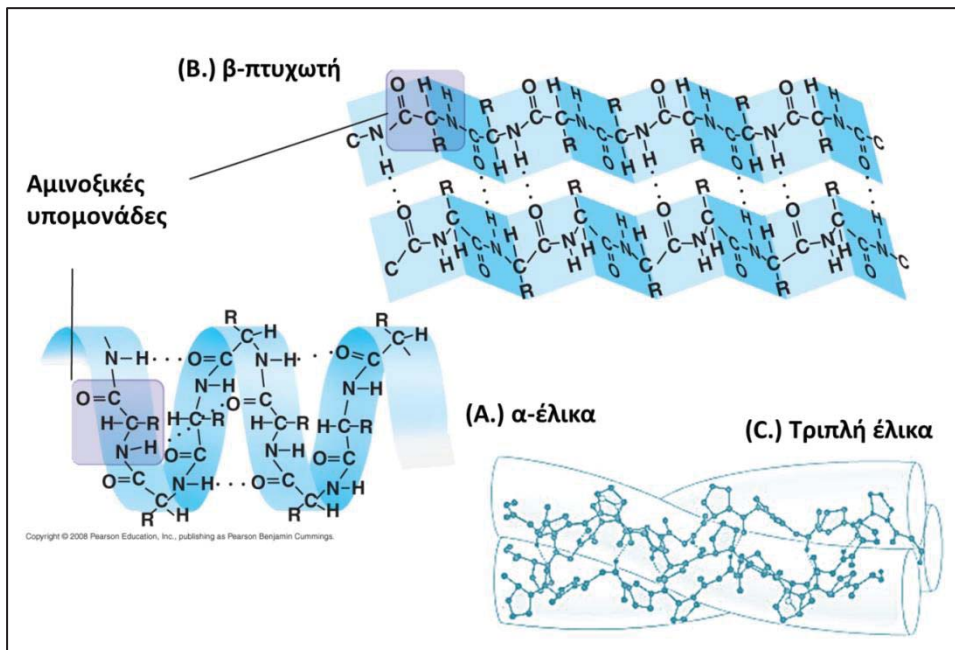
Εικόνα 3: Πρωτοταγής δομή πεπτιδικής αλυσίδας. Με κόκκινο πλαίσιο σημειώνεται ο πεπτιδικός δεσμός μεταξύ των αμινοξέων.

Η πρωτοταγής δομή κάθε πρωτεΐνης είναι μοναδική, αυτό όμως δεν ισχύει για τη **δευτεροταγή δομή** των πρωτεϊνών, καθώς διαφορετικές πρωτεΐνες μπορούν να έχουν την ίδια δευτεροταγή δομή. Η δευτεροταγής δομή των πρωτεϊνών καθορίζεται από κανονικούς και επαναλαμβανόμενους τύπους προσανατολισμών των μερών της πολυπεπτιδικής αλυσίδας. Αναγνωρίζονται τρεις κύριοι τύποι τέτοιων διαμορφώσεων.

- **α-έλικα:** Είναι μία δεξιόστροφη περιέλιξη των αμινοξέων που αποτελούν την πολυπεπτιδική αλυσίδα. Κάθε στροφή αποτελείται από 3 έως 6 αμινοξέα. Η έλικα μιας πολυπεπτιδικής αλυσίδας γύρω από τον άξονά της επιφέρει το σχηματισμό δεσμών υδρογόνου μεταξύ αμινοξέων. Ενδιάμεσα στα αμινοξέα που σχηματίζουν δεσμούς υδρογόνου, παρεμβάλλονται 3 αμινοξέα. Για παράδειγμα στην αμινοξική αλληλουχία **Lys-Ala-His-Gly-Lys-Lys-Val-Leu-Gly-Ala**, δεσμοί υδρογόνου μπορούν να δημιουργηθούν μεταξύ των αμινοξέων με το κόκκινο χρώμα. Καθώς η ικανότητα μιας πρωτεΐνης να

δημιουργεί α-έλικες εξαρτάται από τη πρωτοταγή της δομή, κάποια αμινοξέα ευνοούν τη δημιουργία έλικας ενώ άλλα την παρεμποδίζουν (Εικόνα 4A).

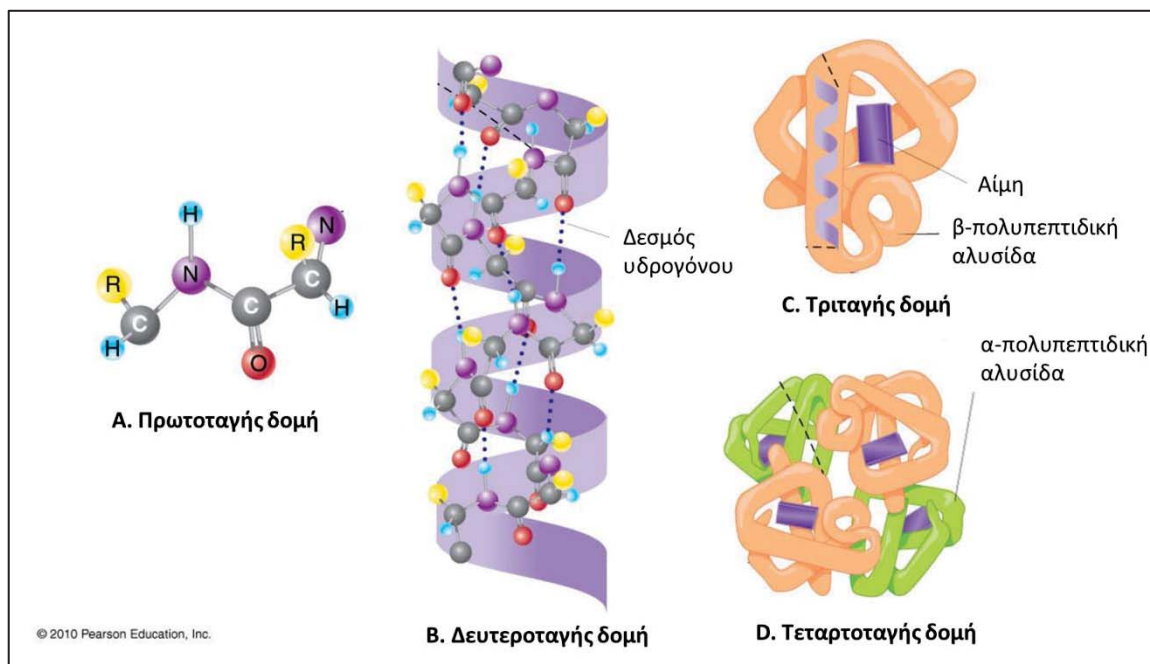
- **Δομή β-φύλλου:** Σε αυτή τη δευτεροταγή διαμόρφωση, οι πολυπεπτιδικές αλυσίδες είναι εκτεταμένες πλήρως και συνδέονται παράλληλα μέσω δεσμών υδρογόνου που ενώνουν τη μια αλυσίδα με την άλλη. Έτσι οι αλυσίδες αποκτούν πτυχωτή διαμόρφωση και για αυτό το λόγο η δομή αυτή καλείται και β-πτυχωτή (Εικόνα 4B).
- **Τριπλή έλικα:** Η δομή αυτή αποτελείται από τρεις πολυπεπτιδικές αλυσίδες οι οποίες περιστρέφονται η μία γύρω από την άλλη (κολλαγόνο). Οι αλυσίδες συνδέονται με δεσμούς υδρογόνου με αποτέλεσμα τη δημιουργία μιας ισχυρής και σταθερής δομής που δεν είναι εκτατή (Εικόνα 4C).



Εικόνα 4: Δευτεροταγής δομή πρωτεϊνών, Α. Δομή α-έλικας, Β. Δομή β-πτυχωτής, C. Δομή τριπλής έλικας.

Η πρωτοταγής δομή της πρωτεΐνης καθορίζει και την **τριτοταγή** της δομή (Εικόνα 5). Στη τριτοταγή δομή μιας πρωτεΐνης η θέση κάθε ατόμου στο μόριο, στον τρισδιάστατο χώρο, καθορίζεται επακριβώς σε σχέση με όλα τα άλλα μόρια. Η πεπτιδική αλυσίδα αναδιπλώνεται με αποτέλεσμα αμινοξέα που απέχουν πολύ μεταξύ τους στην πρωτοταγή δομή να προσεγγίζονται στην τριτοταγή δομή. Η σταθεροποίηση της τριτοταγούς δομής επιτυγχάνεται μέσω αλληλεπιδράσεων μεταξύ των πλευρικών ομάδων. Οι αλληλεπιδράσεις αυτές περιλαμβάνουν τους δισουλφιδικούς δεσμούς (ομοιοπολικός δεσμός μεταξύ των πλευρικών ομάδων της κυστεΐνης), τους δεσμούς υδρογόνου, τις υδρόφοβες αλληλεπιδράσεις (μεταξύ πλευρικών ομάδων μη πολικών αμινοξέων) και τις γέφυρες άλατος (μεταξύ αντίθετα φορτισμένων πλευρικών ομάδων δύο αμινοξέων).

Το ανώτερο επίπεδο οργάνωσης των πρωτεϊνών είναι η **τεταρτοταγής δομή** (Εικόνα 5). Τεταρτοταγή δομή έχουν οι πρωτεΐνες που αποτελούνται από δύο ή περισσότερες πολυπεπτιδικές αλυσίδες. Η τεταρτοταγής δομή καθορίζεται από δύο παράγοντες: από την διαμόρφωση, στον χώρο, των υπομονάδων της πρωτεΐνης μεταξύ τους και από τη δομή της κάθε υπομονάδας.



Εικόνα 5: Διαμόρφωση πρωτεϊνών στον χώρο. Α. Πρωτοταγής δομή, Β. Δευτεροταγής δομή, C. Τριτοταγής δομή (αιμοσφαιρίνη), D. Τεταρτοταγής δομή.

2. Δομική Βιοπληροφορική

Οι πρωτεΐνες αποτελούν κύριο συστατικό κάθε βιολογικής δραστηριότητας. Διαφορετικά μεταβολικά, κυτταρικά και δομικά στοιχεία απαιτούν την σωστή λειτουργία των πρωτεϊνών σε ένα κύτταρο. Η πρόσφατη ανάπτυξη της βιοπληροφορικής αποτελεί ορόσημο για την αποκρυπτογράφηση και αποκωδικοποίηση των χαρακτηριστικών και των λειτουργιών των γονιδίων και των πρωτεϊνών. Το αντικείμενο της δομικής βιοπληροφορικής, είναι **ο καθορισμός των τρισδιάστατων δομών, δηλαδή οι συντεταγμένες των ατόμων ενός βιολογικού μακρομορίου**. Οι διαθέσιμες τρισδιάστατες δομές είναι μια τάξη μεγέθους λιγότερες από τις διαθέσιμες νουκλεοτιδικές αλληλουχίες. Έτσι η αναζήτηση σε ένα αποθετήριο πρωτεϊνικών δομών, όπως η Protein Data Bank (PDB), για μια πρωτεϊνική αλληλουχία θα έδινε εκατοντάδες χιλιάδες νουκλεοτιδικές αλληλουχίες, ενώ θα δημιουργούσε μόνο λίγες πιθανές δομές για την ίδια πρωτεΐνη. Πολλά εργαλεία βιοπληροφορικής ανάλυσης έχουν αναπτυχθεί για την πρόβλεψη της δομής των πρωτεϊνών ώστε το έργο των βιοχημικών και βιοπληροφορικών να γίνει πιο εύκολο. Οι πληροφορίες αυτές έχουν εκχωρηθεί σε βάσεις δεδομένων που είναι ελεύθερα διαθέσιμες στους επιστήμονες. Ανάλογα με το επίπεδο των πληροφοριών που προσφέρουν, οι βάσεις δεδομένων ταξινομούνται σε διαφορετικούς τύπους. Σε αυτή την ενότητα αναφέρονται διάφορες βάσεις δεδομένων για κάθε επίπεδο πληροφοριών που προσφέρουν.

2.1. Πρωτεύουσες βάσεις δεδομένων

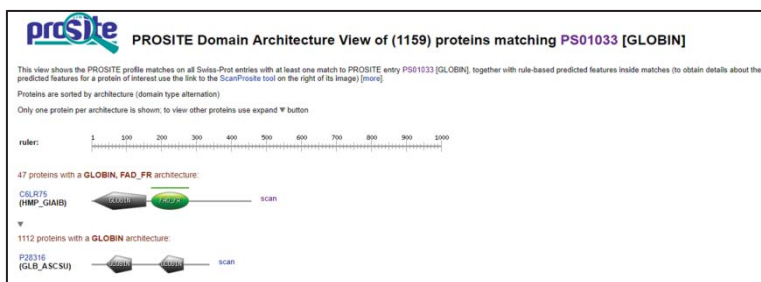
Οι πρωτεύουσες βάσεις δεδομένων πληρούνται με πειραματικά δεδομένα όπως η νουκλεοτιδική αλληλουχία, η αλληλουχία των πρωτεϊνών ή η μακρομοριακή τους δομή. Τα πειραματικά αποτελέσματα υποβάλλονται απευθείας στη βάση δεδομένων από τους ερευνητές και τα δεδομένα είναι ουσιαστικά αρχειακά. Κάποιες από τις πρωτεύουσες βάσεις δεδομένων είναι οι:

- Protein Information Resource (PIR- <https://pir.georgetown.edu/>)
- SWISS-PROT (<https://www.expasy.org/>)
- Translated EMBL (<https://www.ebi.ac.uk/Tools/st/>)
- PubMLST (<https://pubmlst.org/analysis/nrdb.shtml>)
- OWL (<http://130.88.97.239/OWL/>)

2.2. Δευτερεύουσες βάσεις δεδομένων

Οι δευτερεύουσες βάσεις δεδομένων είναι συνέπεια των αναλύσεων των ακολουθιών των πρωτογενών βάσεων δεδομένων, που βασίζονται κυρίως στην SWISS PROT. Με την χρήση αυτών των βάσεων δεδομένων, μια άγνωστη αμινοξική αλληλουχία μπορεί να αναζητηθεί σε μια βιβλιοθήκη που περιέχει τα μοτίβα διατηρημένων περιοχών ευθυγραμμισμένων αλληλουχιών που αντανακλούν σε κάποια βασική βιολογική διεργασία. Με βάση αυτά τα προκαθορισμένα χαρακτηριστικά των συντηρημένων περιοχών, η άγνωστη αμινοξική αλληλουχία μπορεί να εκχωρηθεί σε μια γνωστή οικογένεια πρωτεϊνών. Κάποιες από αυτές είναι οι παρακάτω:

- Prosite (<https://prosite.expasy.org/>) (Εικόνα 6)
- PRINTS (<http://130.88.97.239/PRINTS/index.php>)
- Blocks (<http://130.88.97.239/bioactivity/newblocksrch.html>)
- Pfam (<https://pfam.xfam.org/>)



Εικόνα 6: Αναζήτηση της αμινοξικής ακολουθίας της α-αλυσίδας της αιμοσφαιρίνης (PS01033) στη βάση δεδομένων Prosite.

2.3. Βάσεις δεδομένων ταξινόμησης δομών

Πολλές πρωτεΐνες μοιράζονται δομικές ομοιότητες, αντανακλώντας την κοινή εξελικτική τους προέλευση. Υποθέτουμε λοιπόν ότι όταν η λειτουργία μίας πρωτεΐνης διατηρείται εξελικτικά, τότε, τα δομικά χαρακτηριστικά της στα κατάλοιπα των ενεργών περιοχών της επίσης διατηρούνται. Με βάση την παρατήρηση αυτή είναι δυνατή η ταξινόμηση των πρωτεϊνών σε διαφορετικές, δομικά, οικογένειες, με βάση την αναδίπλωσή τους. Μέσω των συγκεκριμένων βάσεων δεδομένων γίνεται προσπάθεια να κατανοηθεί καλύτερα η σχέση αλληλουχίας και δομής. Η πρωτεϊνική δομή είναι πιο συντηρημένη από τις αλληλουχίες της. Η ταξινόμηση των πρωτεϊνών βασίζεται σε μια ιεραρχία επιπέδων που κατατάσσουν τις πρωτεΐνες σε μία οικογένεια, υπεροικογένεια και με βάση σαφή εξελικτική σχέση, πιθανή κοινή εξελικτική προέλευση και μεγάλη δομική ομοιότητα. Οι μέθοδοι ταξινόμησης των πρωτεϊνικών δομών βασίζονται στις μεθόδους σύγκρισης των

αλληλουχιών και στις μεθόδους σύγκρισης των δομών. Οι κύριες βάσεις δεδομένων ταξινόμησης των πρωτεϊνών είναι οι CATH (<http://www.cathdb.info/>) και SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>) (Εικόνα 7).

Protein: Sex hormone-binding globulin from Human (Homo sapiens) [TaxId: 9606]

Lineage:

- 1 Root: *scop*
- 2 Class: *All beta proteins* [48724]
- 3 Fold: *Concanavalin A-like lectin glycanase* [49898]
- 4 Superfamily: *Concanavalin A-like lectin glycanase* [49899]
- 5 *scop* family: *Lumbricin G-like module* [49944]
- 6 Protein: *Sex hormone-binding globulin* [49945]
- 7 Species: *Human (Homo sapiens)* [TaxId: 9606] [49946]

PDB Entry Domains:

- 1 *1d1a* [PDB] [View](#) [Download](#)
copied with ca. 98% identity to chain A [4231] [PDB]
- 2 *1lhu* [PDB] [View](#) [Download](#)
copied with ca. 97% identity to chain A [7964] [PDB]
- 3 *1lhu* [PDB] [View](#) [Download](#)
copied with ca. 97% identity to chain A [7966] [PDB]
- 4 *1d1d* [PDB] [View](#) [Download](#)
copied with ca. 97% identity to chain A [7235] [PDB]
- 5 *1d1f* [PDB] [View](#) [Download](#)
copied with ca. 98% identity to chain A [4222] [PDB]
- 6 *1lhu* [PDB] [View](#) [Download](#)

Εικόνα 7: Ταξινόμηση της πρωτεΐνης Sex hormone-binding globulin στη βάση δεδομένων SCOP. Αναγράφονται η οικογένεια, υπεροικογένεια, τάξη, είδος, καθώς και οι κύριες περιοχές της πρωτεΐνης.

2.4. Πολλαπλή στοίχιση πρωτεϊνικών ακολουθιών και ομολογοποίηση

Πολλές πρωτεΐνες μοιράζονται δομικά στοιχεία και ομοιότητες που αντανακλούν ίσως, σε κάποιες περιπτώσεις, την εξελικτική τους προέλευση. Οι πρωτεΐνες αυτές καλούνται ομόλογες. Όταν οι αλληλουχίες παρουσιάζουν ομοιότητες, είναι πολύ πιθανό οι ομόλογες πρωτεΐνες να παρουσιάζουν δομικές και λειτουργικές ομοιότητες, γεγονός που πολλές φορές επιτρέπει την πρόβλεψή τους (Εικόνα 8). Βάσεις δεδομένων που επιτρέπουν την εύρεση ομολογίας μεταξύ πρωτεϊνών (σε όλο ή σε περιορισμένο μήκος της πολυπεπτιδικής αλυσίδας) είναι:

- *Basic Local Alignment Search Tool* (BLAST-<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
- *ClustalW* software (<http://www.clustal.org/>)

IL-2 [Notamacropus eugenii]
gi|669633918|gb|AI123403.1

Sequence: M K V P L L S C I A L T L V L V A N G A P T L P P P T T V L Q Y L L R D L M E V Q N K L K I V S E R M K K Y E L V I P S N T S S I E N L Q C F T K E L N P V A G A L K Y E S K D A Q N I Q E Y I N N I N V T V N S L M G P E I T Q C H Y A S K M R I E G F F K E F V S V Q R F M H

BLAST Results for: AI123404:IL-2 [Sarcophilus harrisii]

| Score | Expect | Method | Identities | Positives | Gaps |
|---------------|--------|------------------------------|-------------|--------------|-----------|
| 205 bits(522) | 2e-74 | Compositional matrix adjust. | 99/139(71%) | 122/139(87%) | 2/139(1%) |

Range 1: 1 to 139

| Query | Subject | Score | Identities | Positives | Gaps |
|-------|---------|-------|--------------|--------------|-----------|
| 1 | 1 | 205 | 99/139(71%) | 122/139(87%) | 2/139(1%) |
| 60 | 61 | 118 | 100/139(72%) | 122/139(87%) | 2/139(1%) |
| 119 | 121 | 137 | 100/139(72%) | 122/139(87%) | 2/139(1%) |

Εικόνα 8: Αμινοξική ομολογία της πρωτεΐνης IL-2 του είδους *Sarcophilus harrisii* (μαρσιποφόρο της Αυστραλίας) και του είδους *Notamacropus eugenii* (μαρσιποφόρο της Αυστραλίας). Με κόκκινο σημειώνονται οι αμινοξικές αλλαγές.

2.5. Protein data bank (PDB)

Η *Protein Data Bank* (PDB- <https://www.rcsb.org/>) αποτελεί μία συλλογή δομών και δομικών δεδομένων των πρωτεϊνών, των νουκλεϊνικών οξέων και άλλων βιολογικών μακρομορίων. Η PDB είναι βασικός πόρος στον τομέα της δομικής βιολογίας και της δομικής γονιδιωματικής. Οι δομές που περιέχονται στην βάση δεδομένων της PDB κατατίθενται από ερευνητές παγκοσμίως και τα στοιχεία προέρχονται κυρίως από αποτελέσματα κρυσταλλογραφίας ακτίνων Χ, φασματοσκοπίας NMR, κρυσταλλογραφίας μικροσκοπίας και θεωρητικής μοντελοποίησης. Έτσι χρησιμεύει ως πλατφόρμα για τη συλλογή, οργάνωση και διανομή πληροφοριών που σχετίζονται με την δομή των πρωτεϊνών.

2.6. Μέθοδοι πρόβλεψης δομών

Η πρόβλεψη της δομής αποτελεί σημαντική πτυχή της σύγχρονης βιολογίας και βοηθά στην κατανόηση των λειτουργιών και των μηχανισμών των πρωτεϊνών σε τομείς όπως η ιατρική, η φαρμακολογία και η βιοτεχνολογία. Οι σύγχρονες μέθοδοι πρόβλεψης της πρωτεϊνικής δομής περιλαμβάνουν την ομόλογη μοντελοποίηση πρωτεϊνών, την αναγνώριση διπλώματος και ύφανσης καθώς και την *ab initio* ή *de novo* πρόβλεψη των δομών. Θα αναφερθούμε στη συνέχεια στις βασικές προσεγγίσεις αυτών των μεθόδων.

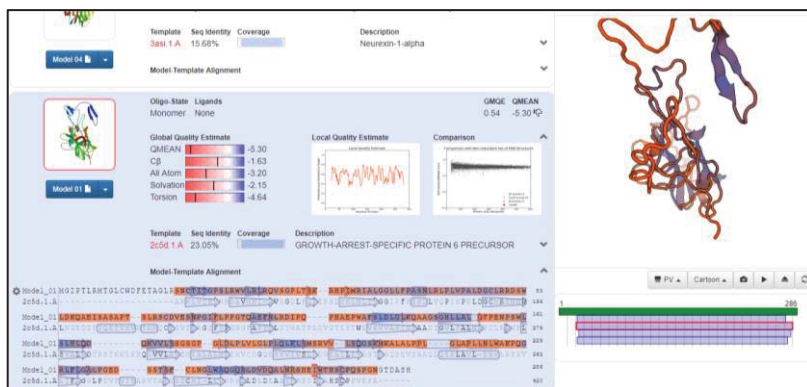
2.6.1 Ομόλογη μοντελοποίηση πρωτεϊνών

Αυτή η μέθοδος βασίζεται στην παρατήρηση ότι αλληλουχίες οι οποίες είναι ομόλογες τουλάχιστον κατά 25% υιοθετούν την ίδια δομή, ενώ αλληλουχίες με ομοιότητα κάτω από 20% μπορεί να έχουν πολύ διαφορετική δομή. Στη μέθοδο αυτή η στοίχιση των αλληλουχιών και μόνο, είναι αυτή που κατευθύνει τη δημιουργία του μοντέλου. Τα βασικά βήματα της ομόλογης μοντελοποίησης των πρωτεϊνών είναι:

- ✓ Εύρεση του πρότυπου και πραγματοποίηση της στοίχισης (BLAST και FASTA)
- ✓ Κατασκευή του σκελετού της κύριας ανθρακικής αλυσίδας
- ✓ Μοντελοποίηση των βρόχων και των πλευρικών αλυσίδων
- ✓ Βελτιστοποίηση του μοντέλου
- ✓ Έλεγχος ποιότητας του μοντέλου

Διαθέσιμα λογισμικά για την ομόλογη μοντελοποίηση πρωτεϊνών (Εικόνα 9) είναι:

- CABS (<http://biocomp.chem.uw.edu.pl/CABSfold/>)
- MODELLER (<https://salilab.org/modeller/>)
- ROSETTA (<http://rosetta.bakerlab.org/>)
- SWISS-MODEL (<http://swissmodel.expasy.org/>)
- WHATIF (<http://swift.cmbi.ru.nl/whatif/>)

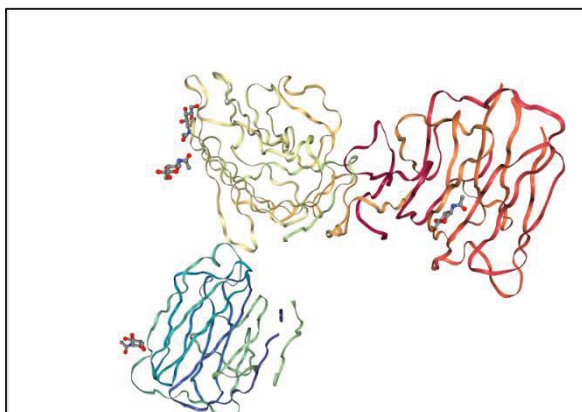


Εικόνα 9: Ομόλογη μοντελοποίηση της πρωτεΐνης Sex hormone-binding globulin στη βάση δεδομένων SWISS-MODEL.

2.6.2. Αναγνώριση διπλώματος και ύφανση

Η τεχνική αυτή χρησιμοποιείται σε περιπτώσεις κατά τις οποίες η πρωτεΐνη στόχος δεν παρουσιάζει ομοιότητα σε επίπεδο αλληλουχίας με κάποια πρωτεΐνη γνωστής δομής, αλλά μοιράζεται την ίδια αναδίπλωση με αυτές. Εδώ λοιπόν δίνεται βάση στην δομή και όχι στην αλληλουχία όπως στην ομόλογη μοντελοποίηση. Η μοντελοποίηση σε αυτή την κατηγορία γίνεται κυρίως με δύο μεθοδολογίες: Η πρώτη στηρίζεται στην μετατροπή της τρισδιάστατης δομής σε μια μονοδιάστατη αλληλουχία, ενώ στη δεύτερη χρησιμοποιείται κατευθείαν η τρισδιάστατη δομή (3D) και η ομοιότητα αξιολογείται με σύγκριση των ατομικών αποστάσεων (Εικόνα 10). Διαθέσιμα λογισμικά για αυτού του τύπου μοντελοποίησης είναι:

- PSI-BLAST (<http://www.biology.wustl.edu/gcg/psiblast.html>)
- 3D-PSSM (<http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html>)
- SUPERFAMILY (<http://supfam.org/SUPERFAMILY/>)
- GenTHREADER (<http://bioinf.cs.ucl.ac.uk/psipred/>)
- HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>)
- Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>)
- RaptorX (<http://raptorx.uchicago.edu/>)
- MUSTER (<http://zhang.bioinformatics.ku.edu/MUSTER>)
- LOMETS (<http://zhanglab.ccmb.med.umich.edu/LOMETS/>)



Εικόνα 10: Αναγνώριση διπλώματος και ύφανσης της πρωτεΐνης Sex hormone-binding globulin στη βάση δεδομένων HHpred.

2.6.3. *Ab initio* και *de novo* πρόγνωση δομής

Με τον όρο *ab initio* πρόγνωση, αναφερόμαστε στην πρόγνωση με χρήση μόνο των βασικών αρχών της φυσικής δηλαδή της αλληλεπίδρασης των ατόμων και τον υπολογισμό της ενέργειας. Έτσι σε αυτή την μεθοδολογία χρησιμοποιούνται οι αρχές των θεωρητικών υπολογισμών στη στατιστική θερμοδυναμική και την κβαντική μηχανική. Από την άλλη, ο όρος *de novo* πρόγνωση (Εικόνα 11), είναι πιο γενικός και αναφέρεται σε όλες τις μεθόδους που επιχειρούν πρόγνωση χωρίς τη χρήση προτύπου με γνωστή δομή. Σε αυτές περιλαμβάνονται οι προσομοιώσεις της μοριακής δυναμικής, προσομοιώσεις Monte Carlo, προσομοιώσεις γενετικών αλγορίθμων και μοντέλα πλέγματος. Οι μεθοδολογίες αυτές χρησιμοποιούνται όταν δεν είναι δυνατή η κατηγοριοποίηση των πρωτεϊνών με βάση την ομολογία ή το δίπλωμα.

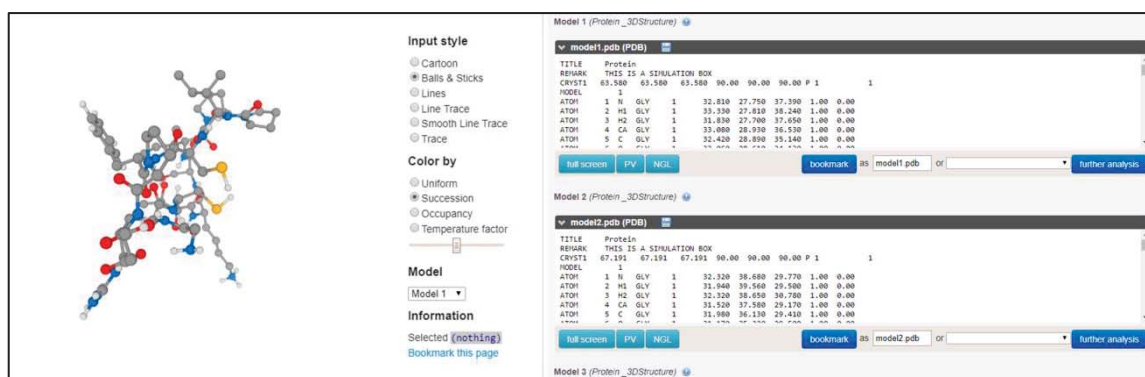
Προγράμματα για *de novo* πρόγνωση είναι τα:

- ROSETTA (<http://rosetta.bakerlab.org/>)
- I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html>)
- ePROPAINOR (<http://www.math.iitb.ac.in/epropainor>)
- PROTinfo (<http://ram.org/compbio/protinfo/>)

- QUARK (<http://zhanglab.ccmb.med.umich.edu/QUARK/>)
- CABSfold (<http://biocomp.chem.uw.edu.pl/CABSfold/>)
- PEP-FOLD (<http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD/>)
- BHAGEERATH (<http://www.scfbio-iitd.res.in/bhageerath/index.jsp>)

Προγράμματα για *ab initio* πρόγνωση είναι τα:

- ROSETTA@home (<http://boinc.bakerlab.org/rosetta/>)
- Folding@home (<http://folding.stanford.edu/>)
- FOLDit (<http://fold.it/portal/>)



Εικόνα 11: *de novo* πρόγνωση δομής της πρωτεΐνης Sex hormone-binding globulin στη βάση δεδομένων PEP-FOLD.

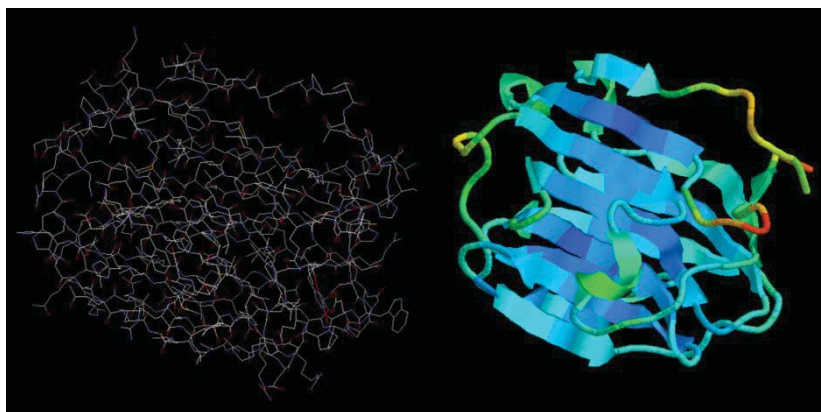
2.7. Πρόγνωση δευτεροταγούς δομής

Η αναγνώριση των πρωτεϊνικών δευτεροταγών δομικών περιοχών κατά μήκος της πρωτεϊνικής αλληλουχίας, αποτελεί προϋπόθεση για την μοντελοποίηση της αναδίπλωσής της ή της κινητικής της. Έτσι, είναι δυνατή η αναγνώριση της τρισδιάστατης τοπολογίας μίας πρωτεΐνης συγκρίνοντας τα προβλεπόμενα δευτερεύοντα δομικά της στοιχεία, με μία βάση δεδομένων που περιέχει γνωστές τοπολογίες πρωτεϊνών. Τέτοιες μέθοδοι είναι οι Chou-Fasman, GOR, PHD κ.α.

2.8. Οπτικοποίηση της δομής

Η απεικόνιση της δομής επιτρέπει την ταυτοποίηση και τον χειρισμό συγκεκριμένων δομών των πρωτεϊνών στην τρισδιάστατη όψη αυτών. Αρκετά προγράμματα έχουν αναπτυχθεί για την προβολή και αναπαράσταση της τρισδιάστατης δομής των πρωτεϊνών (Εικόνα 12). Τέτοια προγράμματα είναι τα παρακάτω:

- RCSB Simple Viewer (http://biojava.org/wiki/RCSB_Viewers>About)
- Jmol (<http://jmol.sourceforge.net/>)
- RasMol (<http://www.bernstein-plus-sons.com/software/rasmol/>)
- OpenRasMol (<http://www.openrasmol.org/>)
- CCP4mg (<http://www.ccp4.ac.uk/MG/>)
- Swiss-PDBviewer (<http://spdbv.vital-it.ch/>)
- Cn3D (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>)
- PyMol (<http://www.pymol.org/pymol>)

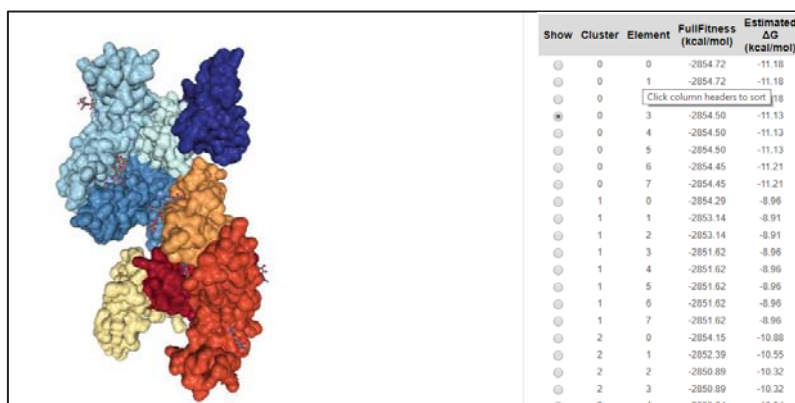


Εικόνα 12: Οπτικοποίηση της δομής της πρωτεΐνης Sex hormone-binding globulin στη βάση δεδομένων RasMol.

2.9. Αγκυροβόληση

Η αγκυροβόληση ή ο ελλιμενισμός (docking) είναι η διαδικασία με την οποία υπολογίζουμε ή προβλέπουμε τον προτιμώμενο προσανατολισμό ενός μορίου σε σχέση με ένα άλλο, όταν σχηματίζουν ένα σταθερό σύμπλοκο. Το σύμπλοκο το οποίο θα επιχειρήσουμε να μοντελοποιήσουμε με τη διαδικασία της αγκυροβόλησης μπορεί να είναι μεταξύ δύο πρωτεϊνών ή μεταξύ μιας πρωτεΐνης και ενός μικρού μορίου (ορμόνη, φάρμακο, αναστολέας, βιταμίνη κ.ο.κ.). Η αγκυροβόληση μας παρέχει πληροφορίες χρήσιμες για την κατανόηση του βιολογικού μηχανισμού λειτουργίας μίας πρωτεΐνης, τον μηχανισμό με τον οποίο αλληλεπιδρούν δυο πρωτεΐνες (ένζυμο-υπόστρωμα, υποδοχέας-προσδέτης) (Εικόνα 13), αλλά και γενικότερα στη μελέτη των πρωτεϊνικών αλληλεπιδράσεων και της τεταρτοταγούς δομής. Κάποια από τα διαθέσιμα πακέτα λογισμικού, κατάλληλα για αγκυροβόληση είναι τα παρακάτω:

- GRAMM (<http://vakser.bioinformatics.ku.edu/resources/gramm/gramm1/>)
- GRAMM-X (<http://vakser.compbio.ku.edu/resources/gramm/grammx/>)
- AutoDock (<http://autodock.scripps.edu/>)
- HADDOCK (<http://haddock.org/>)
- FTDock (<http://www.sbg.bio.ic.ac.uk/docking/ftdock.html>)
- DOT (<http://www.sdsc.edu/CCMS/DOT/>)
- ZDOCK (<http://www.umassmed.edu/zlab/>)
- ClusPro (<http://cluspro.bu.edu/>)
- SwissDock (<http://www.swissdock.ch/>)
- rDock (<http://rdock.sourceforge.net/>)
- RosettaDock (<http://rosie.rosettacommons.org/docking2>)



Εικόνα 13: Η αγκυροβόληση μεταξύ της IL-2 και του υποδοχέα της IL-2R με το λογισμικό πρόγραμμα SwissDock.

3. Βιβλιογραφία

Μπάγκος, Π., 2015. Βιοπληροφορική. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών.

John GSM, Rose C, Takeuchi S. 2011. Understanding Tools and Techniques in Protein Structure Prediction. Systems and Computational Biology-Bioinformatics and Computational Modeling: InTech.