

Βιοπληροφορική

Διάλεξη 10η :

Εφαρμογές στη Δομική Βιολογία.

Εισαγωγή

Το θεμελιώδες πρόβλημα είναι το folding problem. Οι απόπειρες επίλυσης του με καθαρά υπολογιστικές μεθόδους έχουν γίνει γνωστές με τον τίτλο "structure prediction methods". Αυτές οι μέθοδοι ανήκουν τυπικά στον κλάδο της υπολογιστικής βιολογίας και όχι της βιοπληροφορικής, με τη βιοπληροφορική να ασχολείται με φαινομενικά περιφερειακά προβλήματα όπως

- Πρόβλεψη δευτεροταγούς δομής,
- Πρόβλεψη διαμεμβρανικών τμημάτων,
- Πρόβλεψη coiled-coils, κοκ.

Εισαγωγή

Αποτέλεσμα αυτού του διαχωρισμού είναι ότι οι εφαρμογές της βιοπληροφορικής στη δομική βιολογία χαρακτηρίζονται από ένα λογικό παράδοξο : Οι διάφοροι διαθέσιμοι αλγόριθμοι προσπαθούν με βάση τις αλληλουχίες να απαντήσουν ερωτήματα που αφορούν τη δομή των μακρομορίων (όπως π.χ. θέση και έκταση στοιχείων δευτεροταγούς δομής), όταν το καθ'αυτό πρόβλημα (δηλαδή το folding problem) είναι άλυτο. Το αποτέλεσμα είναι ότι όλες αυτές οι μέθοδοι δίνουν απαντήσεις με σημαντικά σφάλματα των οποίων η έκταση είναι εκ προοιμίου άγνωστη.

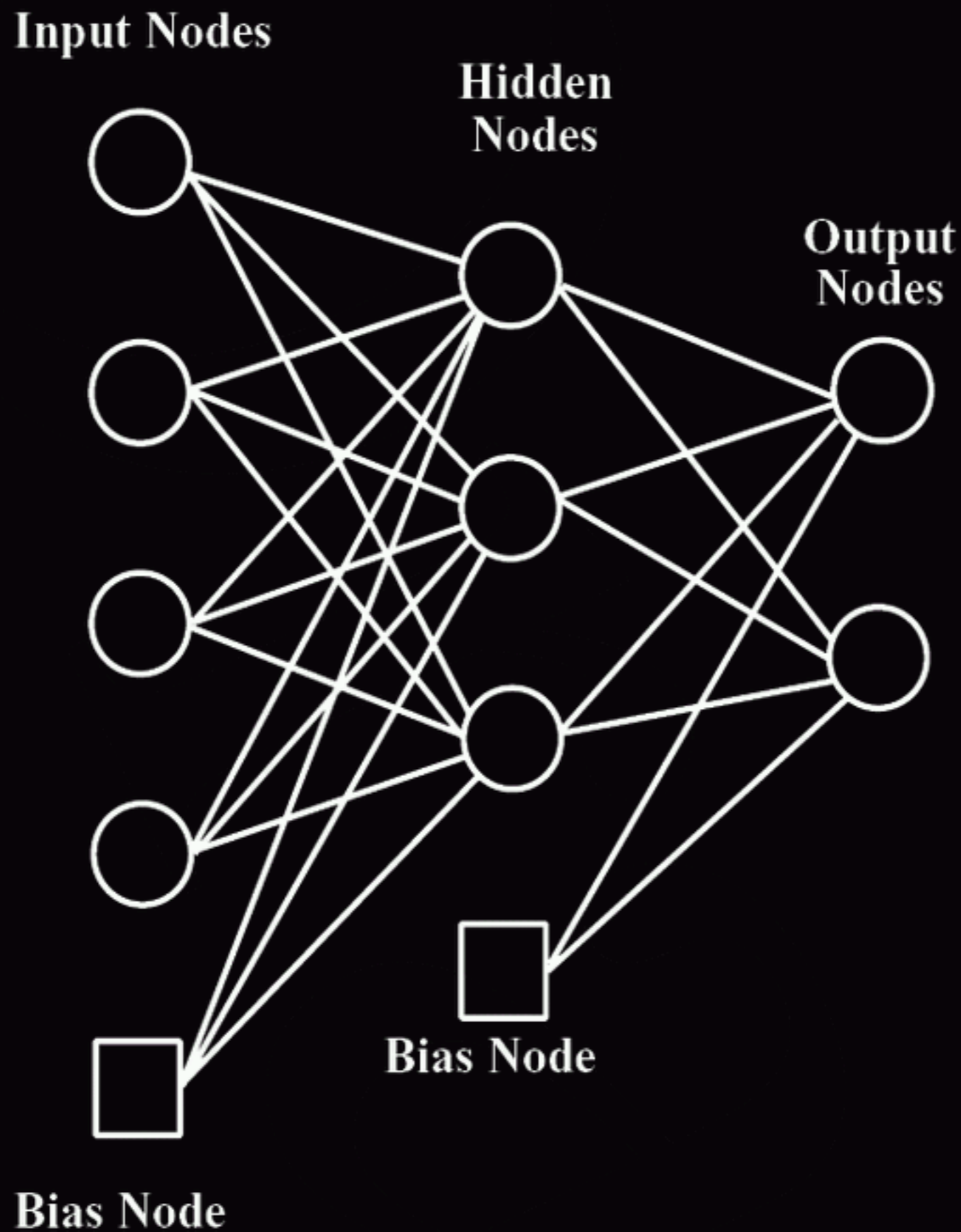
Εισαγωγή

Παρ'όλα αυτά, το σχετικά υψηλό μέσο ποσοστό επιτυχίας μερικών από αυτές τις μεθόδους υποδηλώνει και αποδεικνύει την ύπαρξη μοτίβων και σχέσεων που συνδέουν την πρωτοταγή με τη δευτεροταγή και υπερδευτεροταγή δομή των μακρομορίων χωρίς άμεση αναφορά στην τριτοταγή δομή. Για παράδειγμα, μερικές από τις μεθόδους πρόβλεψης δευτεροταγούς δομής έχουν μέσα ποσοστά επιτυχίας της τάξης του 70-75%. Τέτοια υψηλά ποσοστά όχι μόνο δείχνουν ότι είναι εφικτή η πρόβλεψη της δευτεροταγούς δομής απουσία γνώσης της τριτοταγούς, αλλά ενισχύουν και την άποψη του πρώιμου σχηματισμού των στοιχείων δευτεροταγούς δομής κατά τη διάρκεια της πρωτεϊνικής αναδίπλωσης.

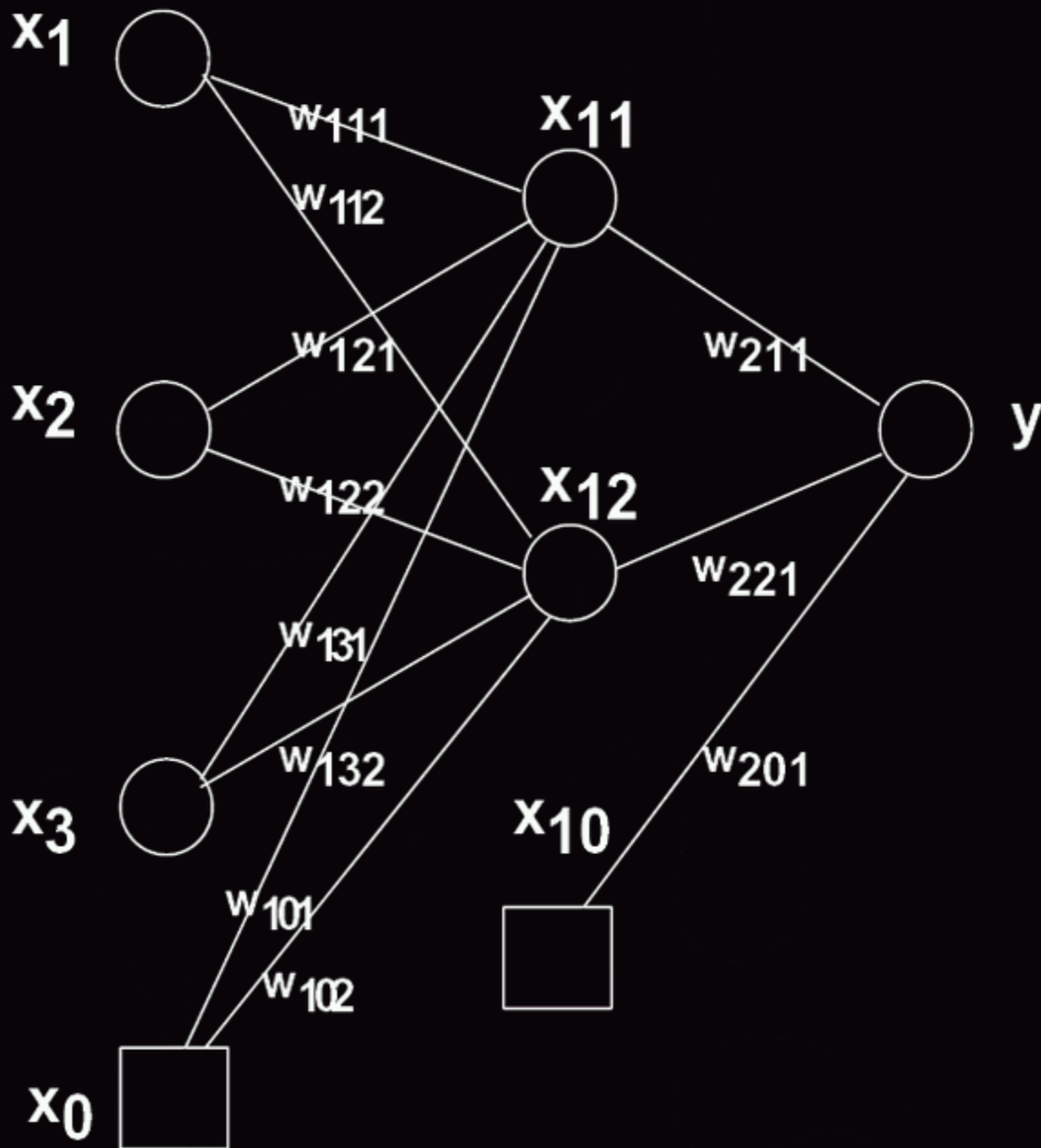
Πρόβλεψη δευτεροταγούς δομής

Οι αρχικές μέθοδοι πρόβλεψης δευτεροταγούς δομής στηρίζονταν κυρίως σε στατιστικά προσδιορισμένες "προτιμήσεις" καταλοίπων για τις τρεις καταστάσεις (α-έλικα, β-κλώνος, θηλιά) και χρησιμοποιούσαν μόνο την προς προσδιορισμό αλληλουχία (αγνοώντας την πληροφορία που περιέχεται σε στοιχίσεις με άλλες ομόλογες αλληλουχίες). Το αποτέλεσμα ήταν χαμηλά ποσοστά επιτυχίας της τάξεως του 60%. Σχεδόν όλες οι σύγχρονες μέθοδοι χρησιμοποιούν νευρωνικά δίκτυα που τόσο στο στάδιο της εκπαίδευσης όσο και της πρόβλεψης στηρίζονται σε multiple sequence alignments.

Τεχνητά νευρωνικά δίκτυα

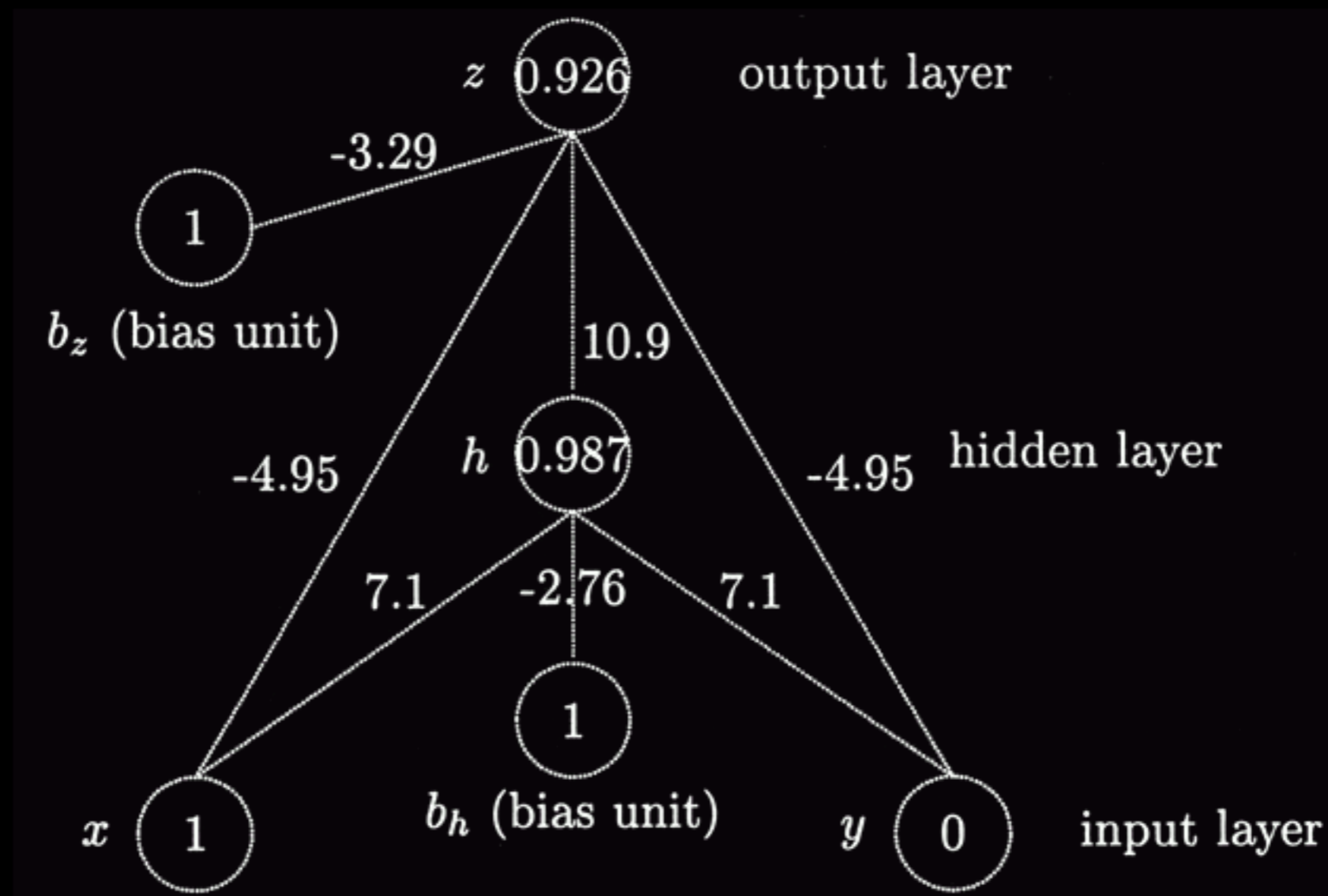


Τεχνητά νευρωνικά δίκτυα



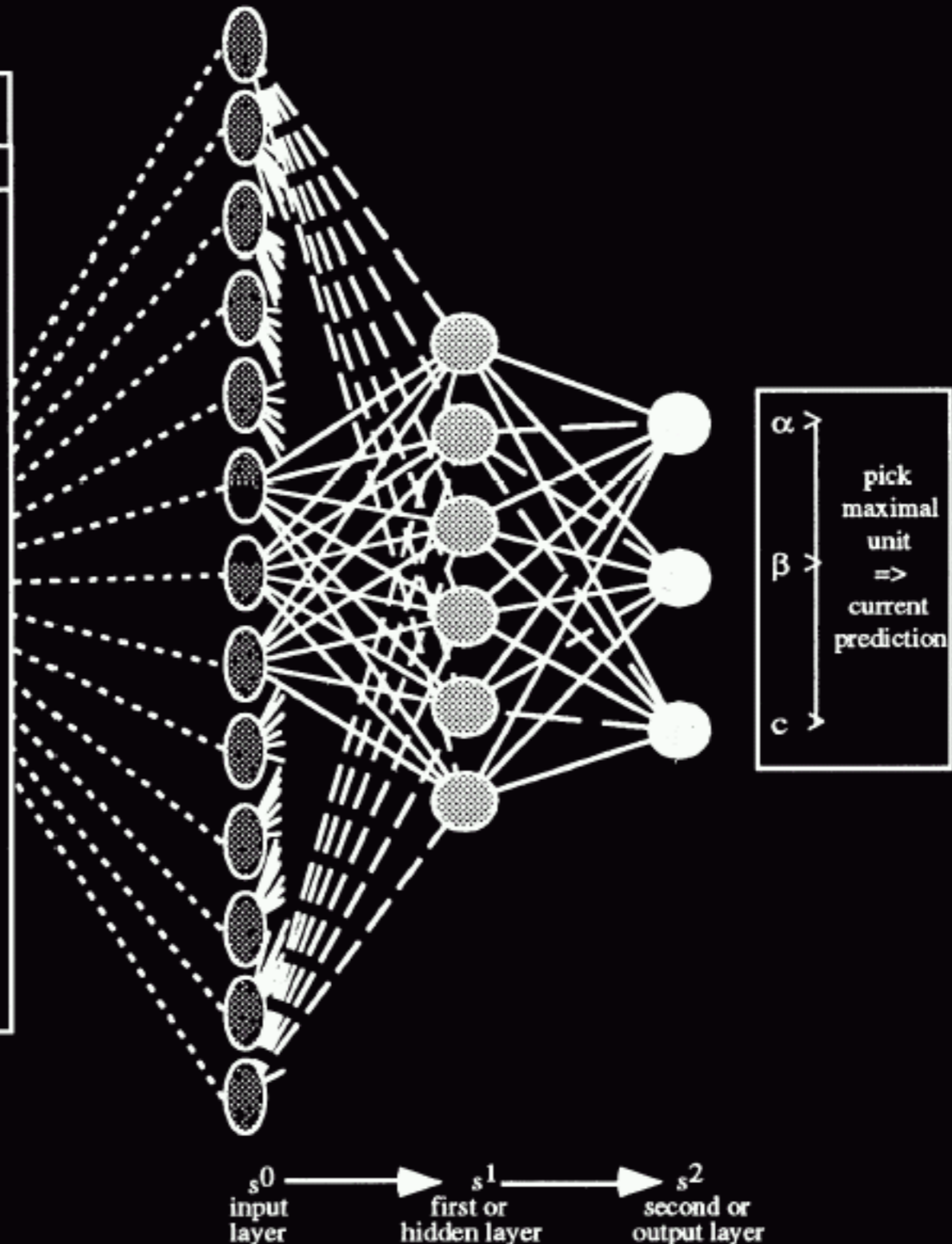
Τεχνητά νευρωνικά δίκτυα

Παράδειγμα : xor

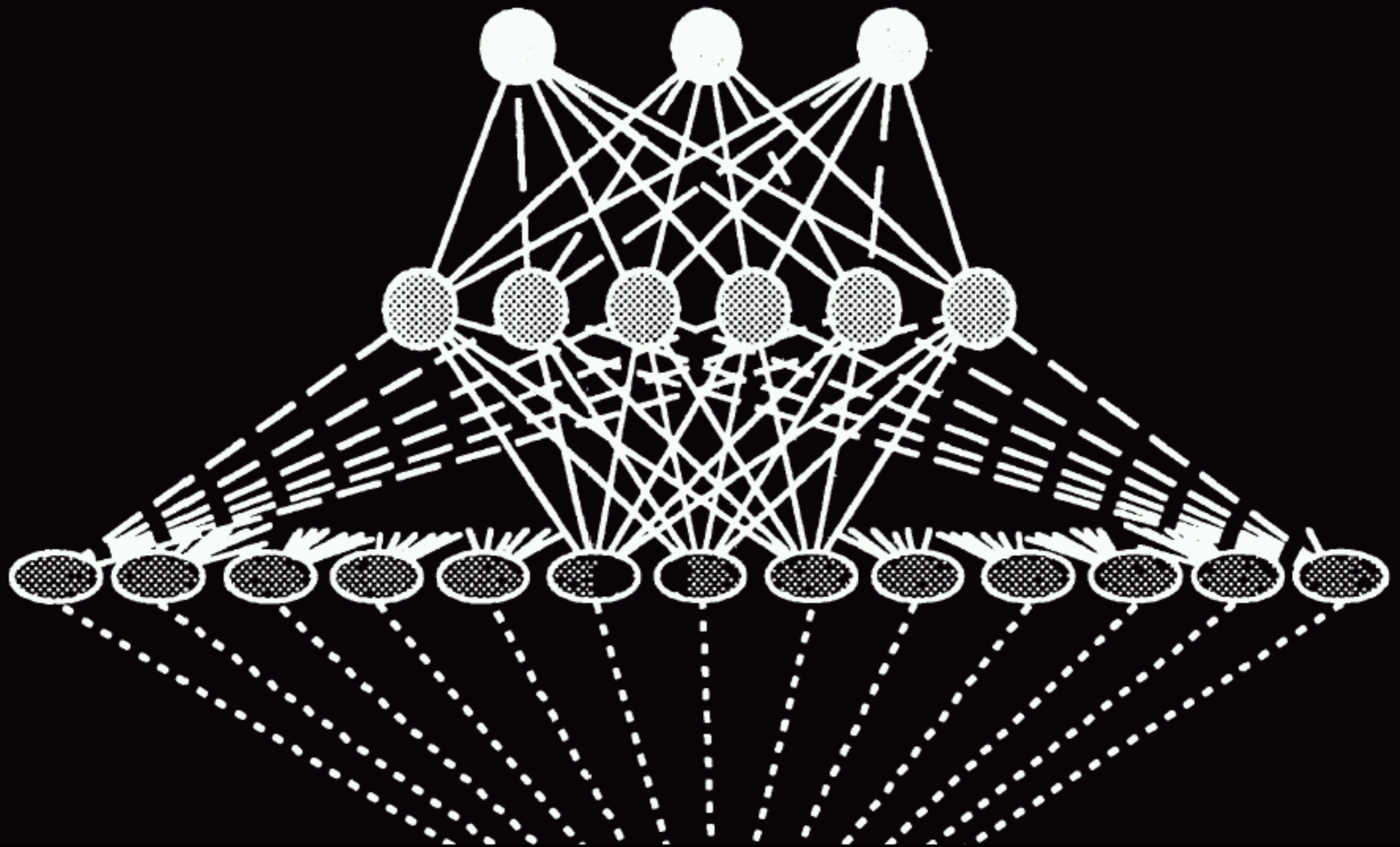


Πρόβλεψη δευτεροταγούς δομής

Protein	Alignments	profile table
		GSAPD NT EKQ C VH IR LM YFW
:	:	:
G	GG GG	5
Y	YY YY 5 . .
I	II EE 2 3
Y	YY YY 5 . .
D	DD DD 5
P	PP PP 5
E	AE AA	. . 3 2
D	VVEE 1 . . 2 2
G	GG GG	5
D	DD DD 5
P	PP PP 5
D	DT DD 4 . . 1
D	NQ NN 1 3 1
G	GN GG	4 1
V	VI VV 1 4 . 1
N	EP KK 1 . 1 . 1 2
P	PP PP 5
G	GG GG	5
T	TT TT 5
D	EK SA	. 1 1 . 1 1 1
F	FF FF 5
:	:	:



Πρόβλεψη δευτεροταγούς δομής



Πρόβλεψη δευτεροταγούς δομής

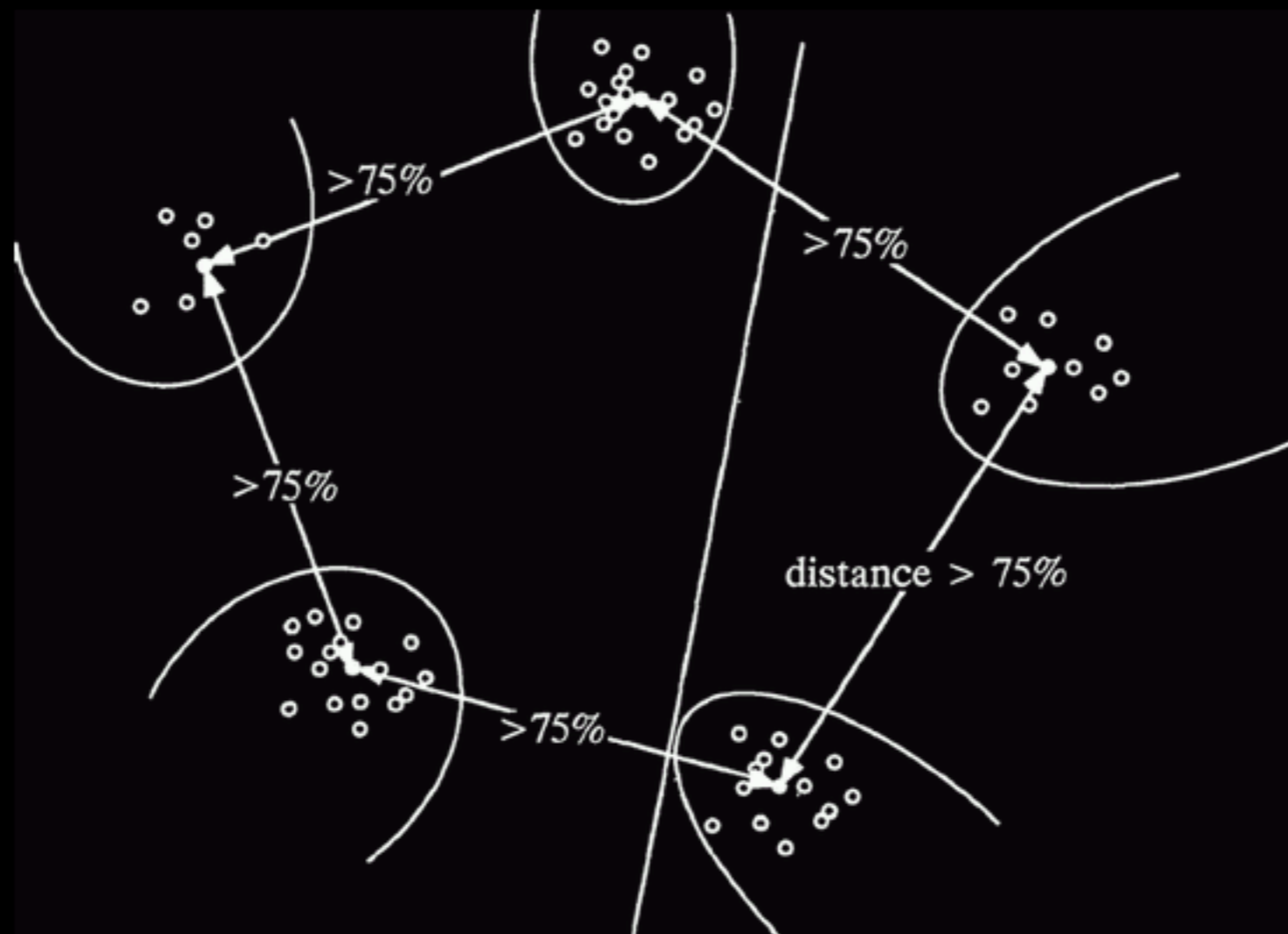
Ένα συνηθισμένο μοντέλο (PHD) είναι :

- Κάθε θέση της πολλαπλής στοίχισης αντιστοιχίζεται σε 21 θέσεις εισόδου στο δίκτυο (20 για τις συχνότερες εμφάνιση των αμινοξέων, 1 για διάκριση N- και C-τελικών άκρων).
- Το δίκτυο "βλέπει" ταυτόχρονα ένα παράθυρο εύρους 13 καταλοίπων από την αλληλουχία (6 κατάλοιπα εκατέρωθεν του υπό εξέταση καταλοίπου). Άρα, το δίκτυο έχει $13 \cdot 21 = 273$ κόμβους εισόδου.
- Το σήμα εισόδου διαδίδεται μέσω ενός δικτύου τριών επιπέδων (εισόδου, hidden, εξόδου). Το επίπεδο εξόδου αποτελείται από τρεις μονάδες οι οποίες αντιστοιχούν στις τρεις υπό πρόβλεψη καταστάσεις.

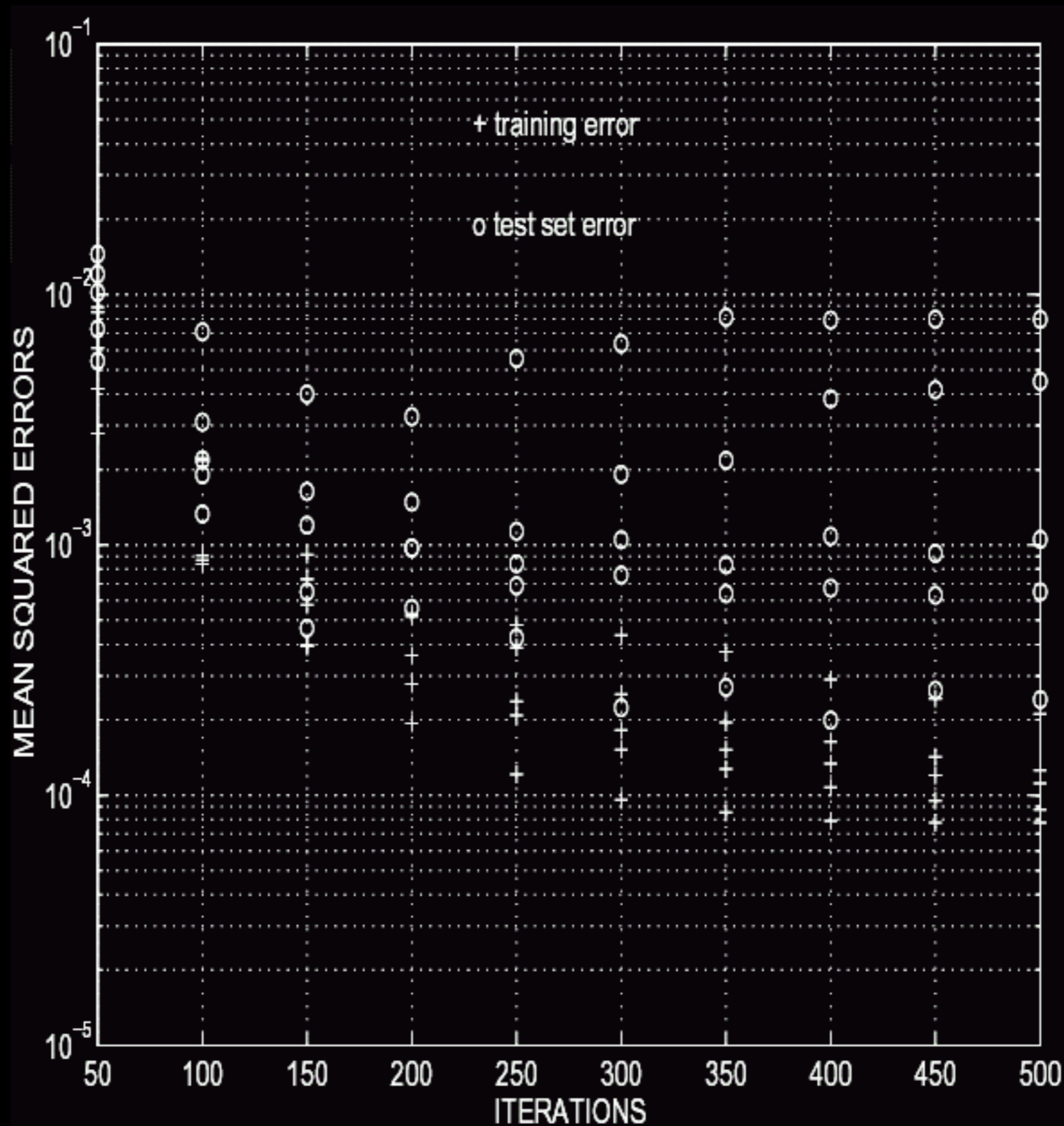
Πρόβλεψη δευτεροταγούς δομής

Training & cross-validation

Προκειμένου να μπορεί να ελεγχθεί η ακρίβεια των προβλέψεων του δικτύου χρησιμοποιείται η μέθοδος του *statistical cross-validation* με ανεξάρτητες ομάδες εκπαίδευσης και ελέγχου.



Πρόβλεψη δευτεροταγούς δομής



Πρόβλεψη δευτεροταγούς δομής

Άλλες βελτιώσεις : **balanced training**

Τα παρατηρούμενα ποσοστά εμφάνισης των τριών καταστάσεων (α,β,loop) στις γνωστές δομές διαφέρουν σημαντικά (α~32%, β~21%, loop~47%). Ένα δίκτυο εκπαιδευμένο με αυτούσιο το δείγμα των γνωστών δομών θα είναι "προκατειλημένο" λόγω της άνισης παρουσίας των τριών καταστάσεων στο δείγμα.

Αποτέλεσμα είναι ότι τα ποσοστά επιτυχίας διαφέρουν από κατάσταση σε κατάσταση (με τη σειρά loop > α > β). Το πρόβλημα διορθώνεται μέσω της εκπαίδευσης του δικτύου με ένα δείγμα στο οποίο τα ποσοστά των τριών καταστάσεων είναι μεταξύ τους ίσα (33%).

Πρόβλεψη δευτεροταγούς δομής

Άλλες βελτιώσεις : εισαγωγή επιπλέον δικτύων

Η πρόβλεψη από το αρχικό δίκτυο μπορεί να χρησιμοποιηθεί ως είσοδος για ένα δεύτερο (structure-to-structure) δίκτυο. Το δεύτερο αυτό δίκτυο εκπαιδεύεται να αναγνωρίζει δομικό περιεχόμενο, μετατρέποντας για παράδειγμα την πρόβλεψη ΗΗΗΗΕΕΗΗ σε ΗΗΗΗΗΗΗΗ, ή διορθώνοντας για την έκταση (μήκος) των στοιχείων δευτεροταγούς δομής στην αρχική πρόβλεψη. Το επιπλέον κέρδος σε ακρίβεια πρόβλεψης είναι μάλλον μικρό, αλλά τα τελικά αποτελέσματα έχουν σαφέστερο δομικό περιεχόμενο (κυρίως από την άποψη της έκτασης των α,β τμημάτων).

Πρόβλεψη δευτεροταγούς δομής

Άλλες βελτιώσεις : "Jury of networks"

Η συνηθέστερες μέθοδοι προσδιορισμού των δεικτών βαρύτητας των συνδέσεων ενός δικτύου είναι μέθοδοι τοπικής βελτιστοποίησης, δηλαδή μέθοδοι οι οποίες βρίσκουν το πλησιέστερο ελάχιστο των διαφορών ανάμεσα στις προβλέψεις του δικτύου και το training set (ως συνάρτηση των τιμών των παραμέτρων). Το αποτέλεσμα είναι ότι δίκτυα εκπαιδευμένα με το ίδιο training set διαφέρουν στις προβλέψεις τους. Η ακριβεία της ολικής πρόβλεψης μπορεί να βελτιωθεί με τη χρήση πλειοψηφικής απόφασης από πολλά δίκτυα εκπαιδευμένα με το ίδιο training set ('Jury of networks').

Πρόβλεψη δευτεροταγούς δομής

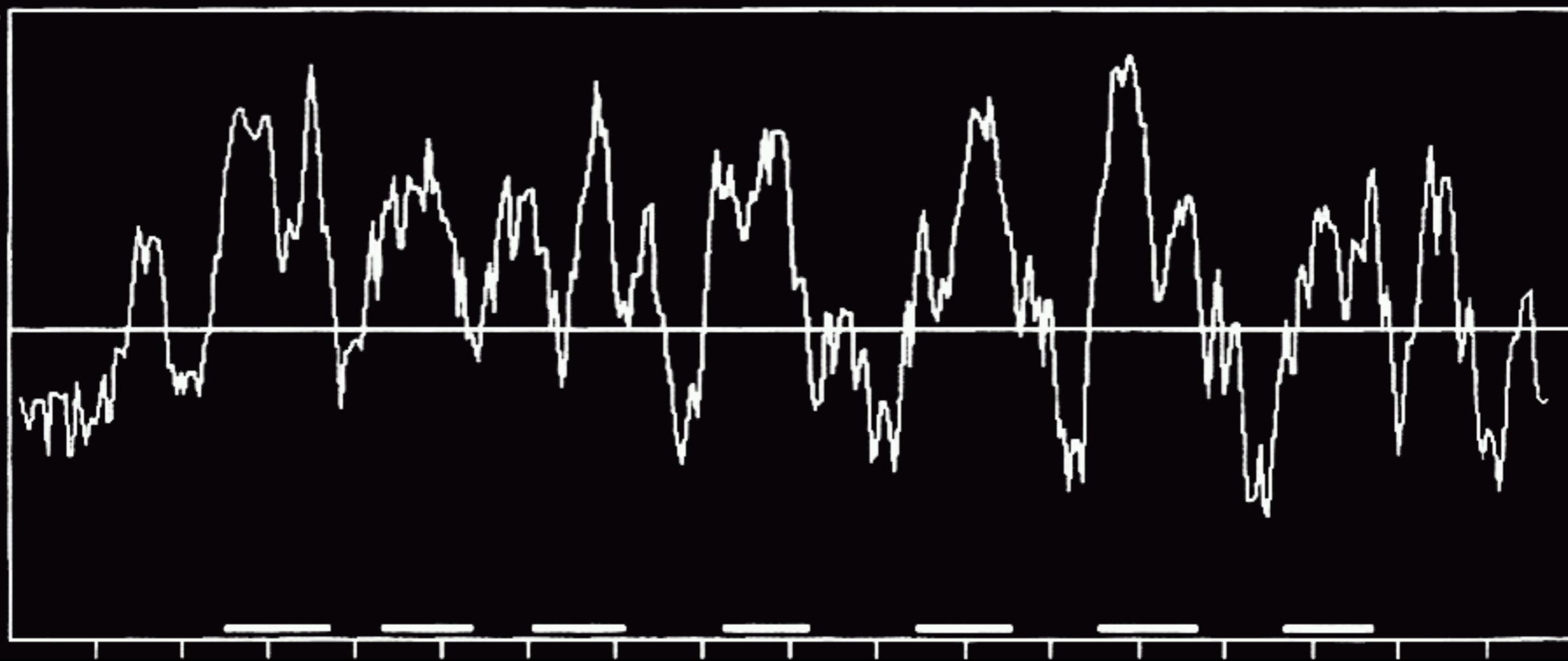
Προγράμματα και εργαλεία

Μερικά από τα πλέον γνωστά προγράμματα [σχεδόν όλα διαθέσιμα μέσω δικτύου (http ή smtp)] είναι :

- PredictProtein (PHD και άλλες μέθοδοι).
- PsiPred (χρήση του PSI-BLAST για την εύρεση ομολογιών και τη δημιουργία της στοίχισης).
- PREDATOR.
- JPred (consensus πολλών μεθόδων).
- SOPMA (consensus πολλών μεθόδων).
- nnpredict.
- PSA (Markov-based).
- GOR (πρόβλεψη με βάση μια αλληλουχία)

Διαμεμβρανικά τμήματα

Όπως και για την πρόβλεψη δευτεροταγούς δομής, έτσι και για την πρόβλεψη διαμεμβρανικών τμημάτων αρχικά χρησιμοποιούνταν αλγόριθμοι βασισμένοι στην υδροφοβικότητα των καταλοίπων (hydropathy plots, π.χ. πρόγραμμα TGREASE).



Διαμεμβρανικά τμήματα

Πιο σύγχρονες μέθοδοι χρησιμοποιούν προφίλ και νευρωνικά δίκτυα για τις προβλέψεις τους.

Γνωστά προγράμματα και μέθοδοι είναι

- TMHMM
- MEMSAT
- TopPred
- PHDtopology

Επειδή το πρόβλημα της πρόβλεψης διαμεμβρανικών τμημάτων (και ιδιαίτερα α-ελίκων) είναι πιο περιοριστικό από την πρόβλεψη δευτεροταγούς δομής (για σφαιρικές πρωτεΐνες) τα ποσοστά επιτυχίας είναι πολύ υψηλά (μέχρι και 95% για έλικα-θηλιά).

Πρόβλεψη τριτοταγούς δομής

Διακρίνονται τρεις βασικές κατηγορίες αλγορίθμων

- Homology modeling,
- Threading (fold recognition), και,
- Ab initio structure prediction

Από αυτές, η τρίτη είναι αυτούσιο το υπολογιστικό τμήμα του προβλήματος της αναδίπλωσης των πρωτεϊνών, η δεύτερη προσπαθεί να εντοπίσει γνωστές δομές οι οποίες να είναι συμβατές με μια νέα αλληλουχία (για την οποία δεν υπάρχουν γνωστές ομολογίες στην PDB), ενώ η πρώτη χρησιμοποιεί τις δομές ομόλογων πρωτεϊνών (από την PDB) για να δημιουργήσει ένα ατομικό μοντέλο για μια νέα αλληλουχία.

Homology modeling

Το πρόβλημα είναι : με δεδομένη μια στοίχιση ανάμεσα σε μια αλληλουχία άγνωστης δομής με μία ή περισσότερες αλληλουχίες γνωστής δομής, ποιά είναι η 'καλύτερη' (πλέον πιθανή) δομή για την νέα αλληλουχία ; (με βάση τη στοίχιση και τις γνωστές δομές).

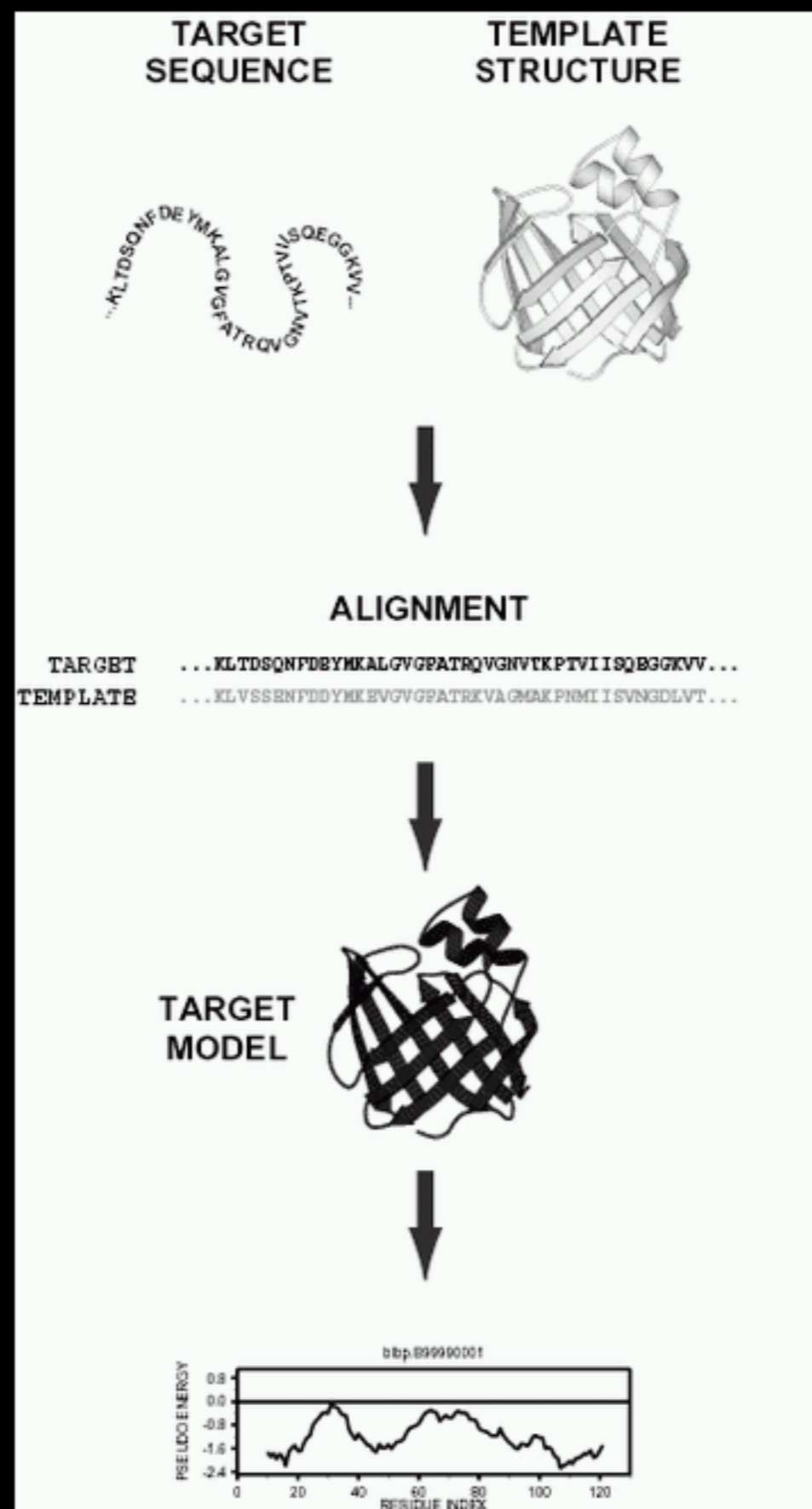
Η βασική υπόθεση για την επίλυση του προβλήματος είναι ότι το μοτίβο αναδίπλωσης της νέας αλληλουχίας είναι το ίδιο με αυτό των αλληλουχιών γνωστής δομής. Τυχόν μεταβολές στις αποστάσεις μεταξύ των στοιχείων δευτεροταγούς δομής (indels) διορθώνονται με την εισαγωγή (στις δομές) βρόχων με κατάλληλο αριθμό καταλοίπων.

Homology modeling

Διαδικασία

1. Εύρεση ομόλογων αλληλουχιών με γνωστή δομή.
2. Στοίχιση των αλληλουχιών.
3. Δημιουργία ενός μοντέλου του πρωτεϊνικού σκελετού της νέας αλληλουχίας χρησιμοποιώντας ως μήτρα τις γνωστές δομές.
4. Στις περιοχές που υπάρχουν εισαγωγές/διαγραφές, εισαγωγή βρόχων κατάλληλου μήκους.
5. Προσθήκη πλευρικών ομάδων.
6. Βελτιστοποίηση των θέσεων των πλευρικών ομάδων.
7. Βελτιστοποίηση ολόκληρης της δομής (energy minimisation, molecular dynamics, ...).

Homology modeling



Homology modeling

Κατασκευή μοντέλου

Οι πλέον συνηθισμένες μέθοδοι είναι :

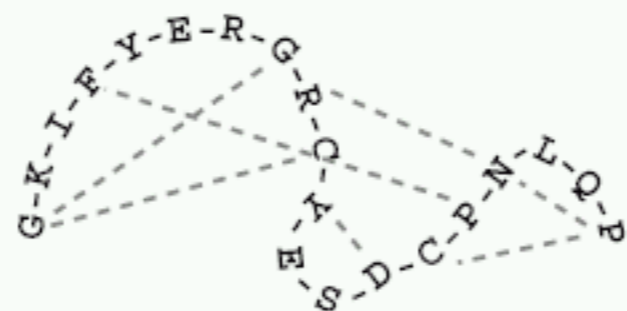
- Συνάθροιση στερεών σωμάτων (rigid-body assembly).
- Segment matching.
- Μέσω εκπλήρωσης χωροταξικών περιορισμών (αντίστοιχο με τον προσδιορισμό δομών με NMR). Οι περιορισμοί αναφέρονται π.χ. σε αποστάσεις μεταξύ καταλοίπων, σε τιμές δίεδρων γωνιών, σε γωνίες και αποστάσεις δεσμών, κοκ. Στόχος είναι να βρεθεί η δομή εκείνη (ή μια οικογένεια συγγενών δομών) που συμφωνεί καλύτερα με το σύνολο των περιορισμών.

Homology modeling

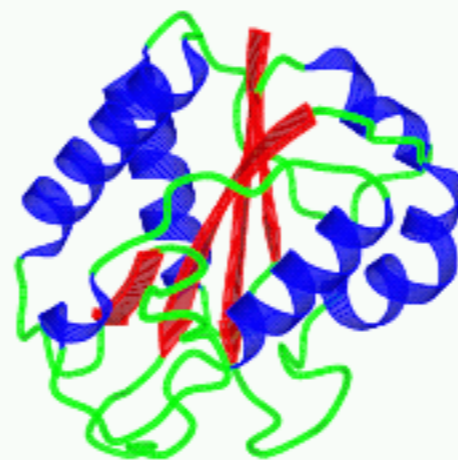
TARGET-TEMPLATE
ALIGNMENT

STRUCTURE GKIFYERGFQGHCHYESDC-NLQP
SEQUENCE GKIFYERG---RCYESDCPNLQP

EXTRACT SPATIAL
RESTRAINTS



SATISFY SPATIAL
RESTRAINTS



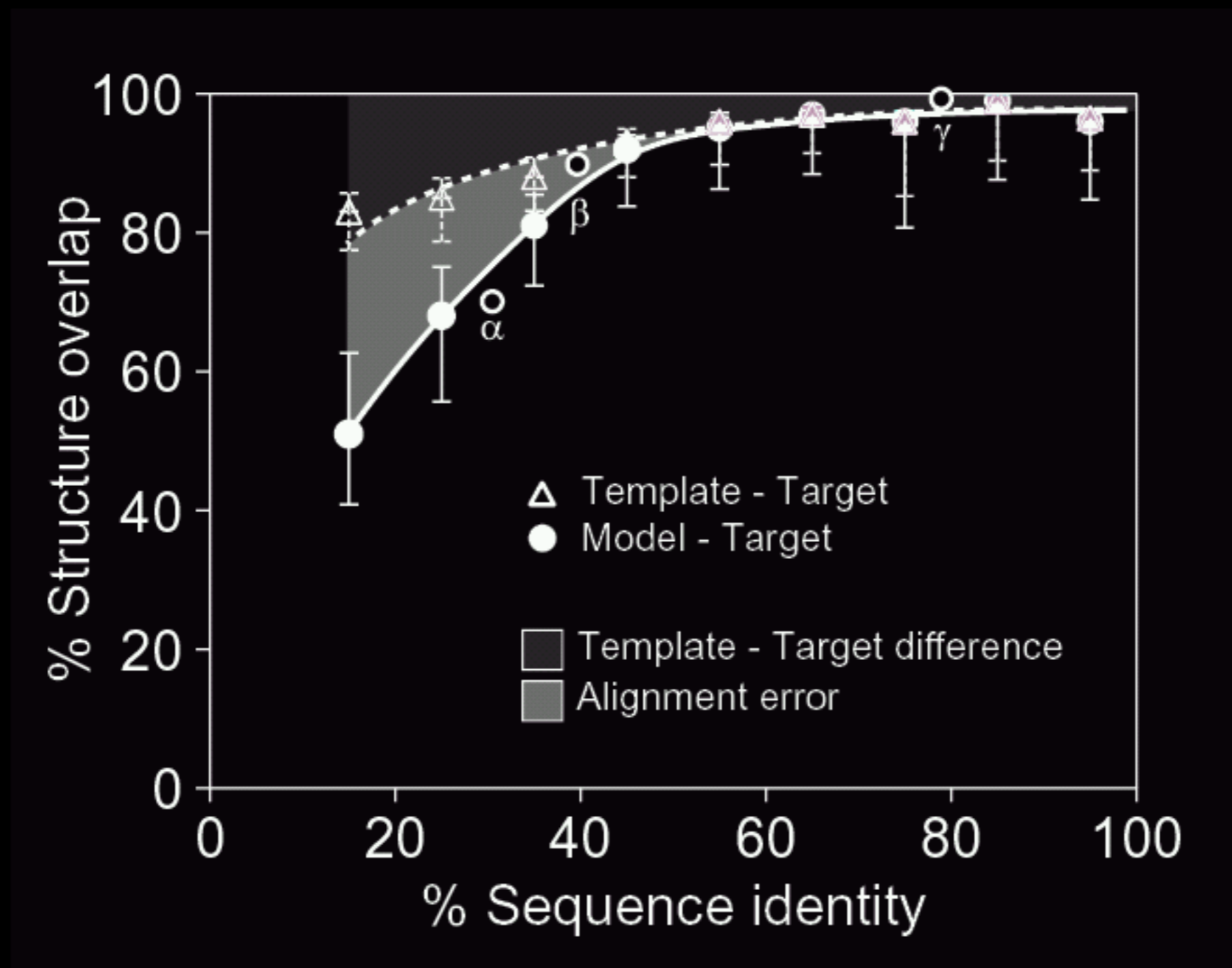
Homology modeling

Loop modelling

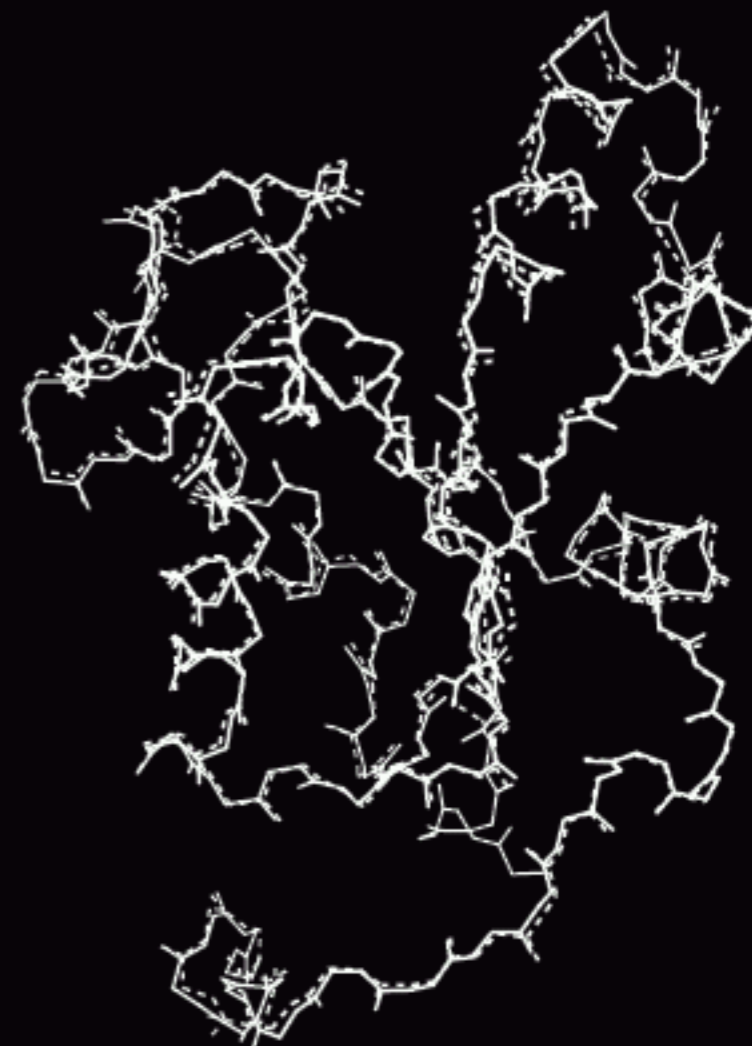
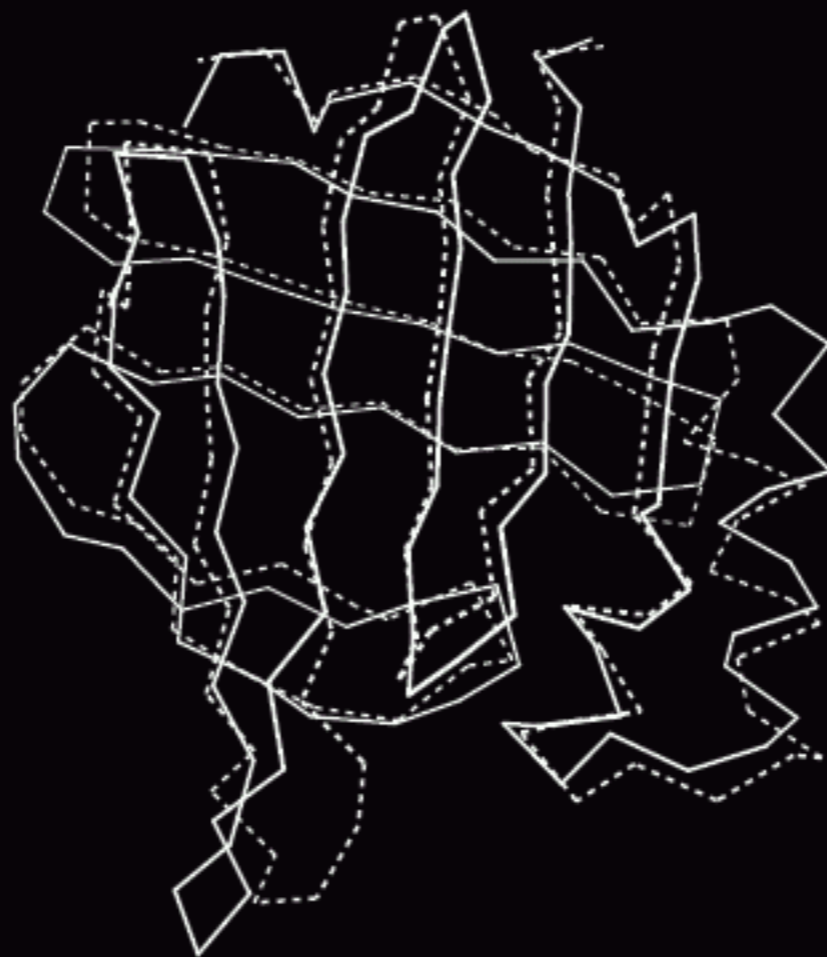
Υπάρχουν δύο βασικές προσεγγίσεις. Η πρώτη βασίζεται σε μία έρευνα της PDB για την εύρεση ενός πρωτεϊνικού τμήματος κατάλληλου μήκους (σε κατάλοιπα) και του οποίου η σχετικές θέσεις των άκρων να ταιριάζουν με το υπο κατασκευή μοντέλο. Η δεύτερη προσέγγιση χρησιμοποιεί ένα δυναμικό πεδίο για να προσδιορίσει (*ab initio*) πιθανές δομές βρόχων συμβατών με το υπάρχον μοντέλο και το περιβάλλον του υπο κατασκευή βρόχου. Η μέθοδος μπορεί να δώσει χρήσιμα αποτελέσματα ($C\alpha$ rmsd $< 2.0\text{\AA}$) ακόμα και για σχετικά μακρύς βρόχους (της τάξεως των 8 καταλοίπων).

Homology modeling

Ομοιότητα vs. ακρίβεια πρόβλεψης



Homology modeling



Homology modeling

Προγράμματα και servers

- Modeller (ModBase)
- Swiss-Model
- 3D-Jigsaw
- Composer
- InsightII (εμπορικό)
- Sybyl (εμπορικό)

Homology modeling

The screenshot shows a Netscape browser window displaying the SWISS-MODEL website. The browser's address bar shows the file path: file:///home/glykos/teach/bioinf/10/SWISS-MODEL.html. The website has a yellow and blue color scheme. On the left, there is a 'MENU' section with two sub-sections: 'Modelling requests:' and 'Interactive tools', each containing a list of links. The main content area on the right features a search bar with 'Swiss-Prot/TrEMBL' selected, a search button, and a clear button. Below the search bar is the SWISS-MODEL logo, which consists of a blue ribbon structure of a protein with a red Swiss cross in the center, followed by the text 'SWISS-MODEL' and 'An Automated Comparative Protein Modelling Server'. The main text describes the server as a fully automated protein structure homology-modeling server, accessible via the ExPASy web server or the program DeepView. It mentions the current version is 3.5 and is under constant improvement. It also states that the server was initiated in 1993 by Manuel Peitsch and is now being further developed within the SIB - Swiss Institute of Bioinformatics in collaboration with Torsten Schwede at the Structural Bioinformatics Group, Biozentrum (University of Basel) and Nicolas Guex at GlaxoSmithKline R&D. The computational resources are provided in collaboration by the Biozentrum (University Basel) and the Advanced Biomedical Computing Center (NCI Frederick, USA). At the bottom, there is a 'Disclaimer' section stating that the result of any modelling procedure is NON-EXPERIMENTAL and MUST be considered with care. The footer of the website lists the authors: Torsten Schwede, Manuel C. Peitsch & Nicolas Guex. The browser's status bar at the bottom shows 'Document: Done (1.383 secs)'.

File Edit View Go Bookmarks Tools Window Help

file:///home/glykos/teach/bioinf/10/SWISS-MODEL.html Go Search

Home Radio Bookmarks Members WebMail Connections BizJournal SmartUpdate Mktplace

SWISS-MODEL

MENU

Modelling requests:

- [First Approach mode](#)
- [Alignment Interface](#)
- [Optimise \(project\) mode](#)
- [Oligomer modelling](#)
- [GPCR mode](#)

Interactive tools


- [Swiss-PdbViewer](#), a tool for viewing and manipulating protein structures and models.
- [Lookup](#) the ExPDB template **codes** accessible to SWISS-MODEL using known PDB codes.
- [Search](#) the template sequences accessible to SWISS-MODEL.
- [Examples](#) using SWISS-MODEL and the Swiss-PdbViewer.
- [ANOLEA](#) Protein structure quality check (atomic non-local environment assessment)
- [News from Swiss-Model](#)

HELP

- [Frequently Asked Questions](#).
- [Visualising 3D models](#).
- [Reliability of models](#).
- [How SWISS-MODEL works](#).
- [How ProModII works](#).
- [Modelling of oligomeric proteins](#).
- [Model Confidence factors](#).
- [About model quality](#)

[ExPASy Home page](#) [Site Map](#) [Search ExPASy](#) [Contact us](#)

Search for



SWISS-MODEL

An Automated Comparative Protein Modelling Server

[SWISS-MODEL](#) is a fully automated protein structure homology-modeling server, accessible via the [ExPASy](#) web server, or from the program [DeepView](#) (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us refine the sequence analysis and modelling algorithms, please [report](#) of possible bugs and problems with the modelling procedure.

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and is now being further developed within the [SIB - Swiss Institute of Bioinformatics](#) in collaboration between Torsten Schwede at the [Structural Bioinformatics Group](#), Biozentrum (University of Basel) and Nicolas Guex at [GlaxoSmithKline R&D](#). The computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel) and the [Advanced Biomedical Computing Center](#) (NCI Frederick, USA).

Disclaimer

The result of any modelling procedure is NON-EXPERIMENTAL and MUST be considered with care. This is especially true since there is no human intervention during model building. Carefully read the header section of the files to know what templates and alignments were used during the model building process.

[Torsten Schwede](#), [Manuel C. Peitsch](#) & [Nicolas Guex](#).

Document: Done (1.383 secs)

Threading

Η βασική ιδέα είναι : αντί να ερευνήσουμε όλες τις πιθανές διαμορφώσεις της νέας αλληλουχίας (δηλ. το folding problem), ερευνούμε πόσο συμβατή είναι η νέα αλληλουχία με κάθε μία από τις υπάρχουσες γνωστές δομές. Αυτός ο έλεγχος μπορεί να γίνει γρήγορα : ο πρωτεϊνικός σκελετός μένει αμετάβλητος και το μόνο που αλλάζει είναι οι πλευρικές ομάδες των καταλοίπων. Δομές που είναι ασύμβατες με κάποια αλληλουχία θα οδηγούν σε ενεργειακά απίθανες διαμορφώσεις (π.χ. καταστροφή του υδρόφοβου πυρήνα). Το threading συνήθως χρησιμοποιείται για fold recognition και όχι για τη κατασκευή ενός πλήρους μοντέλου.

Threading

Προγράμματα και servers

- UCLA-DOE
- TOPITS
- Threader
- 123D+
- 3D-PSSM

The screenshot shows a web browser window displaying the 3D-PSSM Protein Fold Recognition Server. The browser's address bar shows the file path: file://home/glykos/teach/bioinf1/03D-PSSMServer.html. The website header includes the logo "3D-pssm" and the text "Imperial College of Science, Technology & Medicine Fold Recognition Server". A navigation menu on the left lists: Home, Recognise a Fold, Fold Library, Authors, Links, and Help. The main content area features a blue bar with "Fold Library Last Updated: Tue Jan 6 08:00:00 2004" and "(9274) Structures". Below this is a green bar with "Disclaimer and Terms of Use" and "Last updated: Sun Jan 11 20:49:51 2004, last updated: Tue, 06 Aug 2002 12:27:55 GMT". A green bar displays "Visitors To Date: 168,243". The main heading reads "Welcome to the 3D-PSSM Web Server V 2.6.0". Below this is a blue box with the text: "A Fast, Web-based Method for Protein Fold Recognition using 1D and 3D Sequence Profiles coupled with Secondary Structure and Solvation Potential Information." A red text instruction says: "Click on 'Recognise a Fold' in the menubar to the left to submit your sequence". A red text note states: "Please Note: 3D-PSSM is now for academic use only." There are two red text notices at the bottom: "Temporary removal of d3btaa3" dated Mar 26 2002, and "Move to Imperial College of Science, Technology and Medicine" dated Mar 4 2002. The browser's status bar at the bottom shows "Document: Done (2.176 secs)".

Το πρόβλημα της αναδίπλωσης των πρωτεϊνών

Με μοναδικό δεδομένο την πρωτοταγή αλληλουχία μιας πρωτεΐνης πως μπορούμε να προσδιορίσουμε την τρισδιάστατη δομή της ;

Ισοδύναμα : πως η αλληλουχία μιας πρωτεΐνης καθορίζει την τρισδιάστατη δομή της ;

Πριν ασχοληθούμε με το πως η πρωτοταγής δομή μιας πρωτεΐνης καθορίζει την τριτοταγή της δομή, θα πρέπει να γνωρίζουμε ότι η πρωτοταγής δομή όντως αρκεί για να καθορίσει την τριτοταγή δομή.

Τα πειράματα του Anfinsen

The Observation:



Native
(100% active)

1. Reduce
2. 8 M urea



Denatured
(inactive)

1. Remove urea
2. Oxidize



Native
(>90% active)

The Control:



Native

1. Reduce
2. 8 M urea



Denatured

1. Oxidize
2. Remove urea



"Scrambled"
(1-2% active)

Τα πειράματα του Anfinsen

Αυτά τα πειράματα έδειξαν ότι (i) η τρισδιάστατη δομή των πρωτεϊνών καθορίζεται από την πρωτοταγή δομή τους, και, (ii) ότι η διαδικασία αναδίπλωσης είναι αυθόρμητη, δηλ. ότι η θερμοδυναμικά σταθερότερη κατάσταση είναι αυτή της αναδιπλωμένης πρωτεΐνης. Το (ii) επίσης υπονοεί ότι η φυσική δομή της πρωτεΐνης αντιστοιχεί στο ολικό (ή σε πλησίον του ολικού) ελάχιστο της ελεύθερης ενέργειας του συστήματος. Άρα το πρόβλημα της αναδίπλωσης των πρωτεϊνών μπορεί να διατυπωθεί και ως εξής : με δεδομένη την πρωτοταγή δομή μιας πρωτεΐνης, ποιά τριτοταγή της δομή αντιστοιχεί στο ολικό ενεργειακό ελάχιστο;

Το παράδοξο του Levinthal

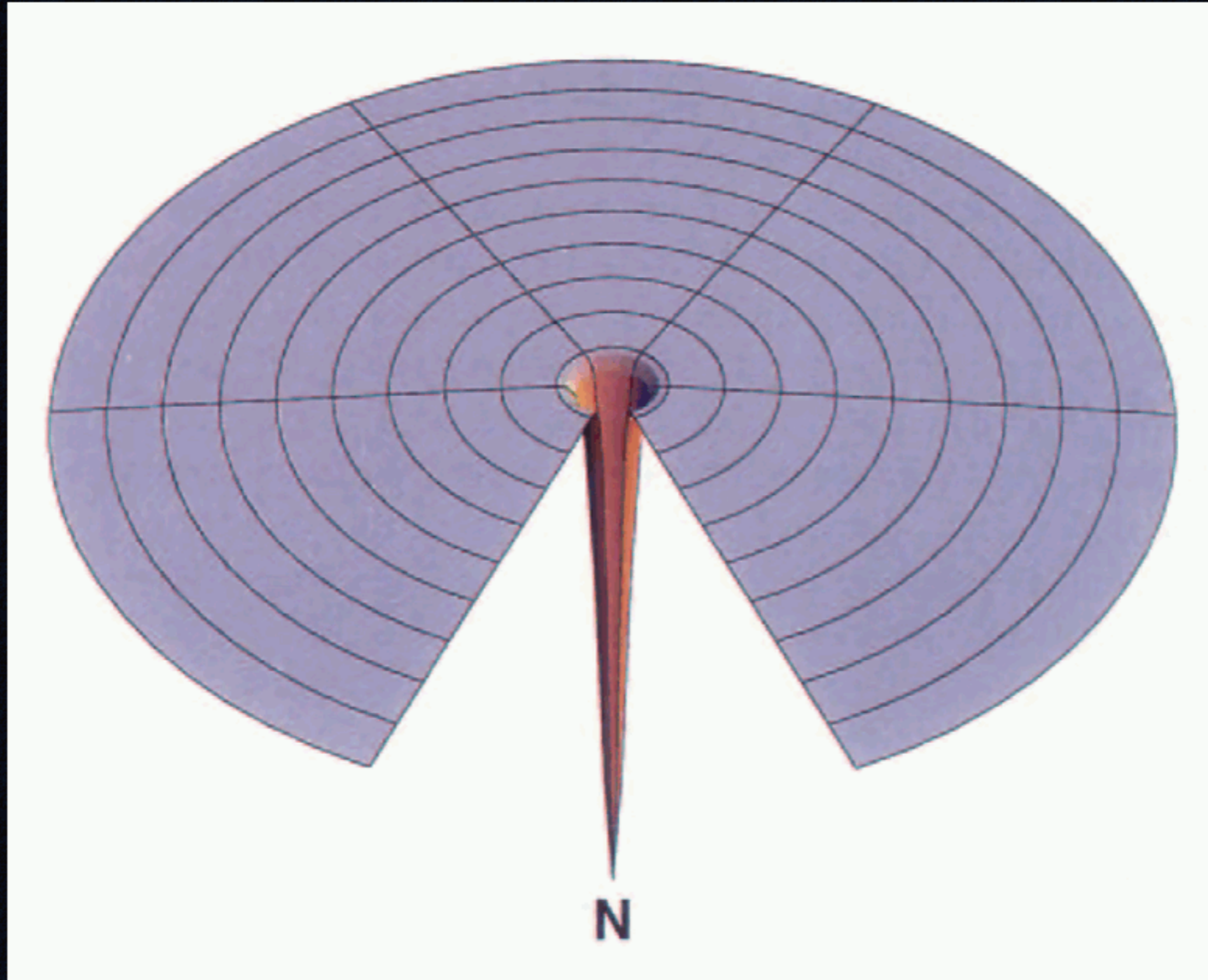
Υποθέστε ότι για κάθε κατάλοιπο υπάρχουν μόνο 10 ενεργειακά αποδεκτοί συνδυασμοί των φ,ψ γωνιών. Τότε, ο ολικός αριθμός πιθανών δομών για μια αλυσίδα 40 αμινοξέων είναι 10^{40} . Ακόμα και εάν μια πρωτεΐνη μπορούσε να περάσει από 10^{14} (100 τρις) δομές ανά δευτερόλεπτο, πάλι θα χρειαζόνταν ~ 67 εκατομμύρια φορές την ηλικία του σύμπαντος για να διπλωθεί. Ευτυχώς, τους παίρνει πολύ λιγότερο από αυτό : από χιλιοστά του δευτερολέπτου έως λεπτά.

Διάκριση θερμοδυναμικής-κινητικής της αναδίπλωσης

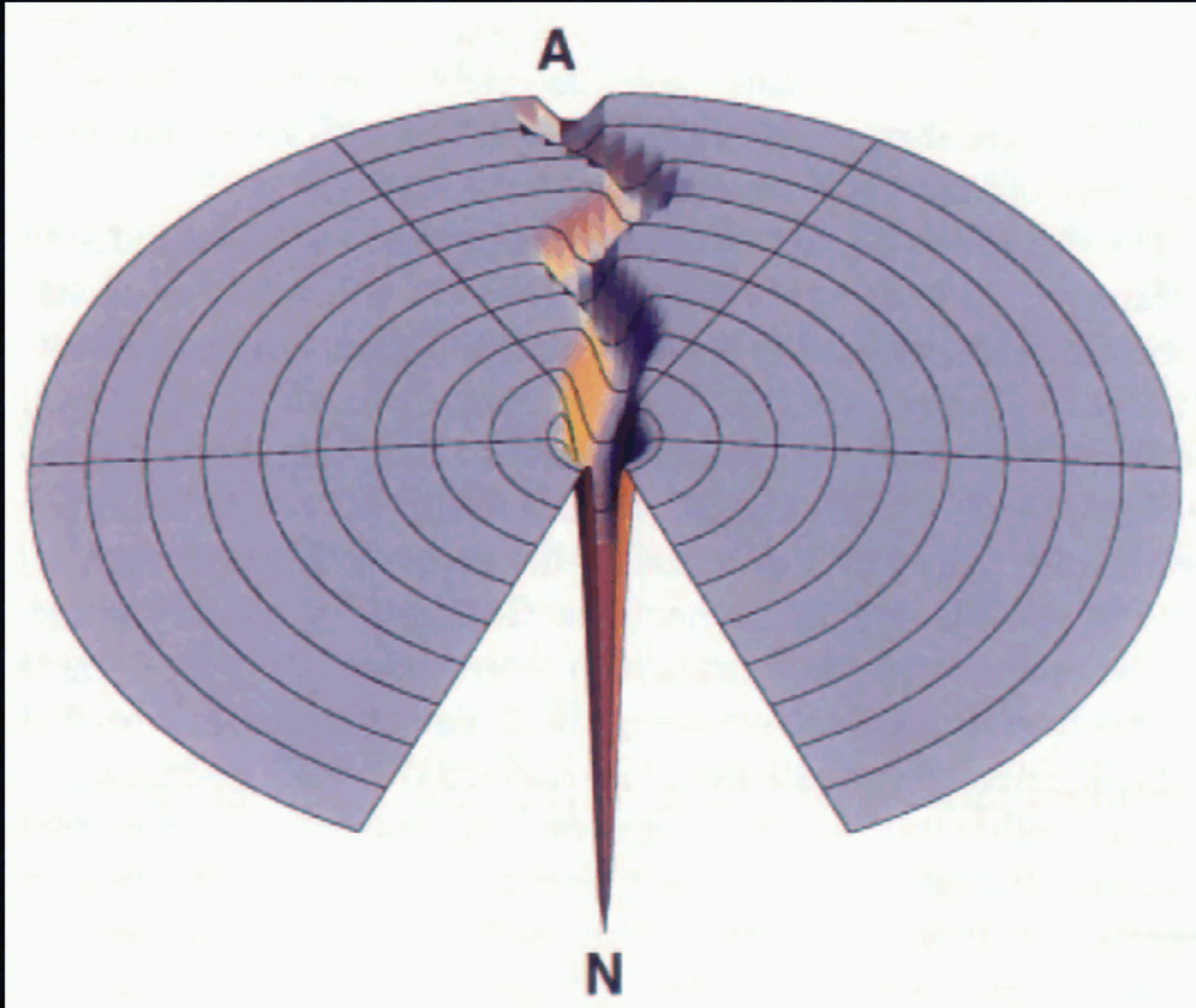
Το πρόβλημα της αναδίπλωσης των πρωτεϊνών

Μετά από 30 χρόνια εντατικής ερευνητικής προσπάθειας, το πρόβλημα της αναδίπλωσης των πρωτεϊνών παραμένει αμετακίνητο στην κορυφή των βασικών άλυτων προβλημάτων της Μοριακής Βιολογίας. Μία περίοδος αισιοδοξίας στις αρχές της δεκαετίας του '90 έχει αντικατασταθεί από τη σχεδόν πλήρη αποδοχή της απιθανότητας να λυθεί το πρόβλημα στο άμεσο μέλλον. Κυρίαρχο ρόλο στην αναγνώριση της πολυπλοκότητας του προβλήματος έπαιξε η ανάπτυξη της θεωρίας του τοπίου ενέργειας (energy landscape theory of protein folding) :

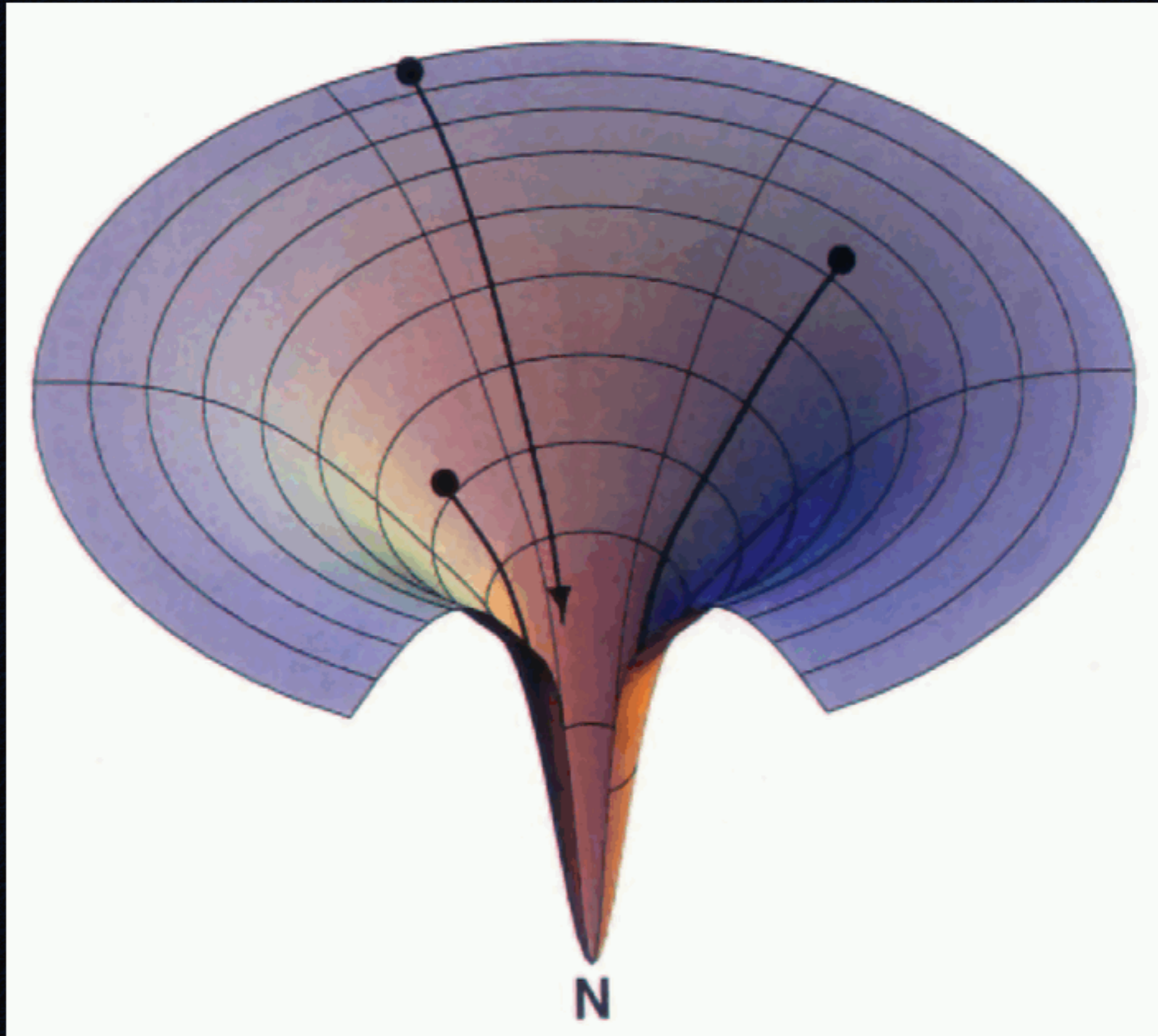
Energy landscape : Levinthal



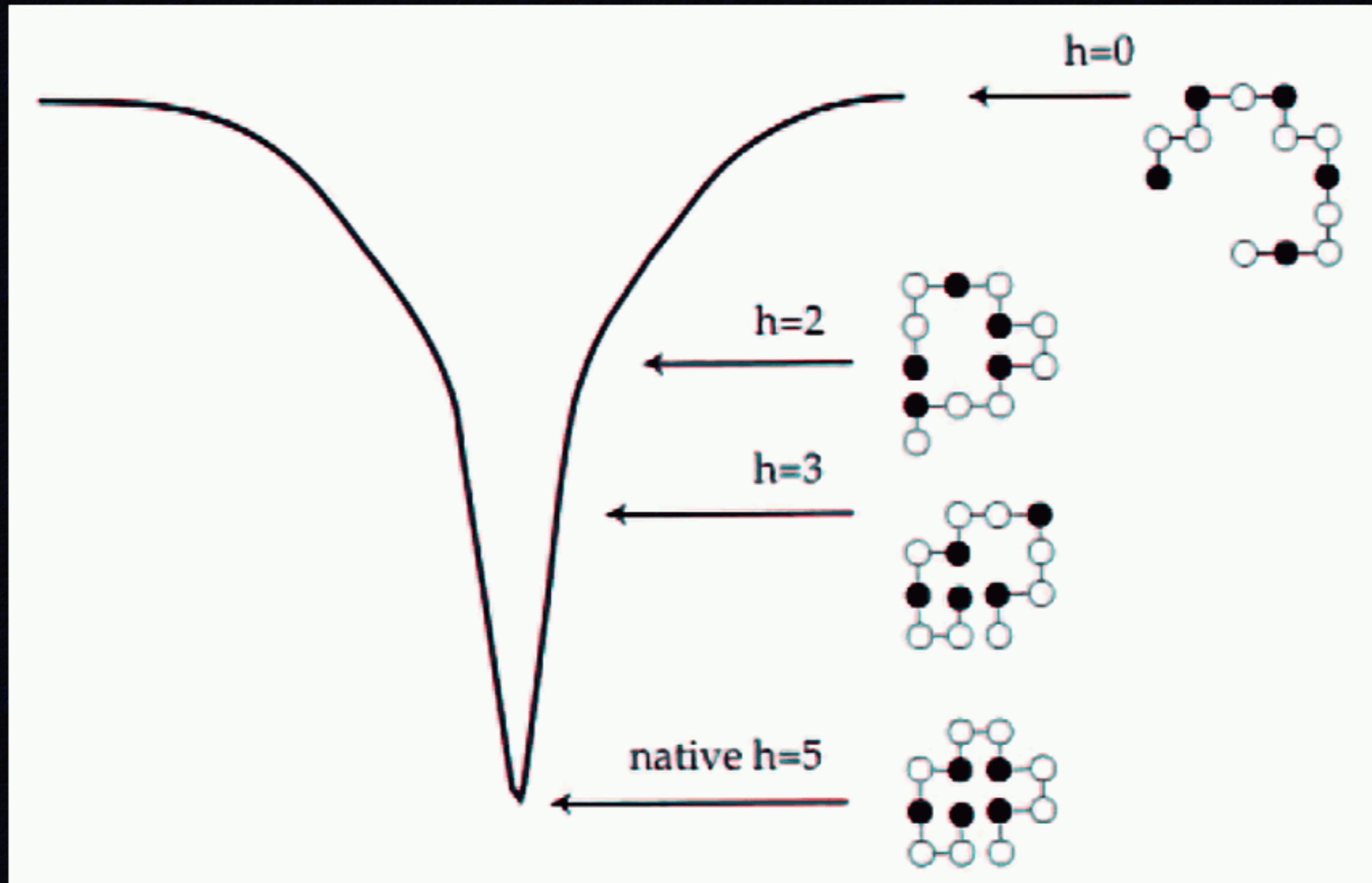
Energy landscape : pathways



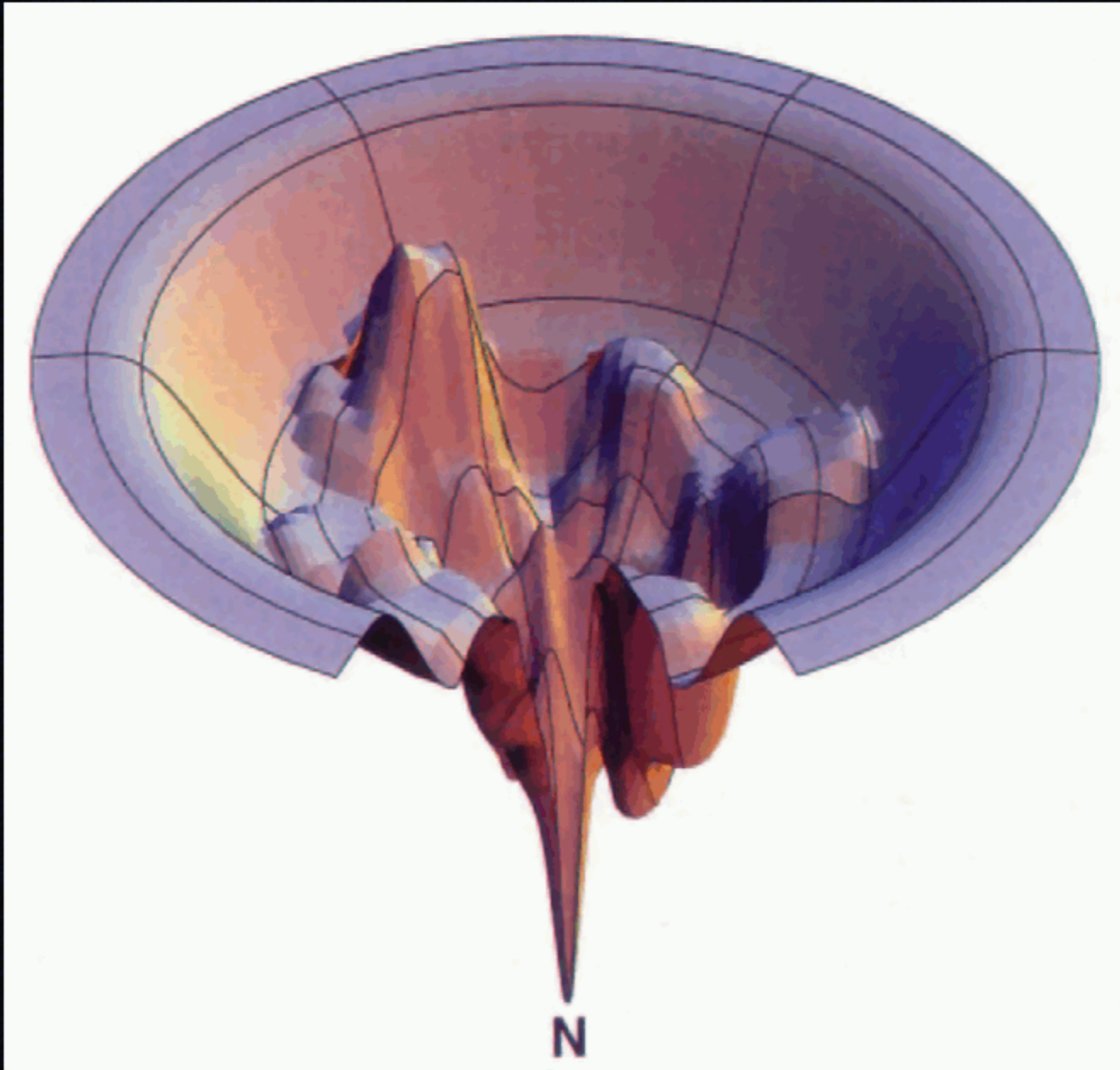
Energy landscape : the funnel



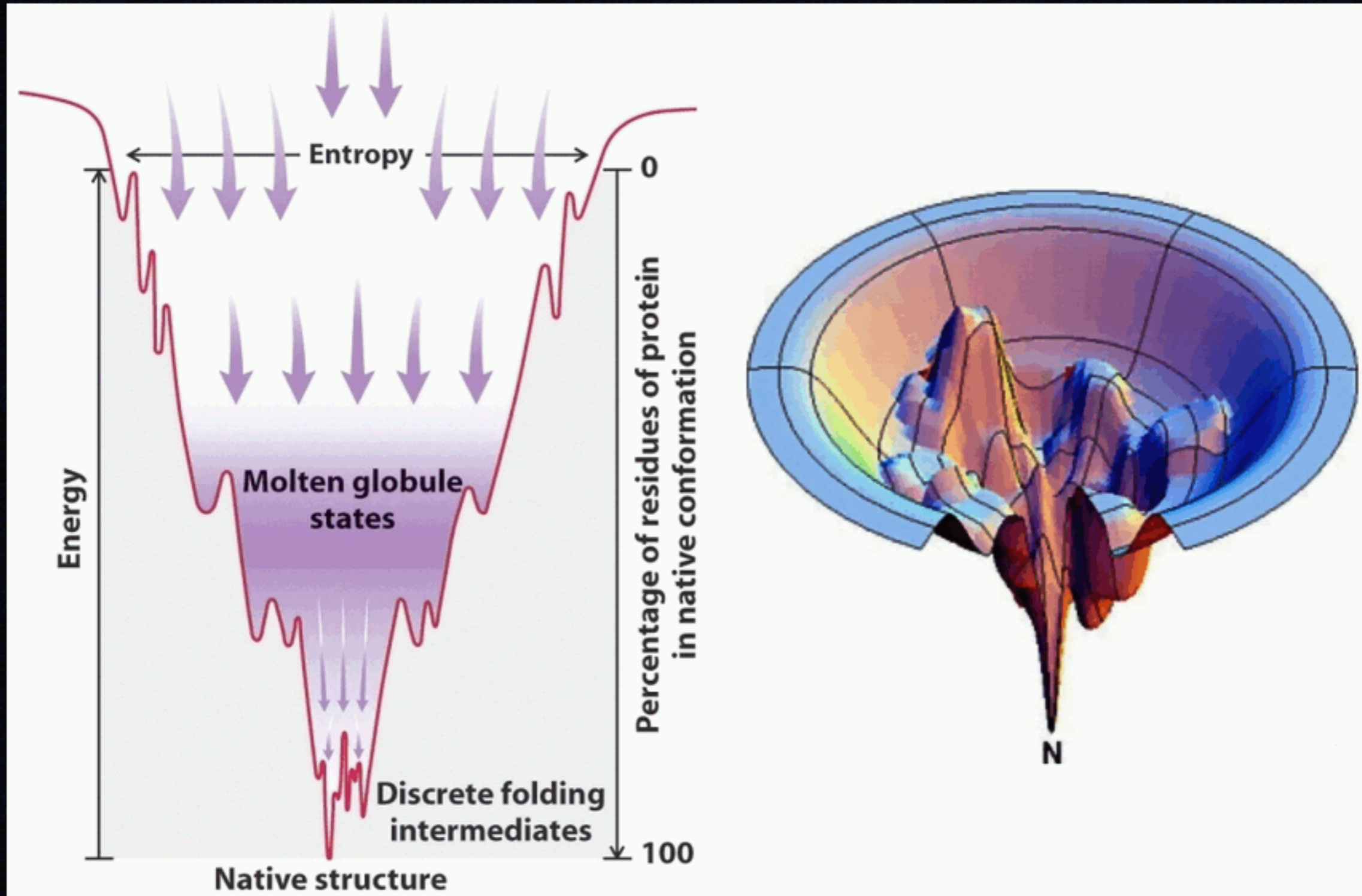
Energy landscape : The connection with evolution



The energy landscape



The energy landscape



The energy landscape



Αναδίπλωση πρωτεϊνών : The physics-based approach

Η βασική υπόθεση είναι : εάν μπορούμε με ακρίβεια να υπολογίσουμε την ενέργεια ενός συστήματος ατόμων ως συνάρτηση των θέσεων και ταχυτήτων τους, τότε θα μπορούσαμε να προσδιορίσουμε τη φυσική κατάσταση (native state) των πρωτεϊνών και πεπτιδίων μέσω μίας προσομοίωσης αυτού του συστήματος ατόμων ως συνάρτηση του χρόνου. Αυτή η προσέγγιση είναι γνωστή ως

Προσομοιώσεις αναδίπλωσης μοριακής δυναμικής

Μοριακή δυναμική

Προκειμένου να καταστεί εφικτή η προσομοίωση (σε ατομικό επίπεδο) ενός συστήματος, θα πρέπει να μπορεί να υπολογιστεί η ενέργεια του συστήματος ως συνάρτηση των ατομικών θέσεων.

Η ενέργεια ενός συστήματος ατόμων αποτελείται από δυο όρους : την κινητική και τη δυναμική ενέργεια.

Η κινητική ενέργεια υπολογίζεται εύκολα ως συνάρτηση των μαζών και των ταχυτήτων των ατόμων.

Η ουσία του προβλήματος βρίσκεται στον υπολογισμό της δυναμικής ενέργειας.

Ενέργεια πρωτεϊνών

Το ζητούμενο είναι να βρεθεί η δυναμική ενέργεια των πρωτεϊνών ως συνάρτηση της δομής τους (της διαμόρφωσης τους στον τρισδιάστατο χώρο). Θεωρητικά, ότι μπορούμε να μάθουμε για το σύστημα (συμπεριλαμβανομένης και της ενέργειας του) βρίσκεται στην επίλυση της εξίσωσης του Schrodinger. Δυστυχώς, οι κβαντομηχανικοί υπολογισμοί είναι τόσο χρονοβόροι που γίνονται πρακτικά ανέφικτοι για περισσότερα από μερικές δεκάδες ή εκατοντάδες άτομα. Για το λόγο αυτό έχουν αναπτυχθεί εμπειρικές μέθοδοι προσδιορισμού της δυναμικής ενέργειας των μακρομορίων, τα λεγόμενα δυναμικά πεδία (force fields).

Force fields

Τυπικά, αυτά τα δυναμικά πεδία περιγράφουν την ολική δυναμική ενέργεια του μορίου ως άθροισμα δύο όρων, ενός για τις δεσμικές (μέσω δεσμών) αλληλεπιδράσεις, και ενός για μη δεσμικές αλληλεπιδράσεις. Οι δεσμικές αλληλεπιδράσεις περιλαμβάνουν όρους για τα μήκη δεσμών, τις γωνίες δεσμών, για άτομα που πρέπει να μείνουν συνεπίπεδα (π.χ. αρωματικά), κλπ. Οι μη δεσμικές αλληλεπιδράσεις είναι το άθροισμα των ηλεκτροστατικών αλληλεπιδράσεων συν ένα Lennard-Jones δυναμικό που περιγράφει τις δυνάμεις διασποράς και την άπωση μεταξύ των ηλεκτρονιακών νεφών.

Force fields

$$\mathcal{H} = \mathcal{K} + \mathcal{V}$$

$$\mathcal{V} = \mathcal{V}_{\text{bonding}} + \mathcal{V}_{\text{non-bonding}}$$

$$\mathcal{V}_{\text{bonding}} = \mathcal{V}_{\text{lengths}} + \mathcal{V}_{\text{angles}} + \mathcal{V}_{\text{dihedral}} + \mathcal{V}_{\text{improper}}$$

$$\mathcal{V}_{\text{non-bonding}} = \mathcal{V}_{\text{electr}} + \mathcal{V}_{\text{L-J}}$$

Force fields

$$\mathcal{V}_{\text{lenghts}} = \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2$$

$$\mathcal{V}_{\text{angles}} = \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2$$

$$\mathcal{V}_{\text{dihedral}} = \sum_{\text{dihedrals}} \frac{1}{2} V_n (1 + \cos(n\phi - \delta))$$

Force fields

$$\mathcal{V}_{\text{electr}} = \sum_{\text{all pairs } s} \frac{q_i q_j}{\epsilon r_{ij}}$$

$$\mathcal{V}_{\text{L-J}} = \sum_{\text{all pairs } s} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

Μοριακή δυναμική

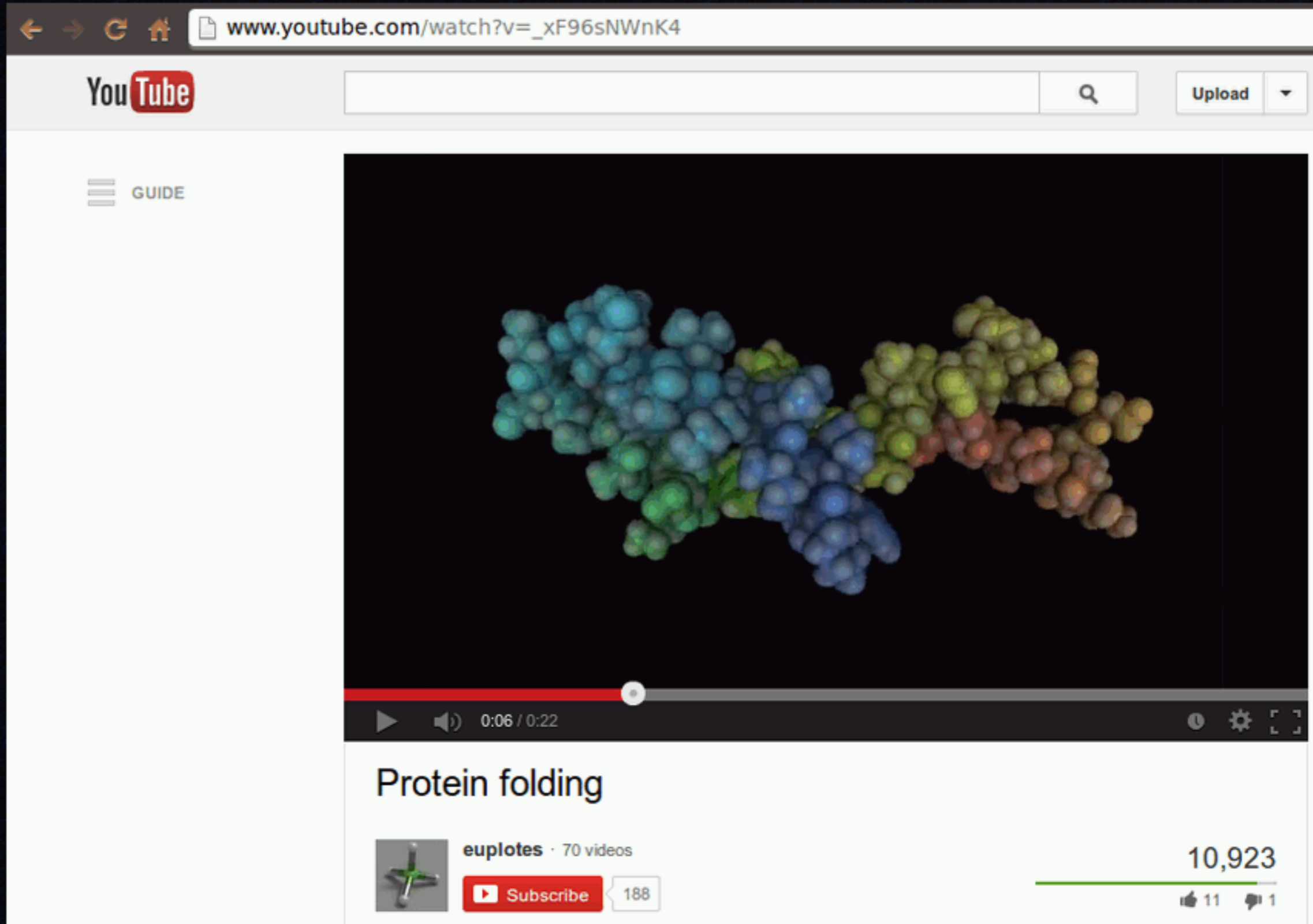
$$\mathcal{H} = \mathcal{K} + \mathcal{V}$$

$$\mathcal{K} = \frac{1}{2} \sum_{\text{all atoms}} m_i \mathbf{u}_i^2$$

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} = - \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i}$$

Μοριακή δυναμική

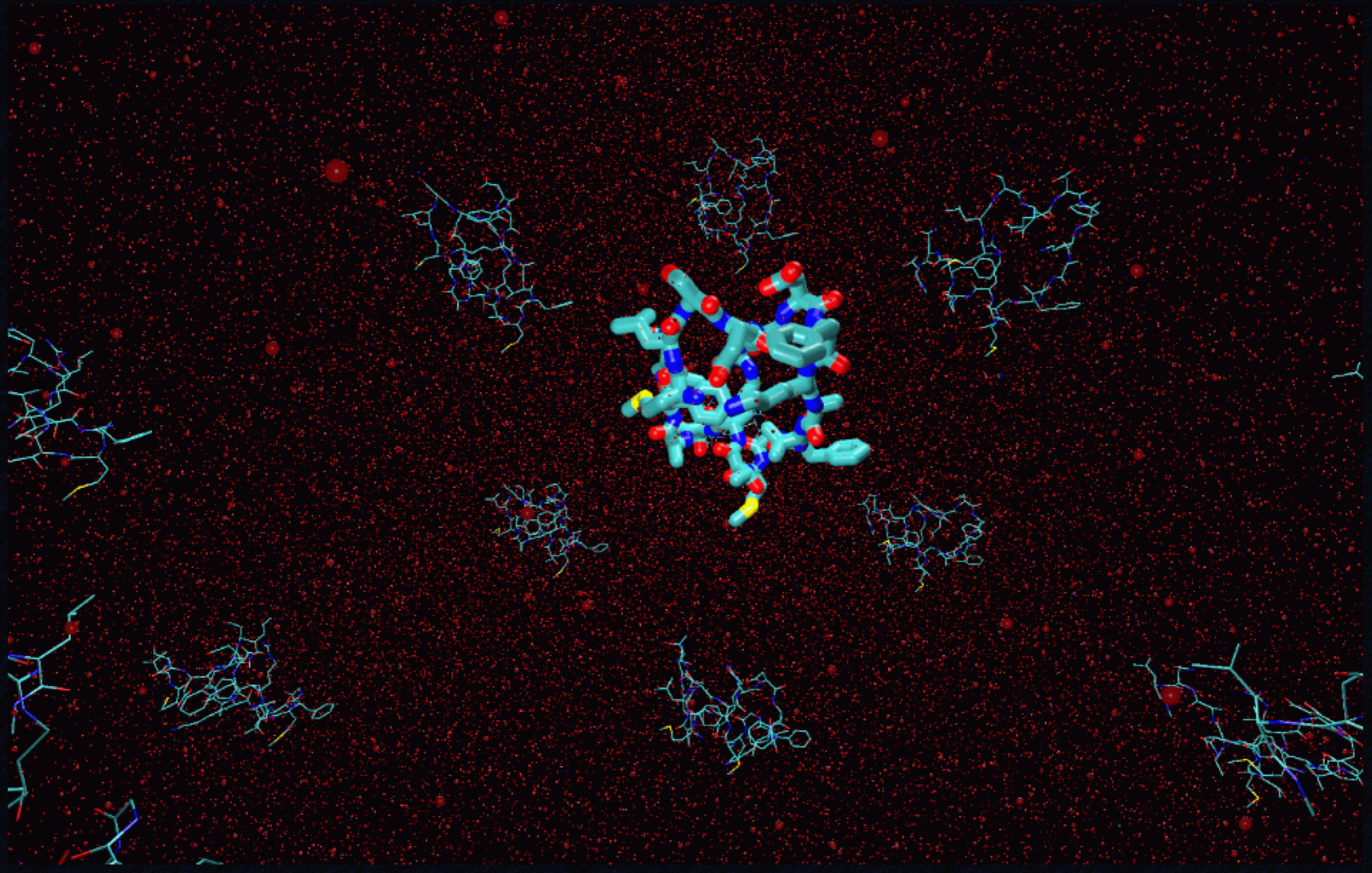
Molecular motion illustrated : a short video.



The image shows a screenshot of a YouTube video player. The browser's address bar at the top displays the URL www.youtube.com/watch?v=_xF96sNWnK4. The YouTube logo is visible in the top left corner of the player interface. A search bar and an 'Upload' button are located in the top right. On the left side, there is a 'GUIDE' menu icon. The main video area shows a 3D molecular model of a protein, represented by a cluster of spheres in various colors (blue, green, yellow, orange, red) against a black background. Below the video, a progress bar indicates the video is at 0:06 of a 0:22 duration. The video title is 'Protein folding'. The channel name is 'euplotes' with 70 videos. There is a 'Subscribe' button with 188 subscribers. The video has 10,923 views, 11 likes, and 1 dislike.

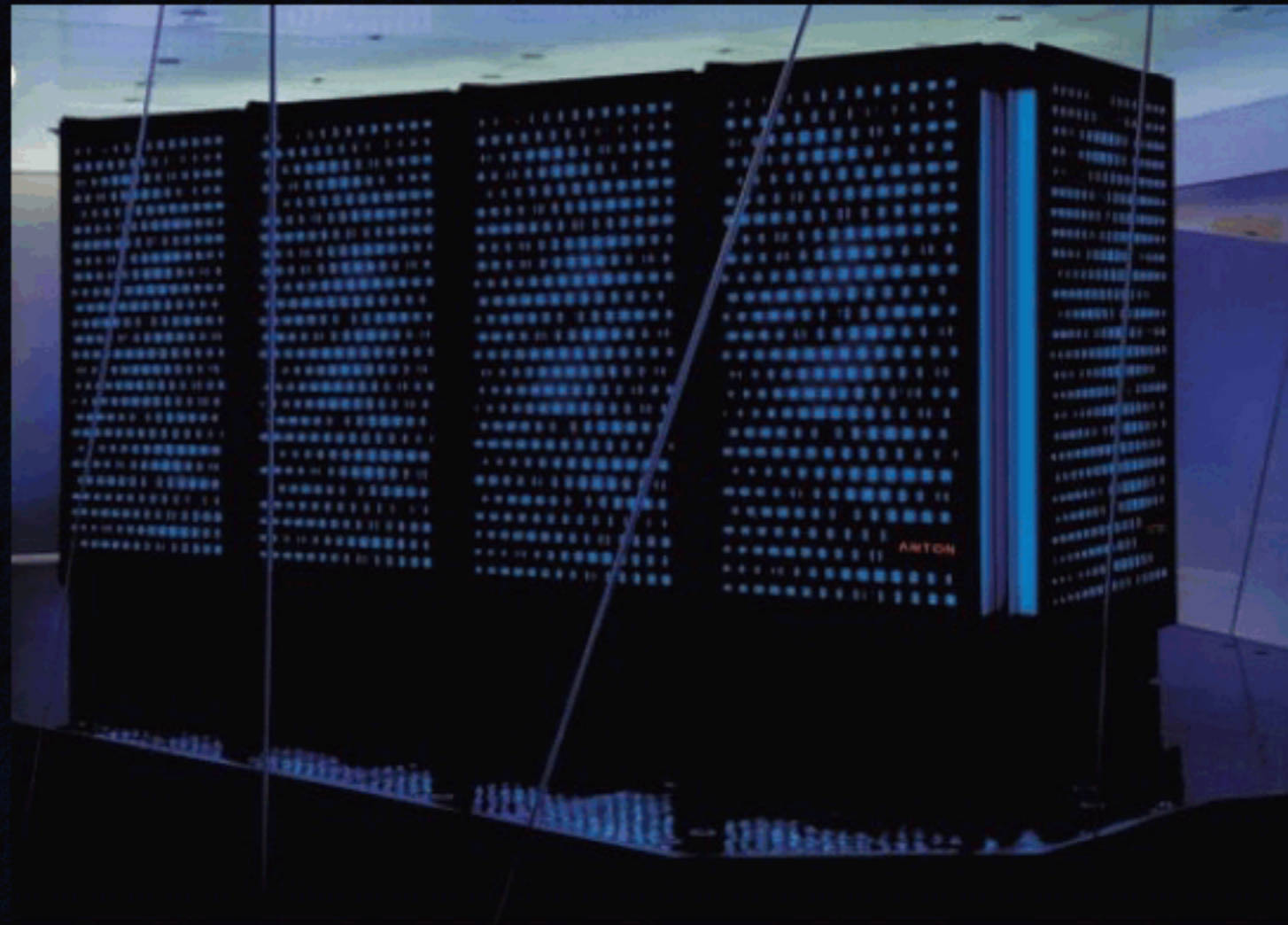
Μοριακή δυναμική

The problem : Too many atoms, too little time



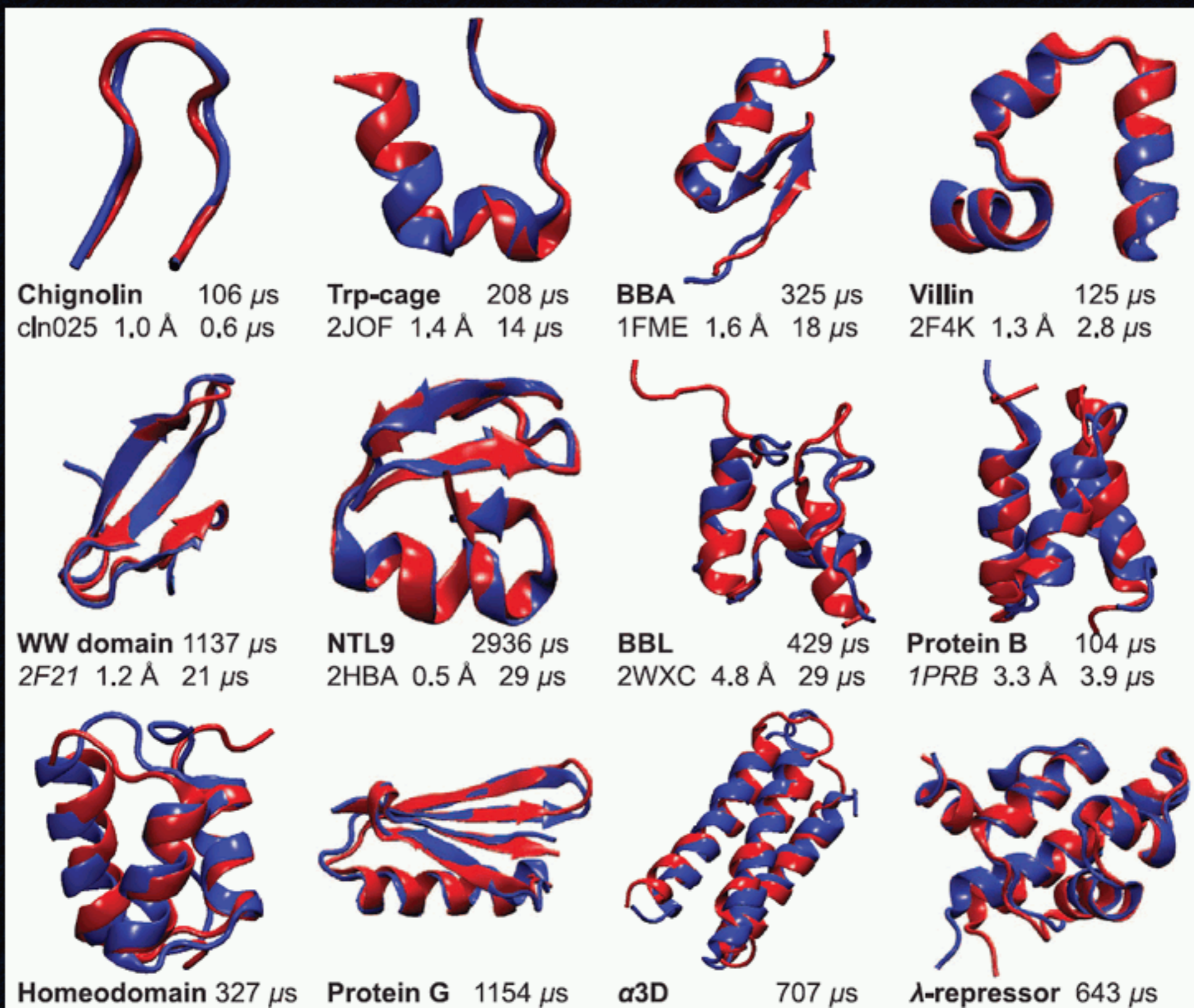
Physics wins ?

The Anton computing machine ...



Physics wins ?

... and its computationally folded proteins.



Physics wins ?

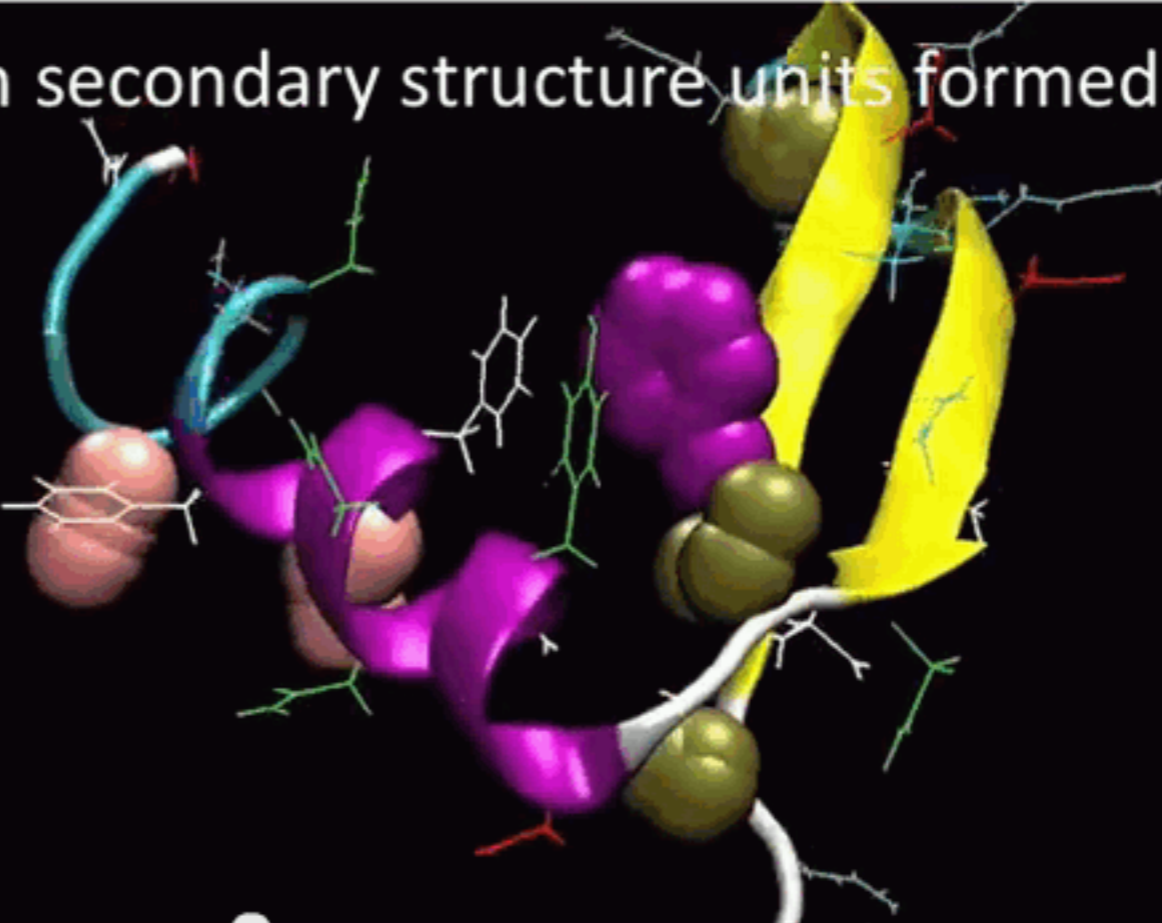
Protein folding in silico : a short video

www.youtube.com/watch?v=gFcp2Xpd29l

YouTube

GUIDE

Clean secondary structure units formed



0:47 / 2:47

Simulation of millisecond protein folding: NTL9 (from Folding@home)

Pande Lab Science · 40 videos

Subscribe 445

57,577

139 2