

Παρουσίαση με θέμα

“Πρόβλεψη της 3D δομής πρωτεϊνών από την ακολουθία των αμινοξέων της”

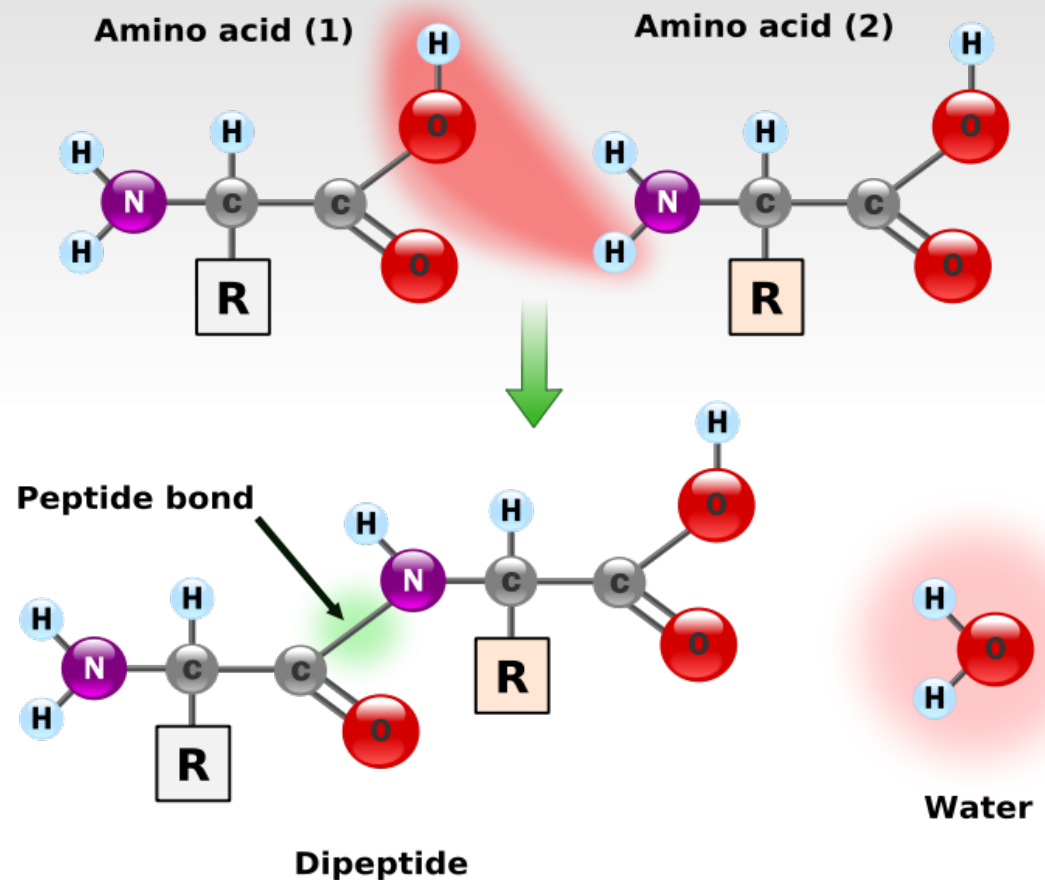
Βαλκανάς Γιώργος

Ορισμός του προβλήματος

- Πρωτεΐνες: δομικές & λειτουργικές μονάδες
- Οι λειτουργίες τους εξαρτώνται από την τρισδιάστατή (3D) δομή τους
- Αλλά από τι καθορίζεται η 3D δομή?
 - Αλληλουχία αμινοξέων
 - Εξωτερικοί παράγοντες
- Θέλουμε: την ανακάλυψη της 3D δομής γνωρίζοντας την αλληλουχία τους

Λίγη Βιολογία.. (1)

- Αμινοξέα:
 - Δομική μονάδα πρωτεϊνών
 - Αμινομάδα
 - Καρβοξυλομάδα
 - Ομάδα *R*
- Πολυπεπτιδικές αλυσίδες
- Λειτουργίες..



Λίγη Βιολογία.. (2)

- **Πρωτοταγής δομή:** η ακολουθία των αμινοξέων που αποτελούν την πρωτεΐνη
- **Δευτεροταγής δομή:** αναδιπλώσεις γειτονικών αμινοξέων στο χώρο
- **Τριτοταγής δομή:** Επιπλέον αναδιπλώσεις, ανεξαρτήτως γειτονιάς
- **Τεταρτοταγής δομή:** Τριτοταγείς δομές για πρωτεΐνες με πολλές αλυσίδες..
- **Ενεργό Μέρος:** Το τμήμα της πρωτεΐνης που επιτελεί τη λειτουργία της..
- **ΠΡΟΣΟΧΗ:** Τριτοταγής \neq 3D

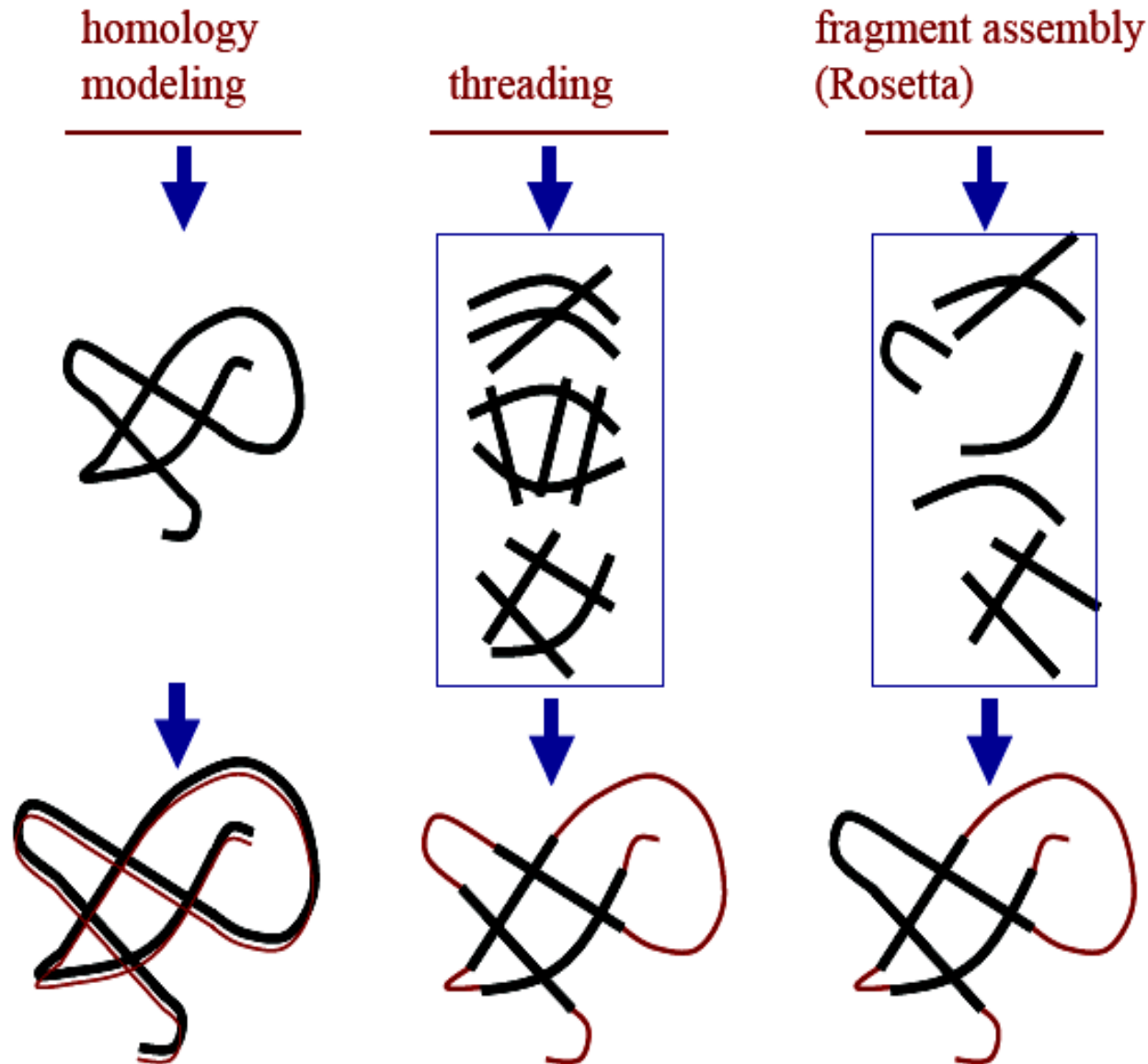
Γιατί μας ενδιαφέρει?

- *Anfinsen, 1973*: “Η λειτουργία της πρωτεΐνης καθορίζεται από την τριτοταγή δομή της”
 - Η τριτοταγής δομή καθορίζει το ενεργό μέρος της
- Οφέλη??
 - Κατασκευή νέων πρωτεϊνών – φαρμάκων
 - Εξελικτική πορεία των ειδών
 - Τι γίνεται με τις μεταλλάξεις? (Alzheimer, mad cow syndrom..)

Γιατί τη Μηχανική Μάθηση?

- Υπάρχουν φυσικές μέθοδοι?
 - NMR, X-Ray πρισματοσκοπία
 - Κοστίζουν
 - Είναι χρονοβόρες
- Άλλος τρόπος?
 - Υπολογισμός όλων των πιθανών καταστάσεων
 - NP-hard
 - 100 αμινοξέα, 3 πιθανές καταστάσεις $\Rightarrow 100^3$ συνδυασμοί
- Οι αναδιπλώσεις καθορίζονται και από παράγοντες του περιβάλλοντος (chaperons)

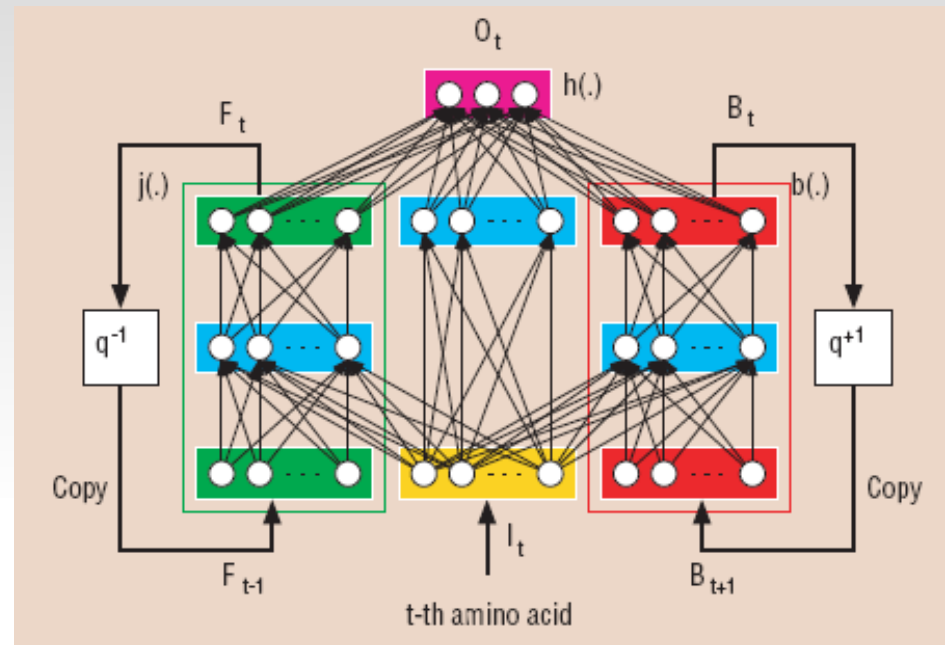
Τεχνικές που υπάρχουν..



- Μοντελοποίηση ομολόγων
- Threading
- Lego
- Ab initio
- Machine Learning

Neural Networks..

- Γιατί?
 - Μεγάλη ικανότητα μάθησης
 - Ικανοποιητικά αποτελέσματα πρόβλεψης
 - Μη γραμμικά δεδομένα
- nnPredict (2 επίπεδα: input, output)



BRNN

Άλλοι τρόποι

Δύο ΤΝΔ:

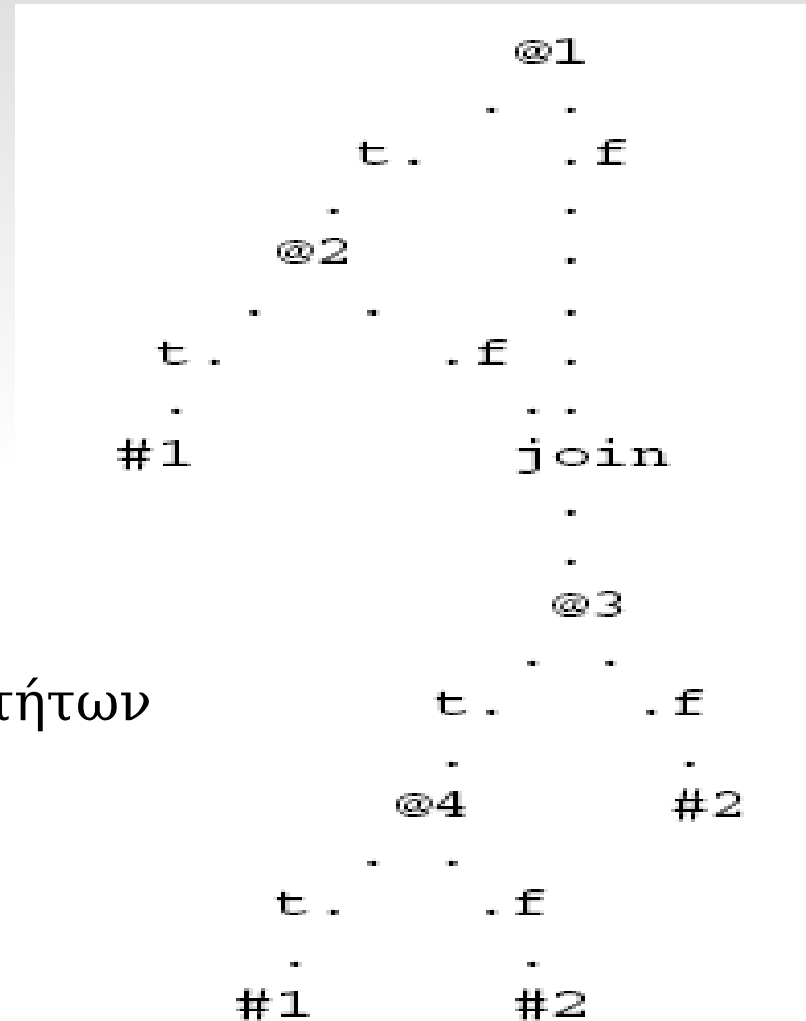
1ο) Πρόβλεψη
2ο) Επαλήθευση

Support Vector Machines..

- Διάφορες αρχιτεκτονικές
 - Χρήση ενός SVM
 - Χρήση 2 SVM (διαφορετικά επίπεδα):
 - 1ο: Γειτονικά αμινοξέα
 - 2ο: Συνολική δομή πρωτεΐνης
- Διαφορετικά kernel functions
 - Χρήση μόνο των ομάδων R
 - Χρήση φυσικοχημικών ιδιοτήτων
- Πώς συνδυάζω τα SVM classifiers?
 - one-versus-one
 - one-versus-rest
- Και ποιο επιλέγω? (vote, SVM_MAX_D, SVM_tree)

Άλλες τεχνικές..

- Decision trees (1992!!!)
 - Ίδιος τρόπος εκπαίδευσης
 - Παράγονται κατανοητοί κανόνες
 - Αντιληπτό αποτέλεσμα
 - Decision Graphs: DT με Joint κόμβους
- Hidden Markov Models
 - Πιθανοτικό μοντέλο, χωρίς μνήμη
 - Υπολογισμός των κατάλληλων πιθανοτήτων
 - trial and error..
- Υβριδικά:
 - Εκπαίδευση με SVM, DT για κανόνες
 - NN όπου κάθε κρυφό επίπεδο στηρίζεται σε SVM
 - Bagging-like, για μεγάλα σύνολα



Ωραία.. και πώς το κάνω?

- Παράθυρο W, γύρω από κάθε αμινοξύ, με κέντρο το αμινοξύ που ελέγχω
 - Ολισθαίνον
 - Στατικό
- Φυσικοχημικές ιδιότητες
 - υδροφιλία, υδροφοβία
 - πολικότητα
- Χρήση στοιχείων ομόλογων πρωτεϊνών (multiple sequence alignment).
- Εμπειρικά στοιχεία από μελέτες των “ειδικών”

$$Q_3 = \frac{N_{\alpha} + N_{\beta} + N_{\epsilon}}{N}$$

Ευχαριστώ!

Folding@Home

Distributed Computing - <http://folding.stanford.edu>

Donator:
lebiathan

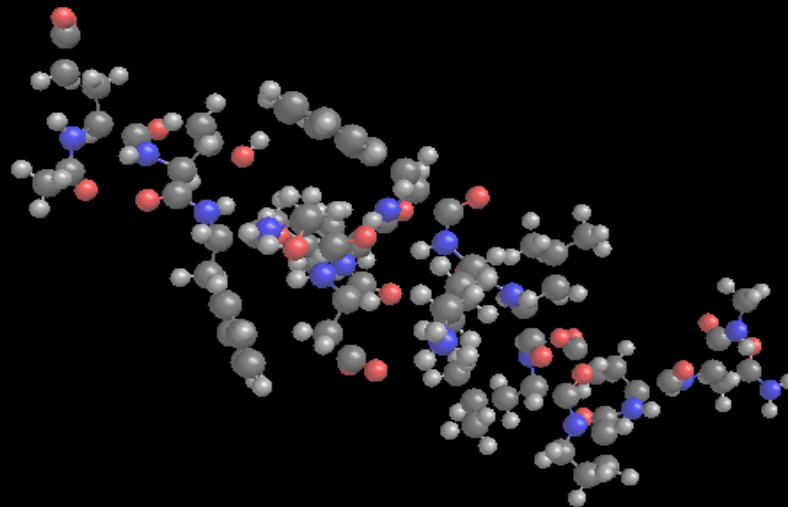
Team Number:
0

Finished WUs:
none

Working on:
p891_p53wtpeptide_tij
AMBER core

Frames completed:
64/500
2m:33s/frame
67/500 (estimate)

WU End:
09:47 Mon 14 Jan 08
00d:18h:19m:28s



Working ...

Απορίες??