



Βιοπληροφορική

Ενότητα 21:

Υπολογιστικός Προσδιορισμός Δομής (3/3), 1 ΔΩ

Τμήμα: **Βιοτεχνολογίας**

Όνομα καθηγητή: **Τ. Θηραίου**



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





Μαθησιακοί Στόχοι

- επισκόπηση των μεθόδων αναγνώρισης διπλώματος και απ' αρχής πρόγνωσης πρωτεϊνικών δομών.



Λέξεις Κλειδιά

- Λέξεις κλειδιά: Αναγνώριση διπλώματος, Απ' αρχής πρόγνωση δομής.
- Key words: Fold recognition, De novo protein structure prediction, Critical Assessment of protein Structure Prediction (CASP).

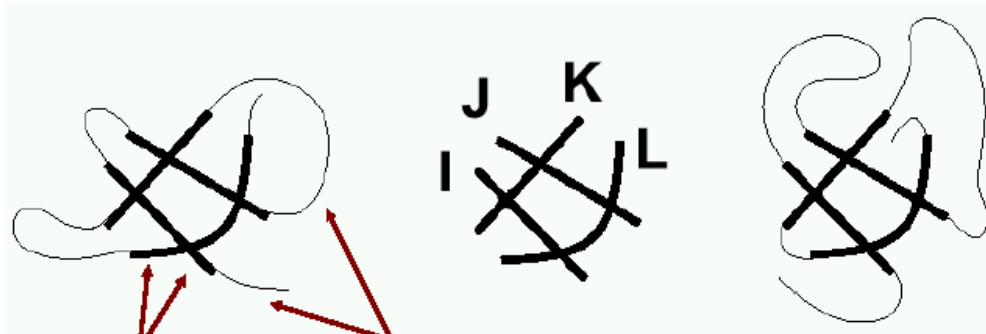


Αναγνώριση Διπλώματος 1/3

protein A threaded
on template

template

protein B threaded
on template



core secondary
structure segments

loops





Αναγνώριση Διπλώματος 2/3

- **Περιορισμένος αριθμός** τρόπων αναδίπλωσης (folds) των πειραματικά λυμένων δομών
- Αναζήτηση του τρόπου αναδίπλωσης μιας ακολουθίας άγνωστης δομής, όταν δεν υπάρχει πρωτεΐνη γνωστής δομής με υψηλή ομοιότητα σε επίπεδο ακολουθίας.
- Εφαρμογές
 - **Αναγνώριση προτύπου** για την προτυποποίηση πρωτεϊνών με ομολογία
 - Αναζήτηση στοιχείων για την **ανάθεση πρωτεϊνικής λειτουργίας**



Αναγνώριση Διπλώματος 3/3

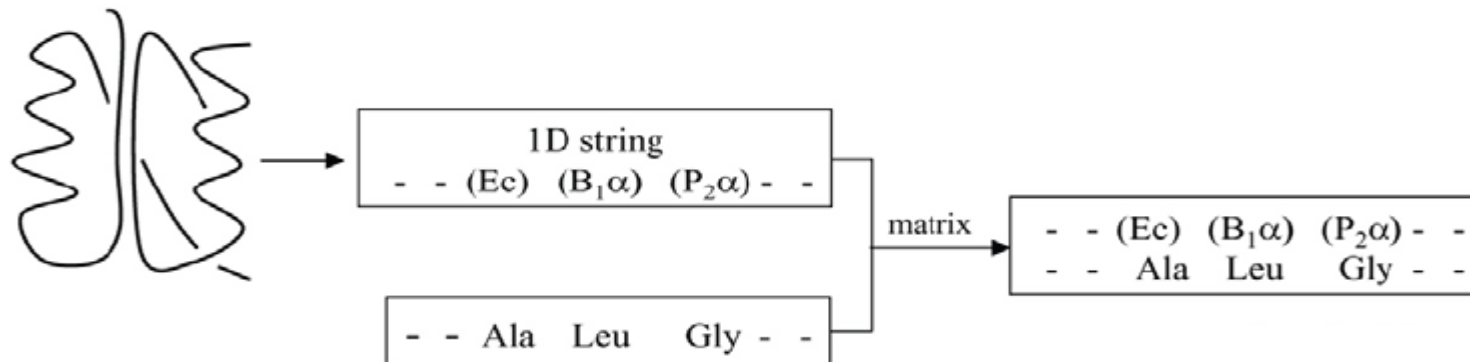
- Fold recognition / Threading
- Προβολή της ακολουθίας μιας πρωτεΐνης άγνωστης δομής σε διαφορετικές πειραματικά λυμένες υποψήφιας δομές.
- Αξιολόγηση της ικανότητας της ακολουθίας να υιοθετήσει κάθε δίπλωμα βάσει κριτηρίων.
- Κατάταξη των πιθανών τρόπων αναδίπλωσης.
 - 1D-3D profiles
 - knowledge based (pair) potentials



1D-3D profiles 1/4

● Στοίχιση ακολουθίας (1D) – δομής (3D)

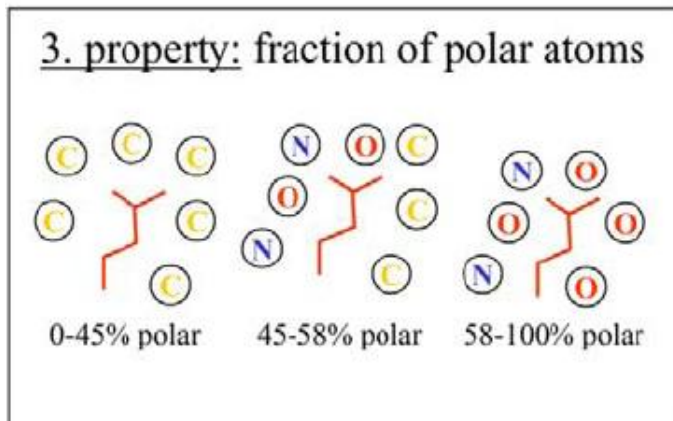
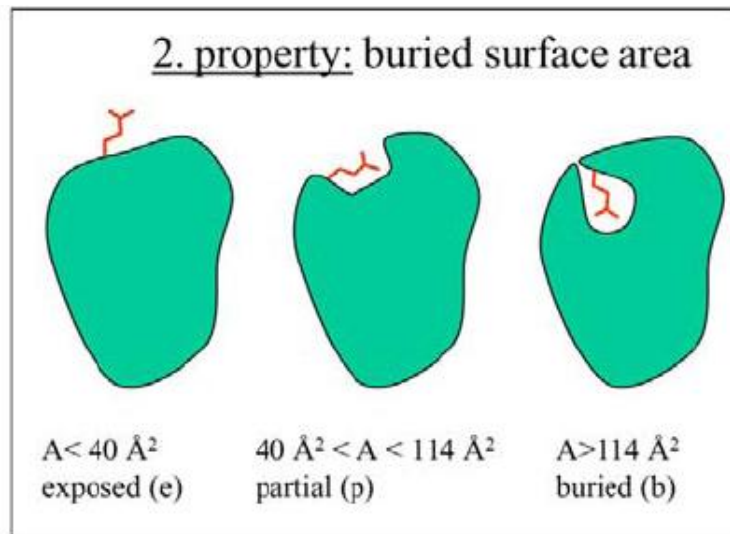
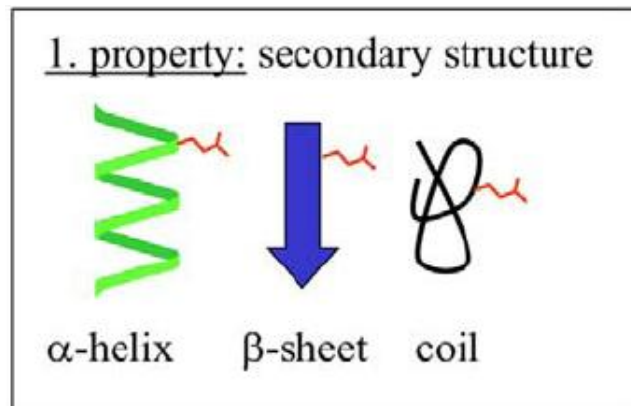
- Μετατροπή της γνωστής δομής σε μία ακολουθία από δομικά περιβάλλοντα
- Στοίχιση της αμινοξικής ακολουθίας άγνωστης δομής με την ακολουθία από δομικά περιβάλλοντα με τη χρήση πίνακα βαθμολόγησης
- Εκτίμηση της στατιστικής σημαντικότητας της στοίχισης





1D-3D profiles 2/4

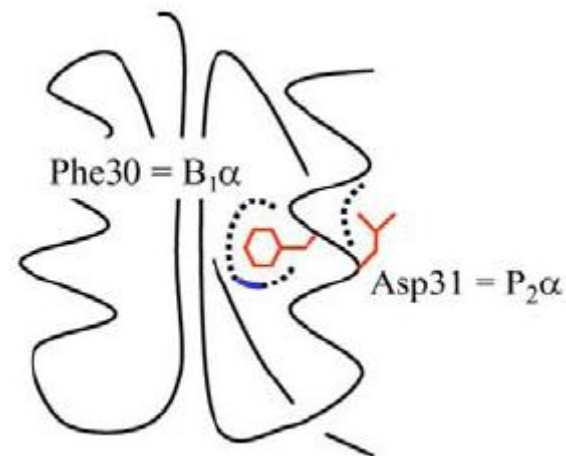
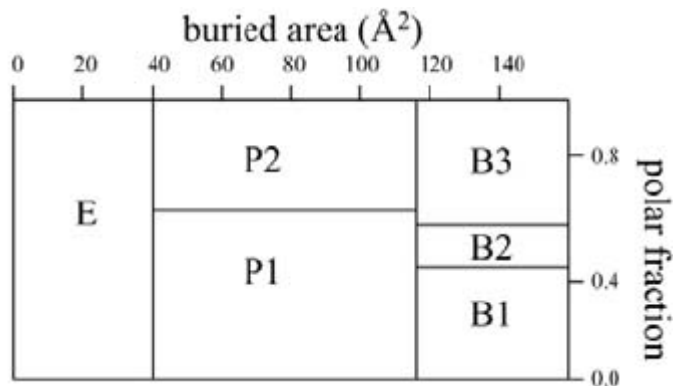
- Μετατροπή της δομής σε μία ακολουθία από **δομικά περιβάλλοντα**





1D-3D profiles 3/4

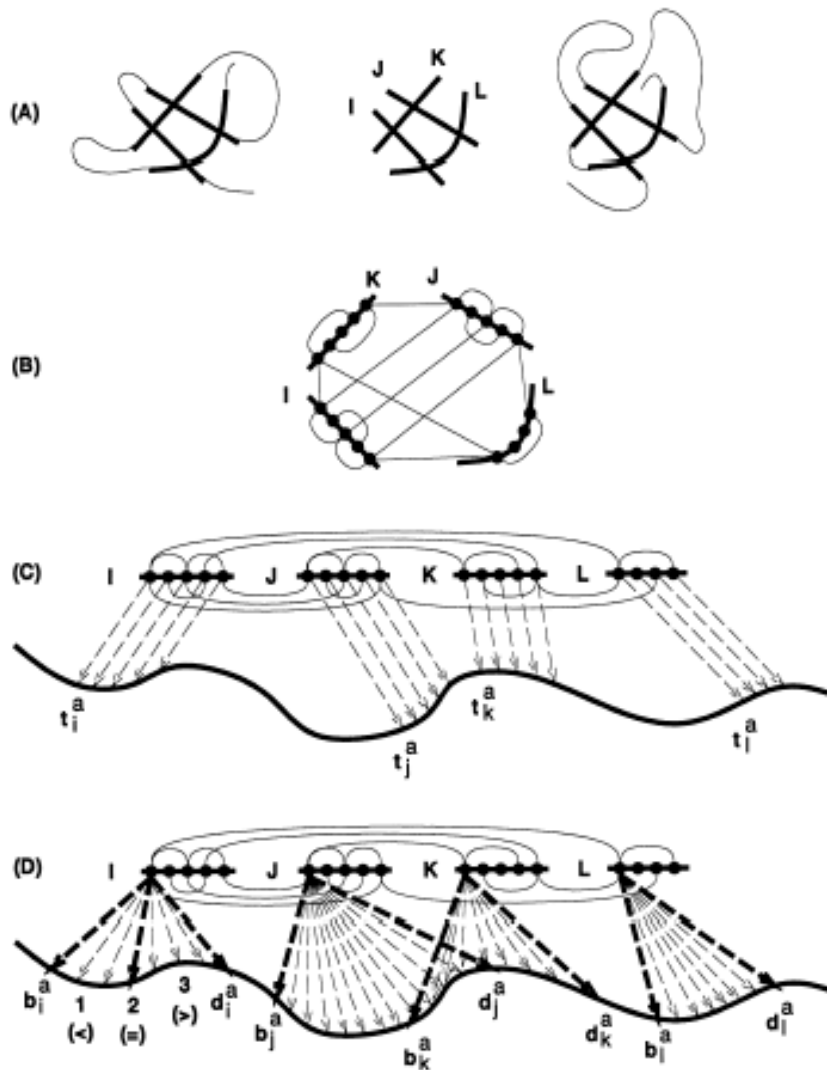
- Μετατροπή της δομής σε μία ακολουθία από δομικά περιβάλλοντα
 - Κάθε αμινοξύ προτιμά να βρίσκεται σε διαφορετικό δομικό περιβάλλον
 - Δευτεροταγής δομή
 - Στον πυρήνα ή στην επιφάνεια
 - Ποσοστό πολικών ατόμων





Knowledge Based (Pair) Potentials

1/2





Knowledge Based (Pair) Potentials

2/2

- Η ακολουθία προβάλλεται απευθείας στις γνωστές δομές χρησιμοποιώντας πληροφορίες για την **αλληλεπίδραση ζευγών αμινοξέων**.
 - π.χ.
 - X = παρατήρηση (Ala – Val σε απόσταση 20 καταλοίπων στην ακολουθία και 3Å στο χώρο)
 - Y = παρατήρηση (δύο οποιαδήποτε αμινοξέα σε απόσταση 20 καταλοίπων στην ακολουθία και 3Å στο χώρο)
 - Potential: $\log (X/Y)$



Αναγνώριση Διπλώματος

- Υβριδικές τεχνικές.
- Σύγκριση των αποτελεσμάτων από διαφορετικά προγράμματα αναγνώρισης διπλώματος.
- Phyre <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>
- MUSTER <http://zhanglab.ccmb.med.umich.edu/MUSTER/>
- RaptorX <http://raptorx.uchicago.edu/>
- SPARKS^X <http://sparks.informatics.iupui.edu/sparks-x/>



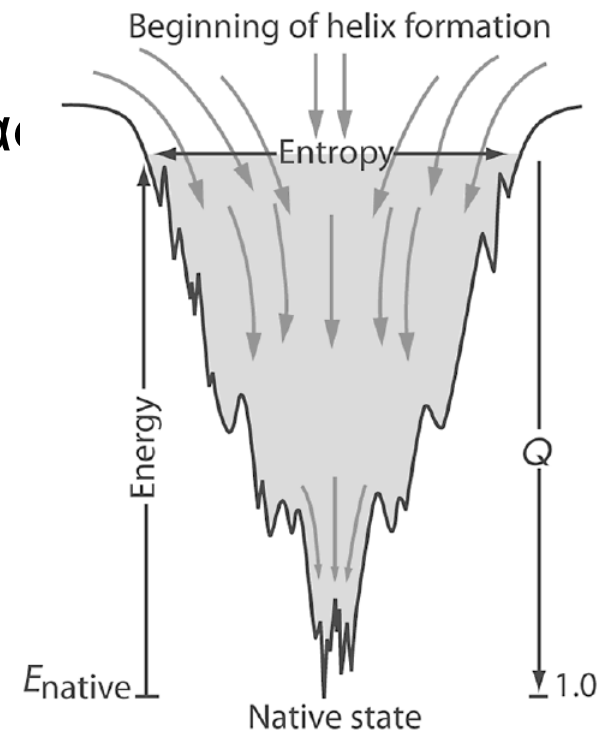
Ab Initio Prediction 1/4

- Προσδιορισμός της δομής μιας πρωτεΐνης βάσει της ακολουθίας της και των φυσικο-χημικών νόμων, **χωρίς** τη χρήση πρότυπων δομών.
- Στην πράξη, όλες οι μέθοδοι στηρίζονται σε **εμπειρικά δεδομένα από λυμένες πρωτεΐνες**.
 - Πεδία δυνάμεων
 - Εμπειρικές ενέργειες αλληλεπίδρασης
 - Δεδομένα εκπαίδευσης
 - Δομικά στοιχεία (fragment structures)



Ab Initio Prediction 2/4

- Προσομοίωση της αναδίπλωσης της πολυπεπτιδικής αλυσίδας
 - Μοριακή Δυναμική με πλήρη αναπαράσταση της πρωτεΐνης και του διαλύτη
 - διαμόρφωση ελάχιστης ενέργειας
 - θεωρητικά ζητήματα
 - υπολογιστική ισχύς / χρόνος
 - αποτελέσματα

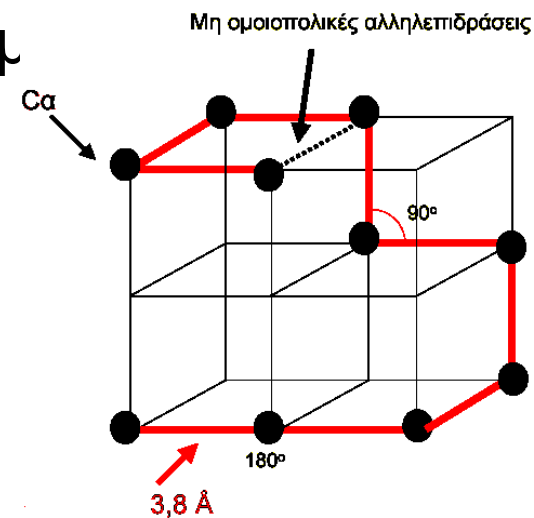




Ab Initio Prediction 3/4

● Απλοποιημένα μοντέλα

- κύρια αλυσίδα ή $C\alpha$
- τρισδιάστατο πλέγμα / περιορισμός βαθμών ελευθερίας
- εμπειρικές ενέργειες αλληλεπίδρασης
- περιορισμός του χώρου των πιθανών διαμορφώσεων και εν συνεχεία λεπτομοντέλα





Ab Initio Prediction 4/4

- ROSETTA <http://www.rosettacommons.org/>
 - Οι πιθανές διαμορφώσεις μικρών πεπτιδικών τμημάτων (3 - 9 κατάλοιπα) αντιπροσωπεύονται στις λυμένες δομές της PDB.
 - **Δημιουργία βιβλιοθήκης "δομικών στοιχείων"**: πιθανών διαμορφώσεων κάθε τμήματος ακολουθίας.

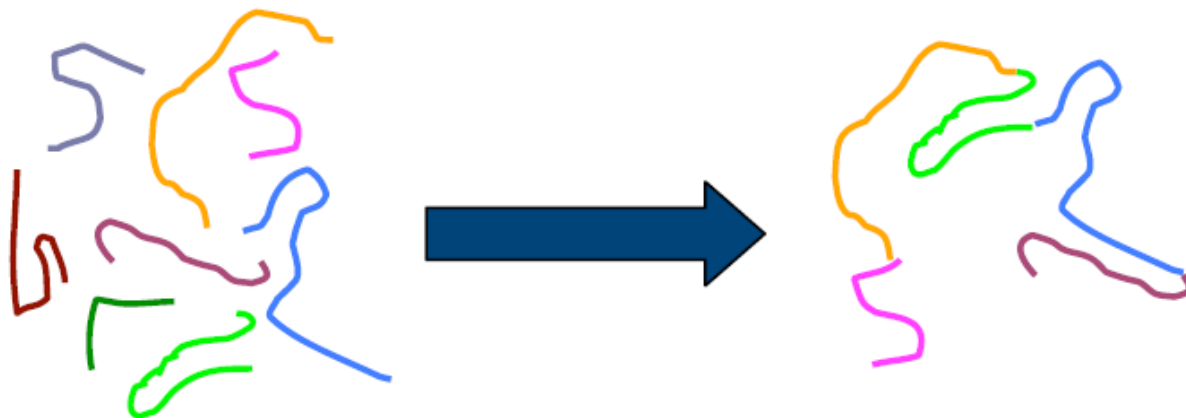
LSERTVARS





ROSETTA

- **Συνδυασμός δομικών στοιχείων** για τη δημιουργία της πρωτεϊνικής δομής.
 - Μέθοδος Monte Carlo
 - Δημιουργία πολλών μοντέλων
 - Ομαδοποίηση των μοντέλων (RMSD)
 - Επιλογή εκείνων που ανήκουν στις πολυπληθέστερες ομάδες

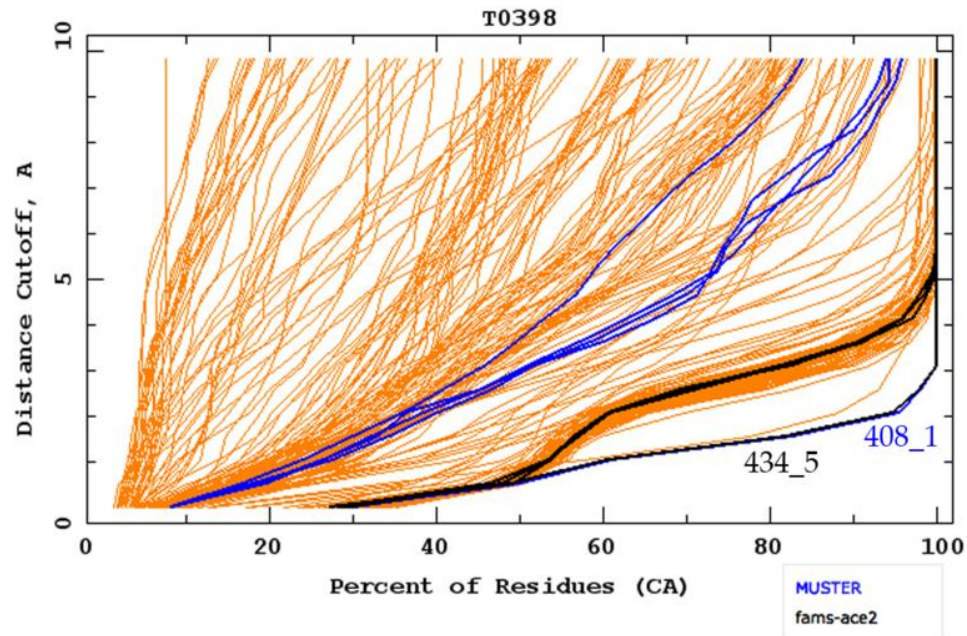
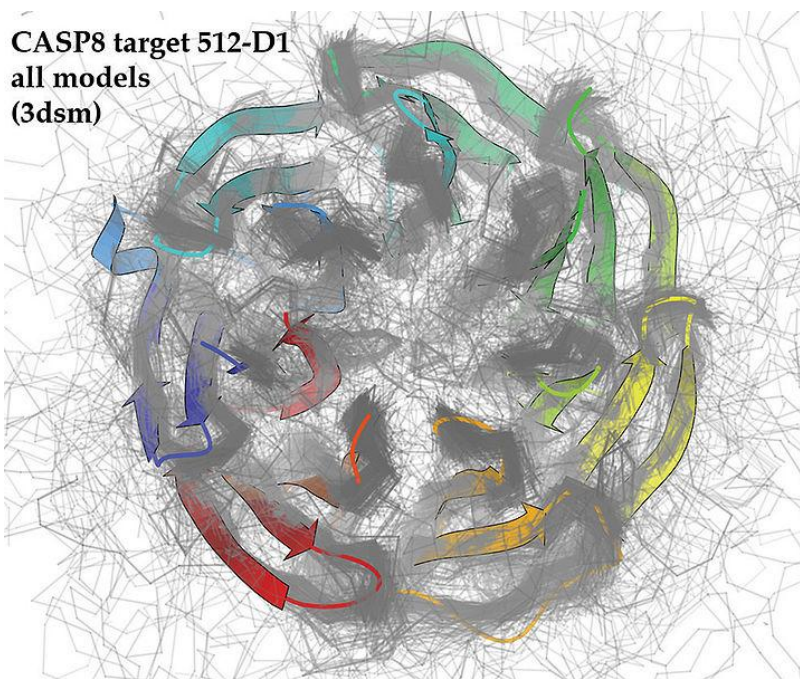




CASP 1/3

- Critical Assessment of protein Structure Prediction (CASP)

CASP8 target 512-D1
all models
(3dsm)





CASP 2/3

- Critical Assessment of protein Structure Prediction (CASP)
- Εκτίμηση ακρίβειας πρόγνωσης και περιορισμών των μεθόδων υπολογιστικού προσδιορισμού δομών πρωτεϊνών.
 - Ακολουθίες πρωτεϊνών, η πειραματικά προσδιορισμένη δομή των οποίων πρόκειται να δημοσιευθεί μετά το διαγωνισμό.
 - Πρόβλεψη της δομής των πρωτεϊνών από διάφορες ερευνητικές ομάδες.
 - Σύγκριση των μοντέλων με τα πειραματικά δεδομένα και αξιολόγηση της ποιότητάς τους από έμπειρους κριτές με τη βοήθεια προγραμμάτων ανάλυσης δομών.



CASP 3/3

Alexey Murzin (Proteins Volume 45, Issue S5, 2001. Pages:76-85)

- "In 1996, in CASP2, we presented a semimanual approach to the prediction of protein structure that was aimed at the recognition of probable distant homology, where it existed, between a given target protein and a protein of known structure (Murzin and Bateman, [Proteins 1997; Suppl 1:105-112]). Central to our method was the **knowledge of all known structural and probable evolutionary relationships among proteins** of known structure classified in the SCOP database (Murzin et al., J Mol Biol 1995;247:536-540). It was demonstrated that a knowledge-based approach could **compete successfully with the best computational methods** of the time in the correct recognition of the target protein fold."



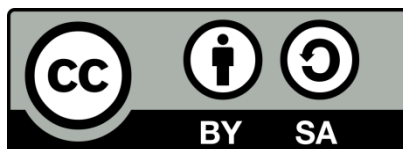
Βιβλιογραφία

- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2nd edition (March 12, 2013)
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2nd edition (May 4, 2009)
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3rd edition (October 29, 2004)
- Jenny Gu, Philip E. Bourne, "Structural Bioinformatics", Wiley-Blackwell; 2nd edition (March 16, 2009)



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.





Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



Σημείωμα Αναφοράς

Copyright Γεωπονικό Πανεπιστήμιο Αθηνών 2015. Τμήμα Βιοτεχνολογίας, Θηραίου Τριάς. «Βιοπληροφορική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://mediasrv.aua.gr/eclass/courses/OCDB100/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
 - το Σημείωμα Αδειοδότησης
 - τη δήλωση Διατήρησης Σημειωμάτων
 - το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)
- μαζί με τους συνοδευόμενους υπερσυνδέσμους.