

# Βιοπληροφορική

# 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

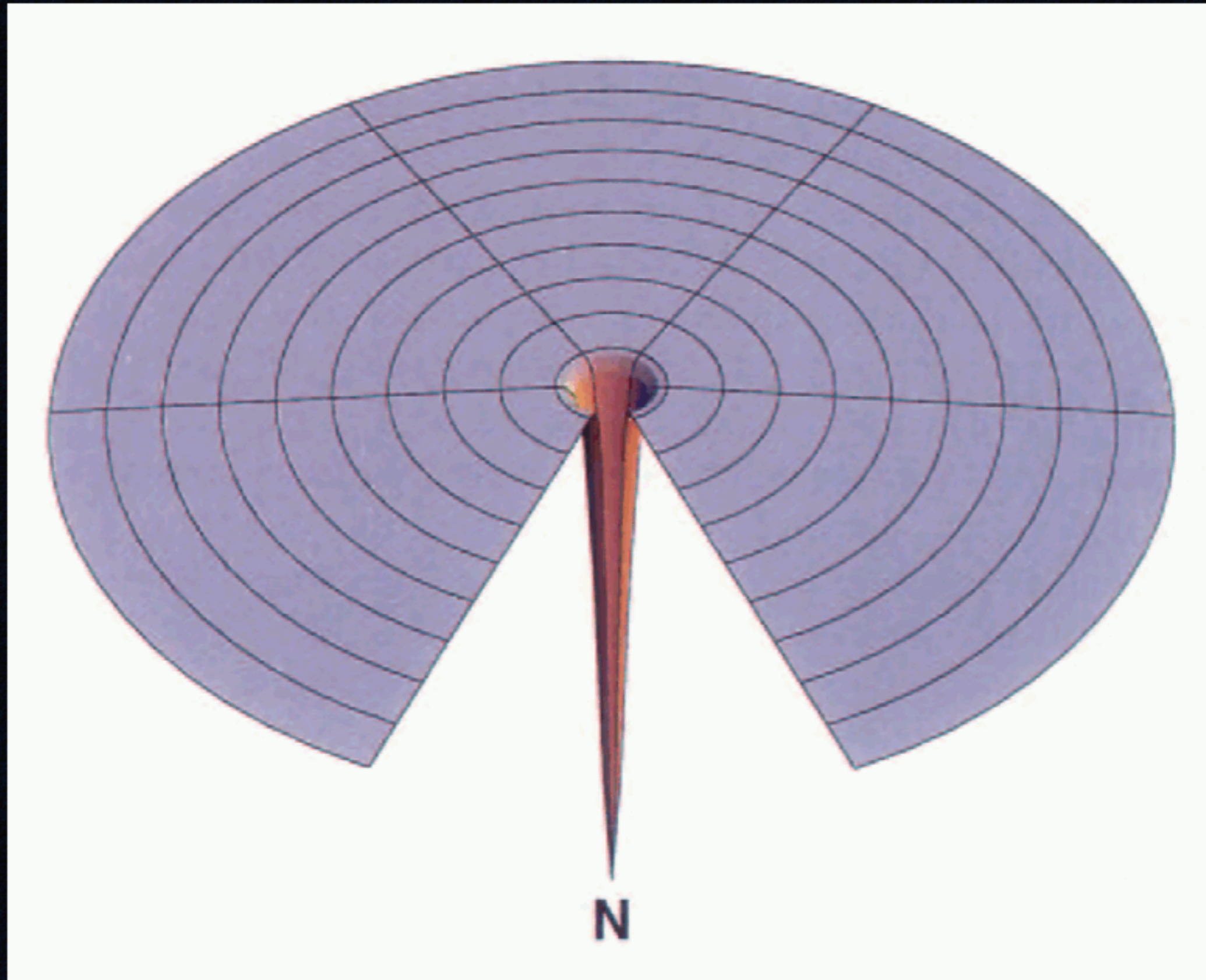
Υποθέστε ότι για κάθε κατάλοιπο υπάρχουν μόνο 10 ενεργειακά αποδεκτοί συνδυασμοί των φ,ψ γωνιών. Τότε, ο ολικός αριθμός πιθανών δομών για μια αλυσίδα 40 αμινοξέων είναι  $10^{40}$ . Ακόμα και εάν μια πρωτεΐνη μπορούσε να περάσει από  $10^{14}$  (100 τρις) δομές ανά δευτερόλεπτο, πάλι θα χρειαζόνταν  $\sim 67$  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

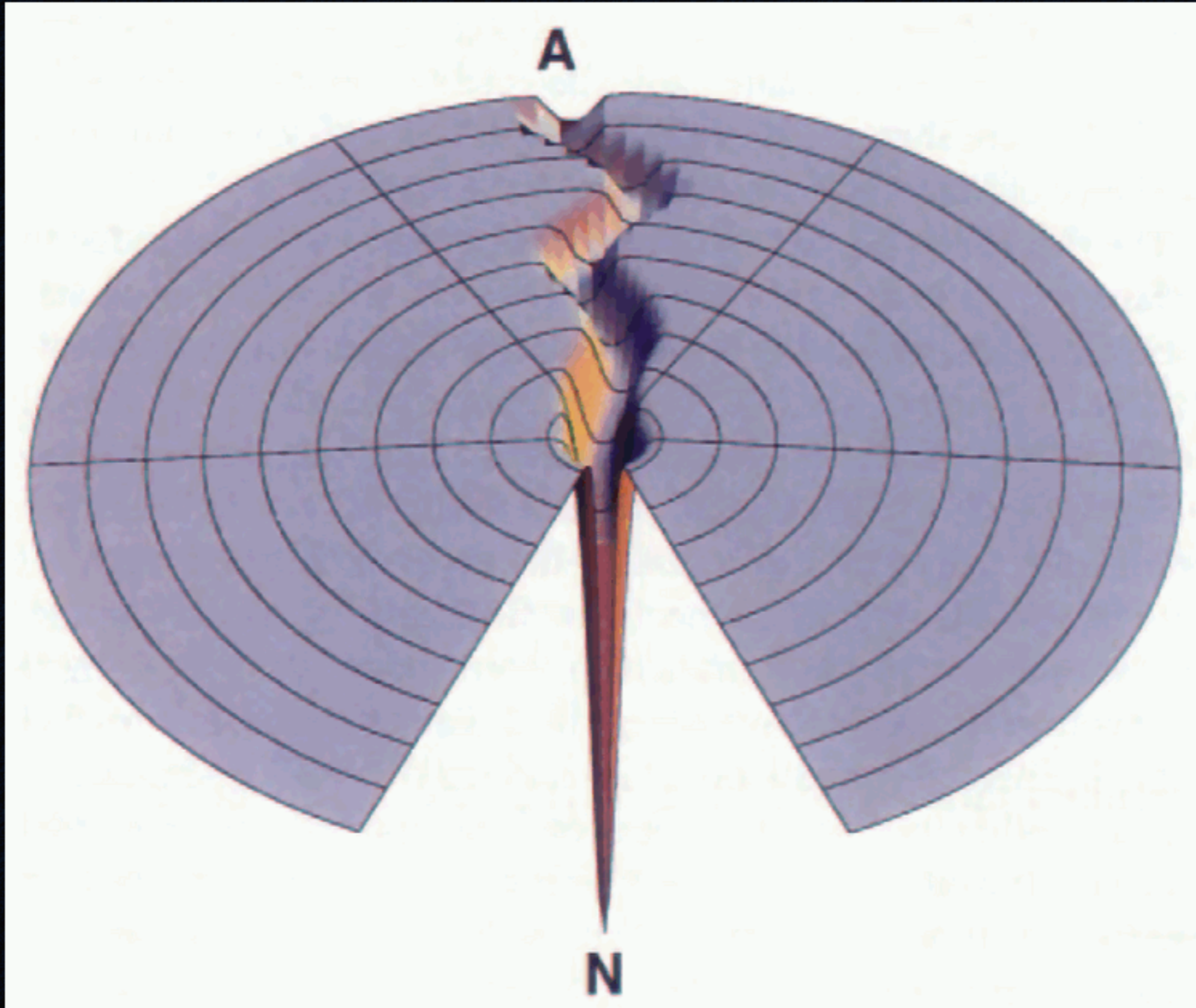
# Το πρόβλημα της αναδίπλωσης των πρωτεϊνών

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

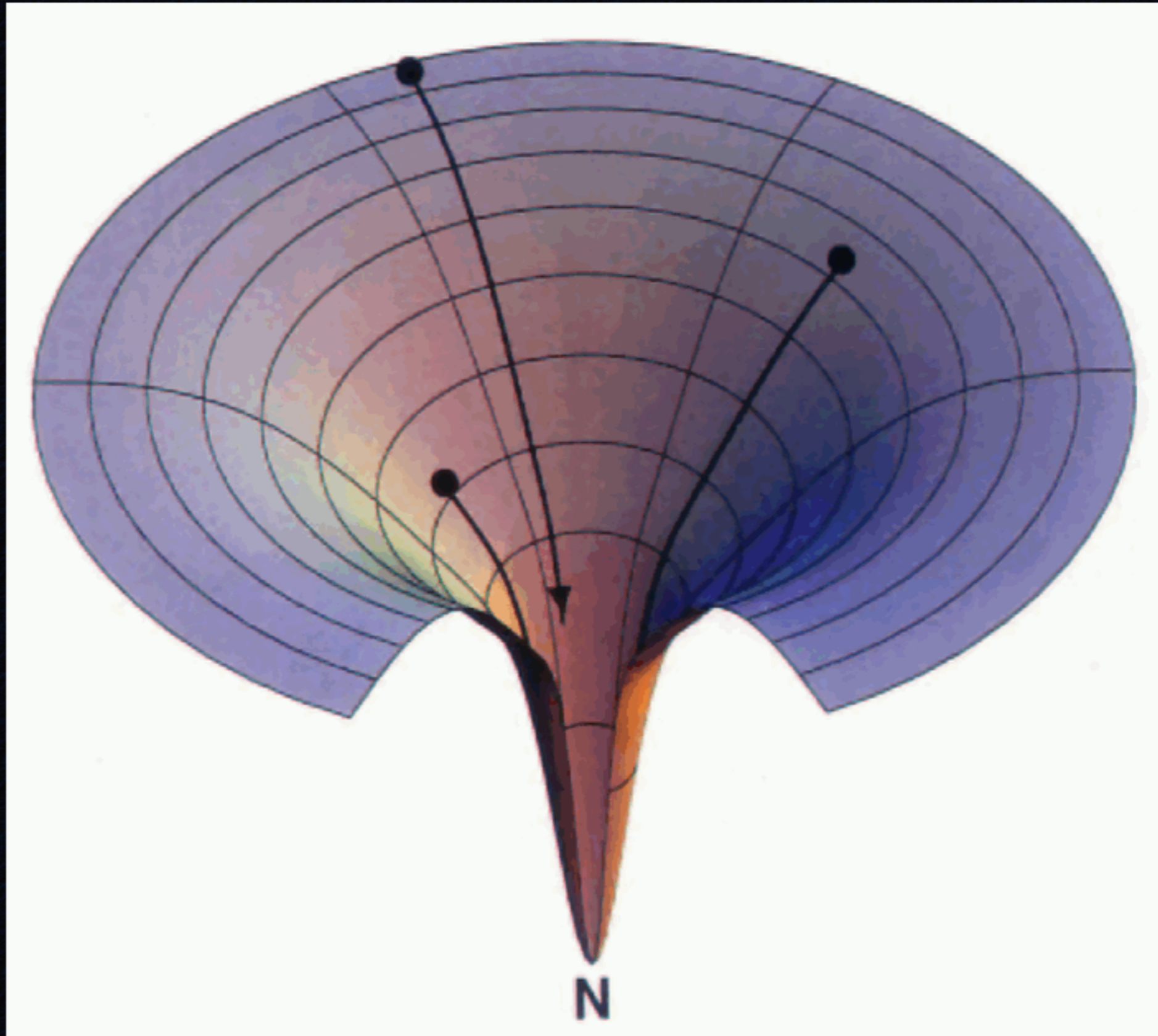


# Energy landscape : pathways

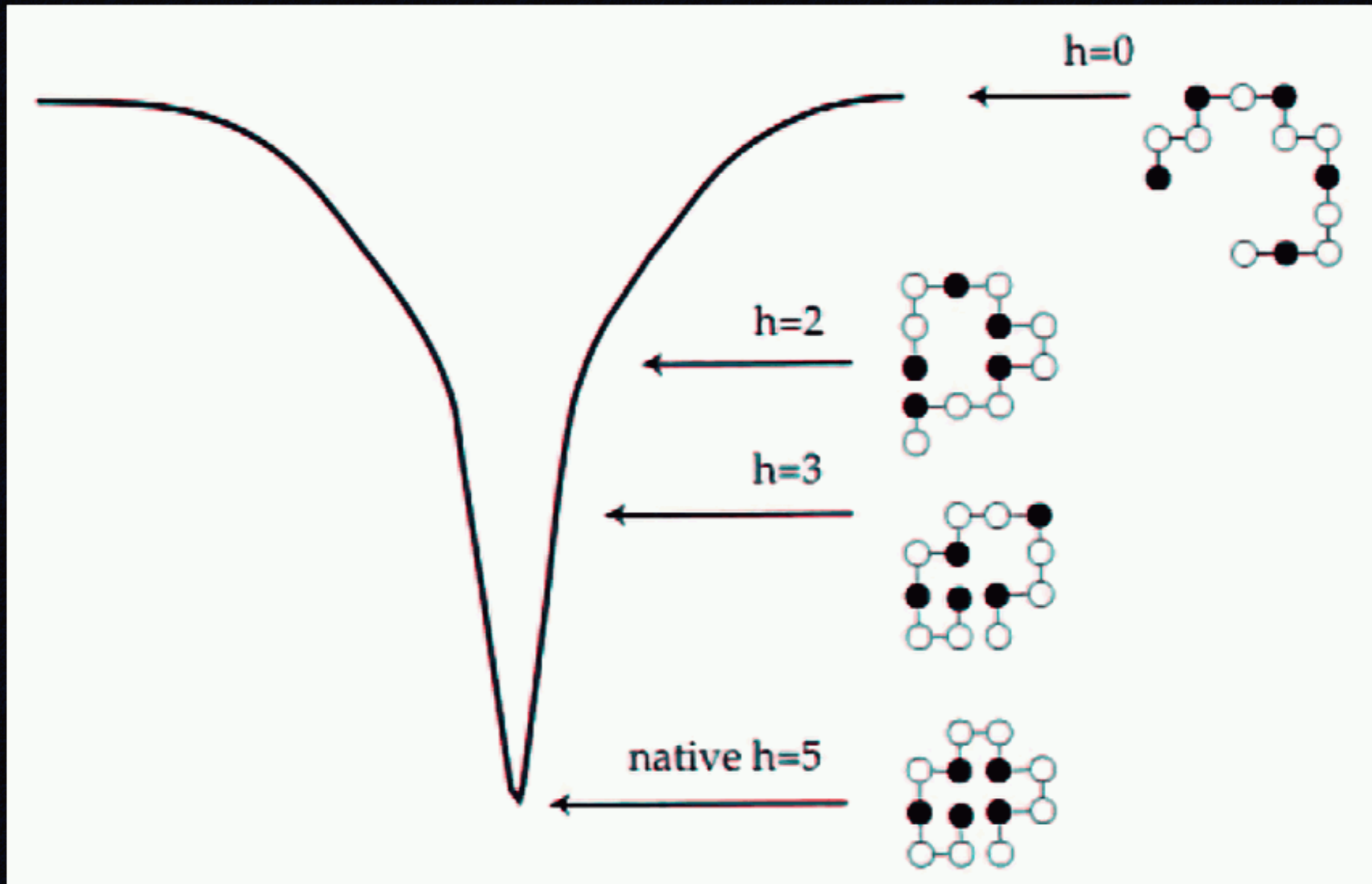




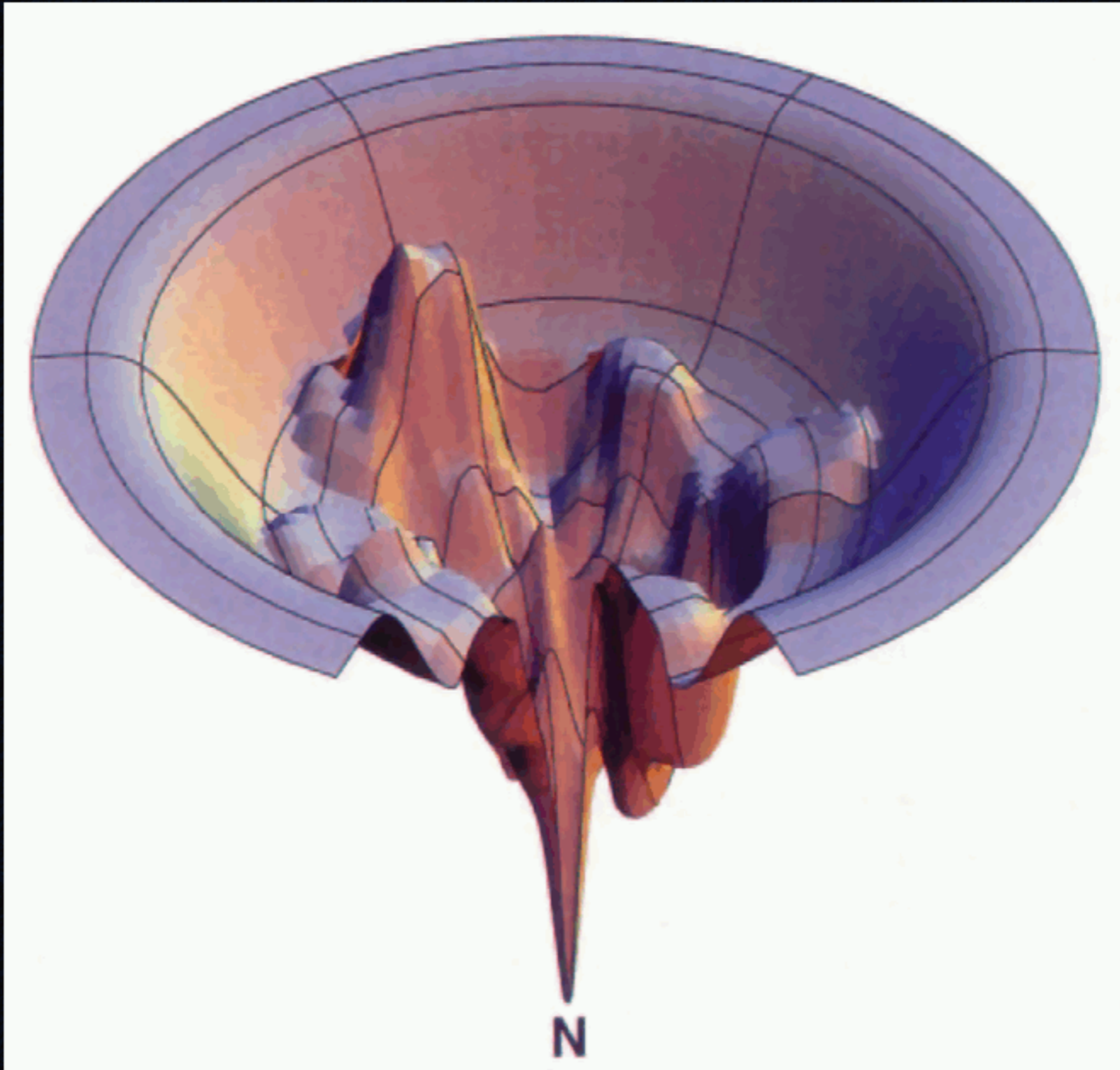
# Energy landscape : the funnel



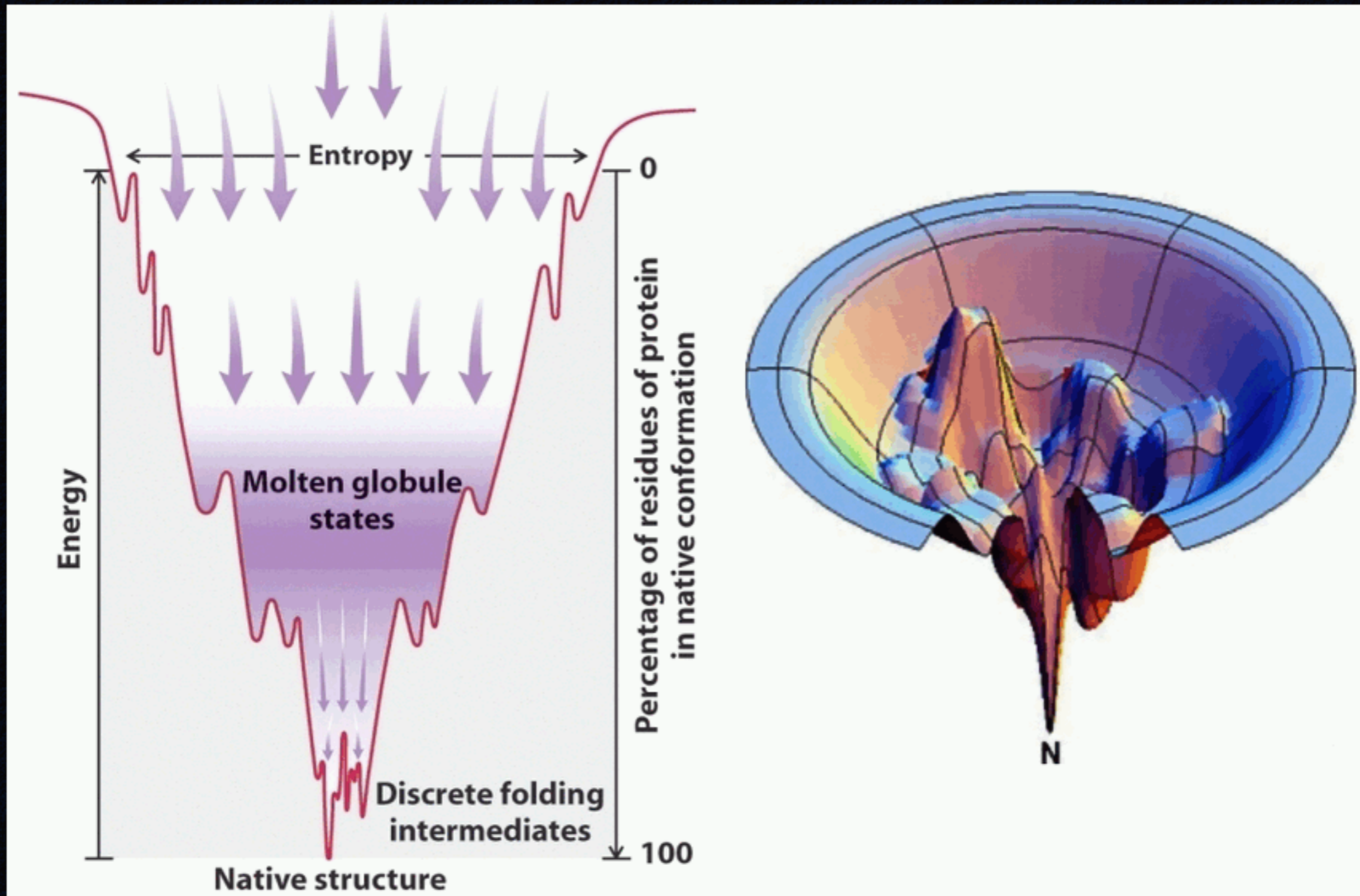
# Energy landscape : The connection with evolution



# 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

---

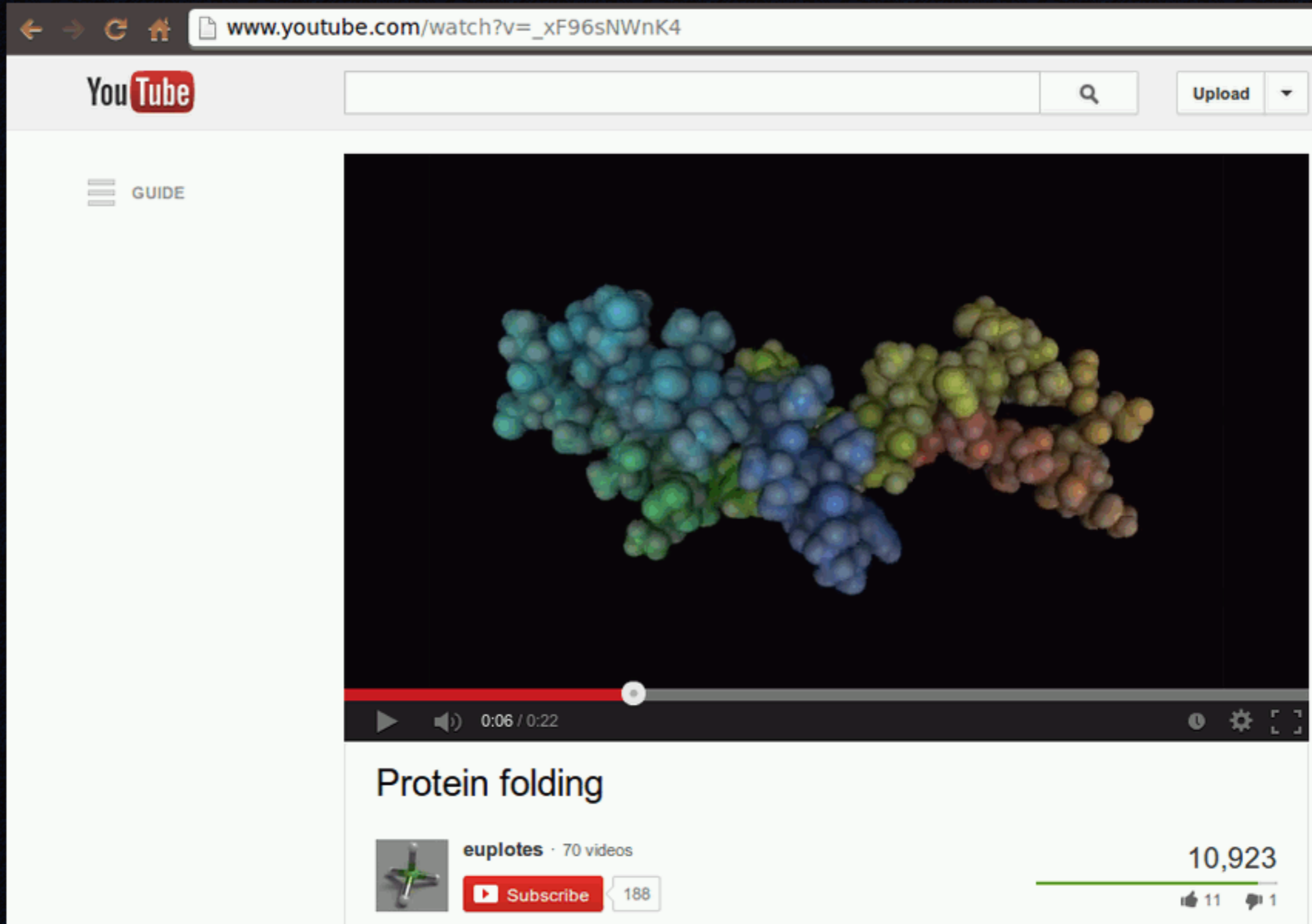
$$\mathcal{H} = \mathcal{K} + \mathcal{V}$$

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

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

# Μοριακή δυναμική

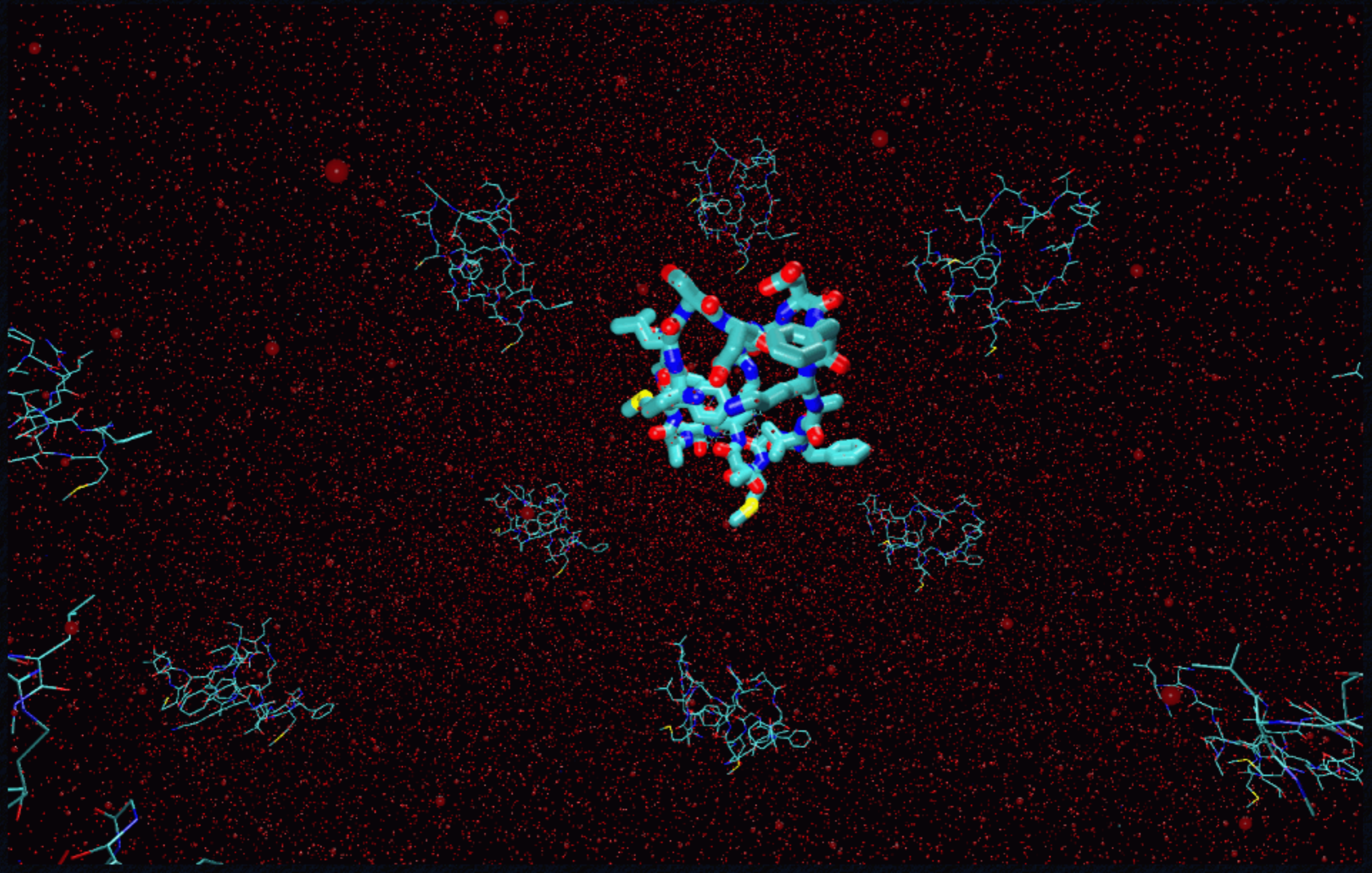
Molecular motion illustrated : a short video.



The image shows a screenshot of a YouTube video player. The browser address bar at the top displays the URL [www.youtube.com/watch?v=\\_xF96sNWnK4](http://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 player, the title "Protein folding" is displayed. The channel name "euplotes" is shown with a profile picture of a green and white molecular structure and the text "70 videos". A red "Subscribe" button is next to the channel name, with a "188" notification badge. The video has 10,923 views, 11 likes, and 1 dislike. The video progress bar at the bottom indicates the video is at 0:06 out of 0:22.

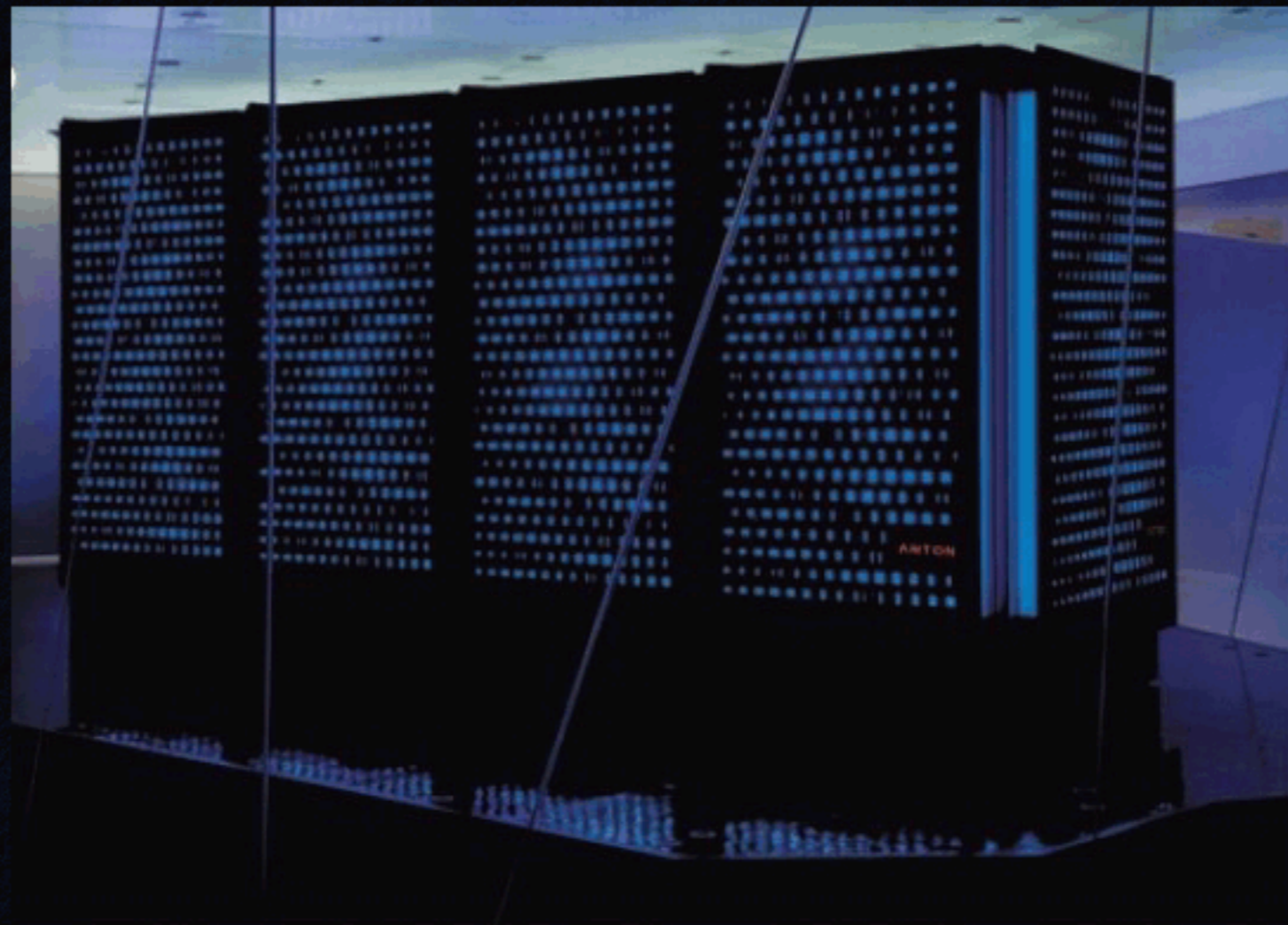
# Μοριακή δυναμική

The problem : Too many atoms, too little time



# Physics wins ?

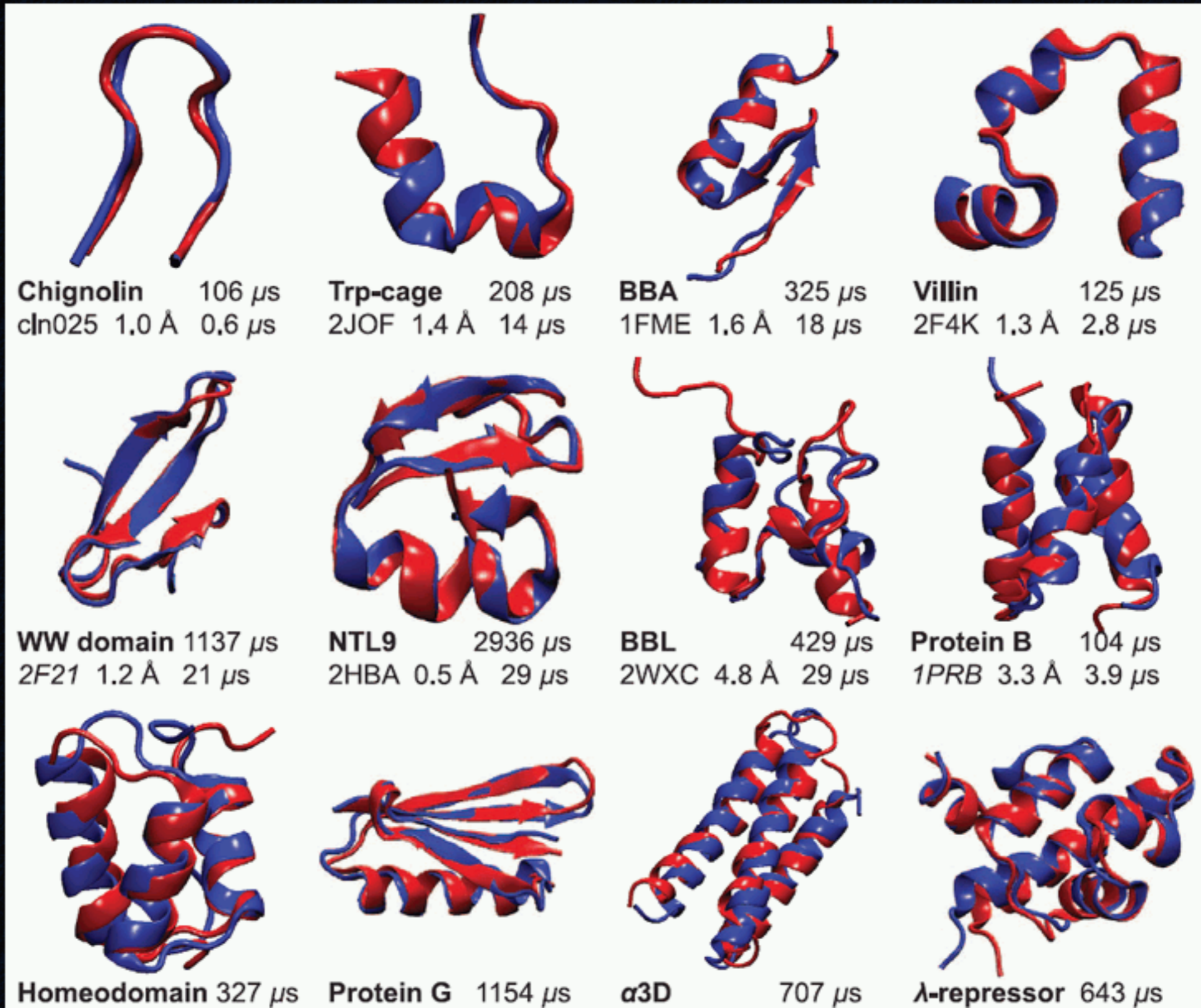
The Anton computing machine ...





# Physics wins ?

... and its computationally folded proteins.



# Physics wins ?

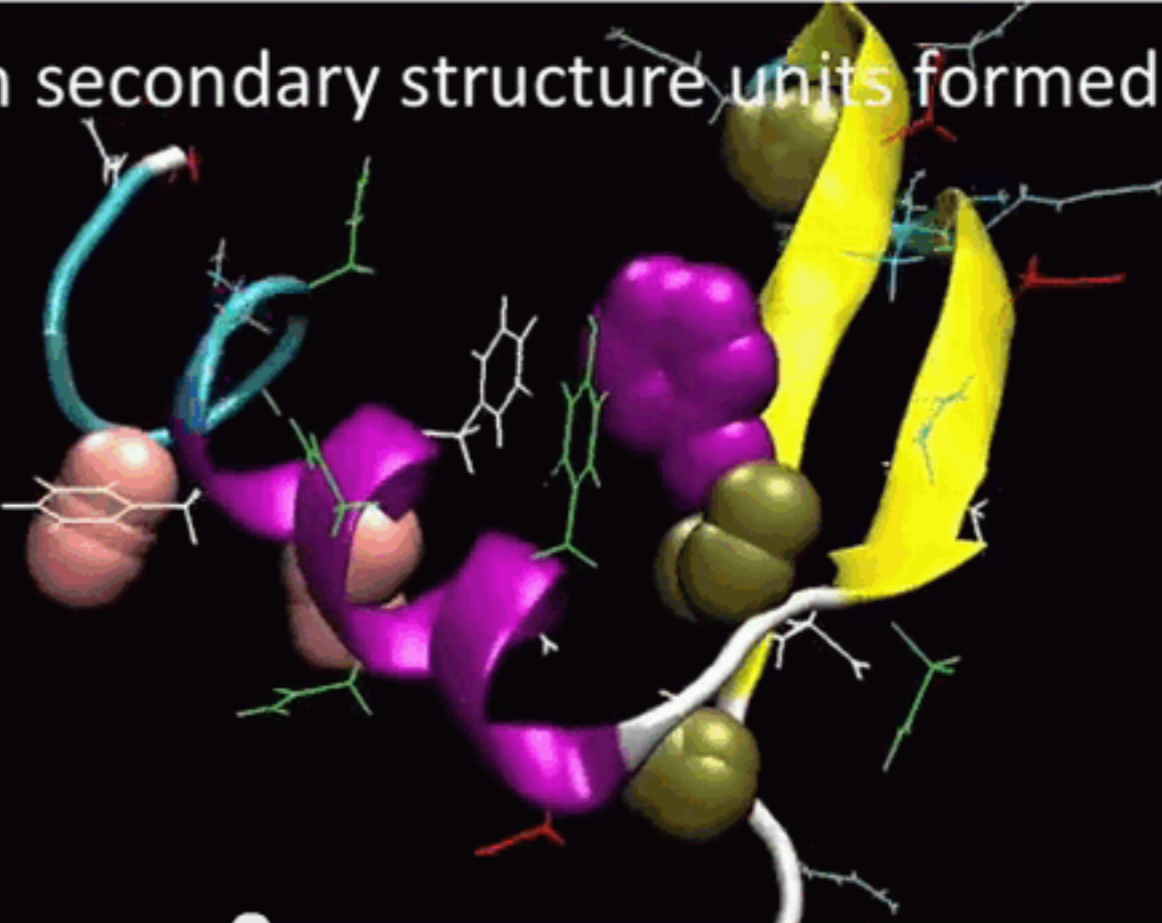
## Protein folding in silico : a short video

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