PROTEIN FOLDING: A PARADIGM FOR SOLVING HARD PROBLEMS IN BIOLOGY

Michael Levitt
Structural Biology & Computer Science
Stanford

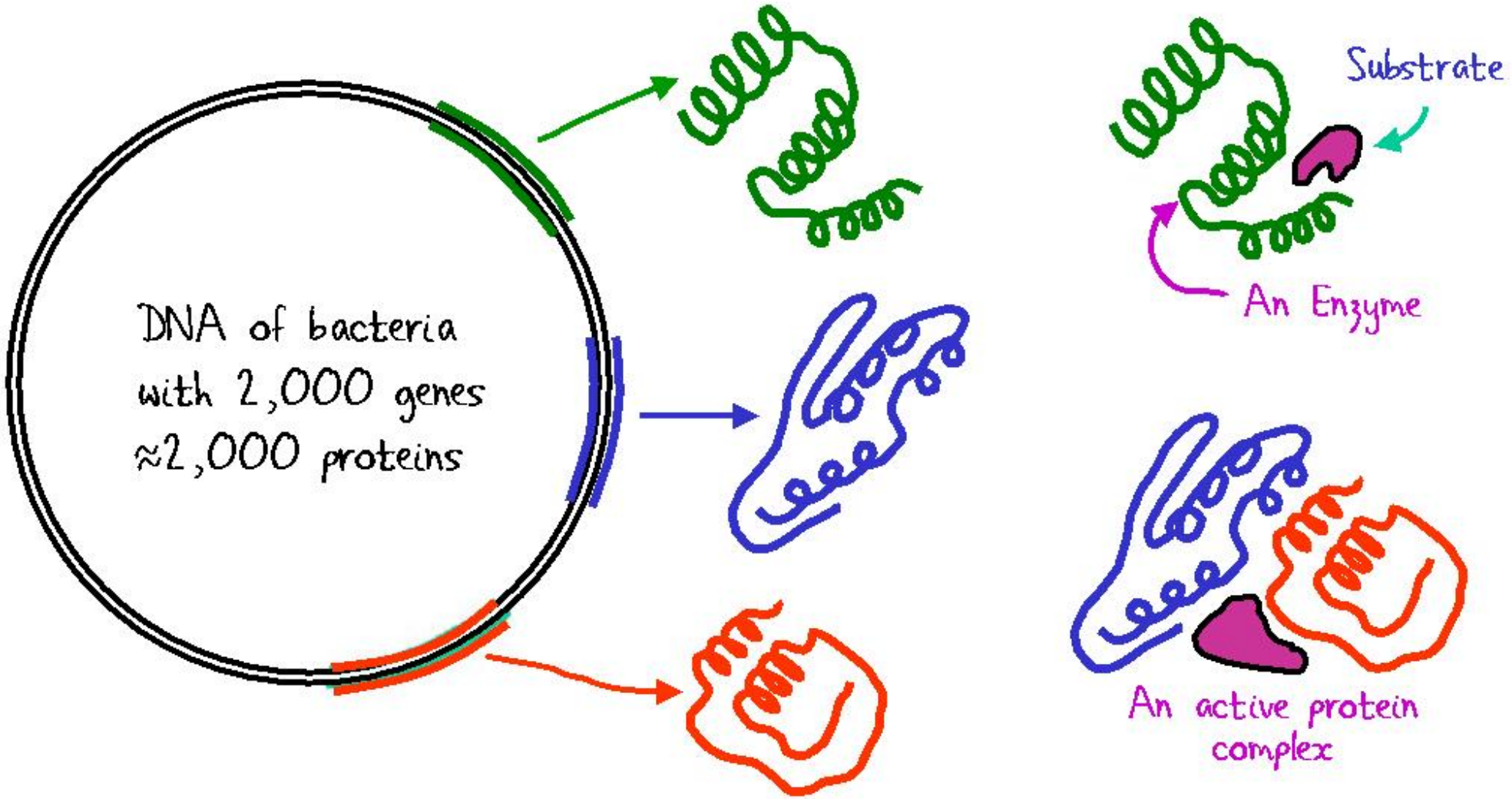
http://csb.stanford.edu/levitt

OUTLINE

- Simulation
  - Basic methods
  - Hydrophobic effect
  - Unfolding, folding
- Prediction
  - Special potentials
  - Minimization
  - Monte Carlo
- Hard Problems

INTRODUCTION

STRUCTURAL OVERVIEW OF BIOLOGY

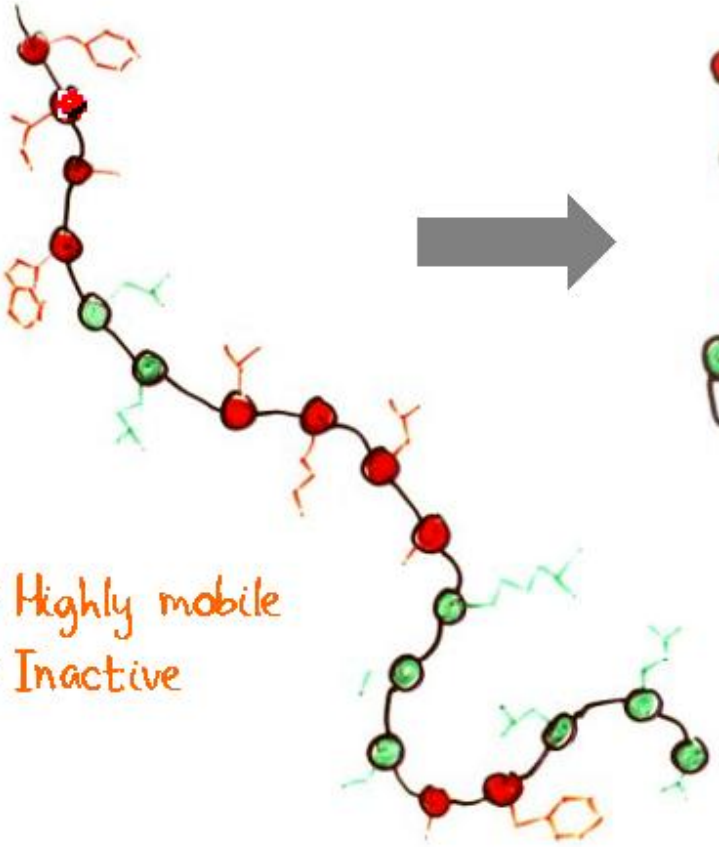


- From DNA sequence, predict all protein structures
- From protein structures predict all function.

PROTEIN FOLDING IS CENTRAL

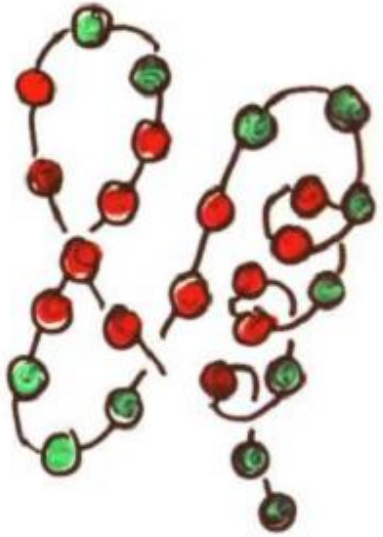
Sequence → Structure → Function

- Unfolded protein is a chain of amino acids



- Highly mobile
- Inactive

- Folded protein



- Almost unique shape
- Precisely ordered
- Stable
- Active

- Function depends on protein shape

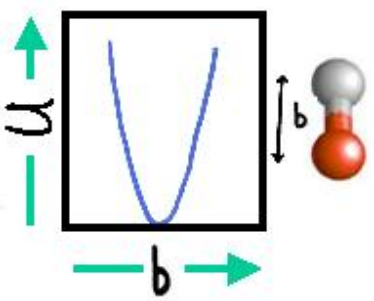
- Specific associations
- Precise reactions

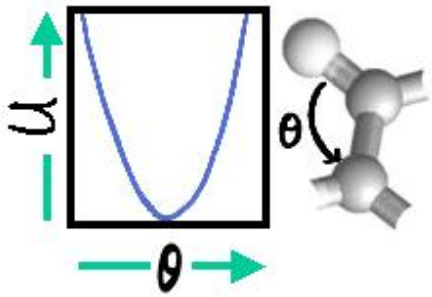
SIMULATION

TOTAL POTENTIAL ENERGY

$$U = \sum \tfrac{1}{2} K_b (b - b_o)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

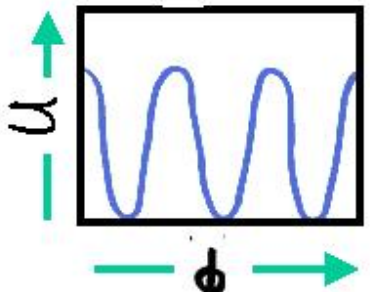
All Bonds — All Angles

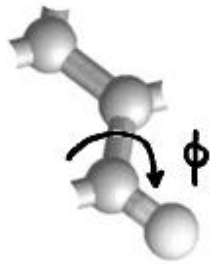




$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

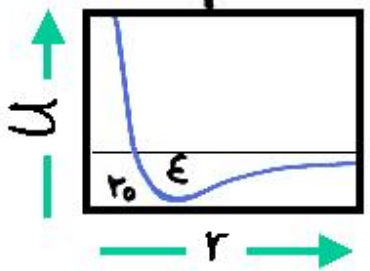
All Torsion Angles

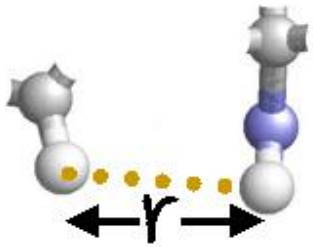




$$+ \sum \varepsilon \left[\left(r_o / r \right)^{12} - 2 \left(r_o / r \right)^{6} \right]$$

All nonbonded pairs

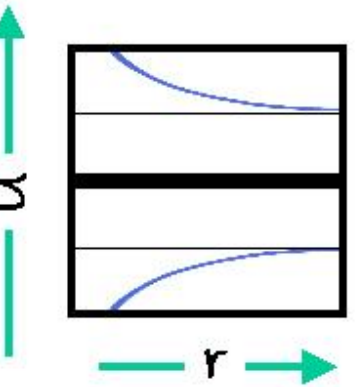


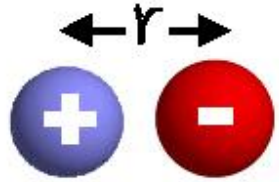


$$+ \sum 332 q_i q_j / r$$

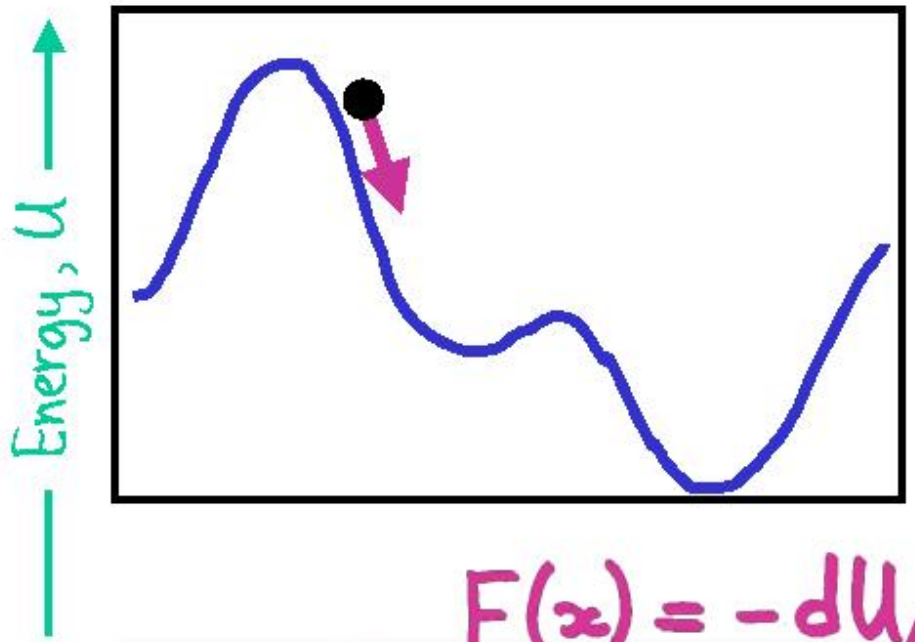
All partial charges

ENCAD. Parameters from 1979 (Lifson)

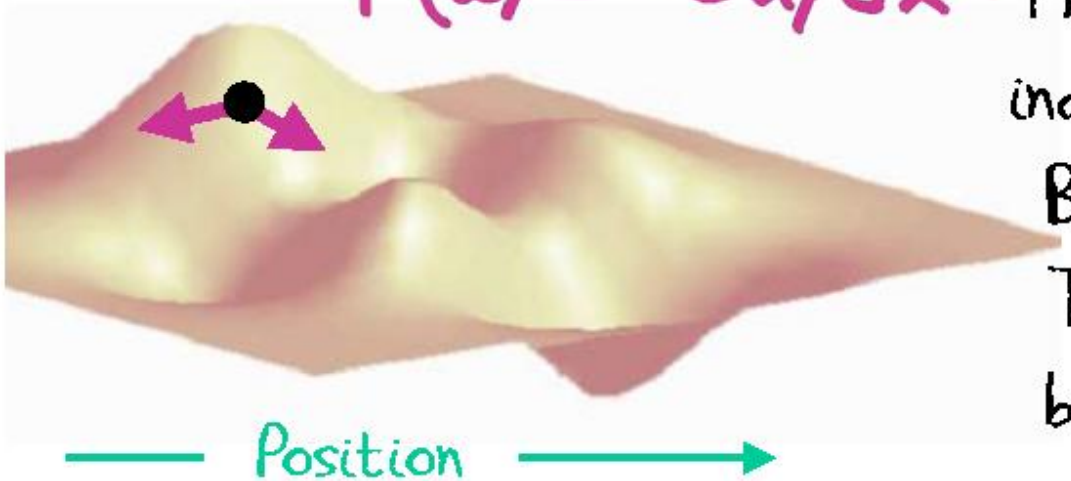




TOTAL POTENTIAL ENERGY . 2

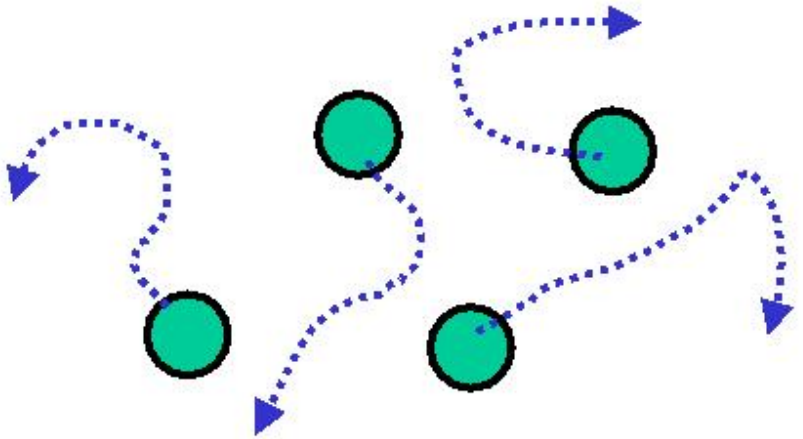


$$F(x) = -dU/dx$$



- The total potential energy or enthalpy fully defines the system, U.
- The forces are the gradients of the energy.
- The energy is a sum of independent terms for: Bonds, Bond angles, Torsion angles and non-bonded atom pairs.

MOLECULAR DYNAMICS THEORY



- Force = $-dU/dx$ (slope of potential, U); acceleration, $m\,a(t)$ = Force.
- All atoms move together so force between atoms change with time.
- Analytical solution for $x(t)$ and $v(t)$ is impossible; numerical solution is trivial.

$$x(t+\Delta t) = x(t) + v(t)\Delta t + [4a(t) - a(t-\Delta t)]\,\Delta t^2/6$$

New position — Old position — Old velocity — Acceleration

$$v(t+\Delta t) = v(t) + [2a(t+\Delta t) + 5a(t) - a(t-\Delta t)]\,\Delta t/6$$

New velocity — Old velocity — Acceleration

$$U_{kinetic} = \frac{1}{2}\sum m_i v_i(t)^2 = \frac{1}{2} n k_B T$$

Kinetic energy — Atomic masses, velocities — Number of coordinates (not atoms) — Boltzmann's Constant — Temperature

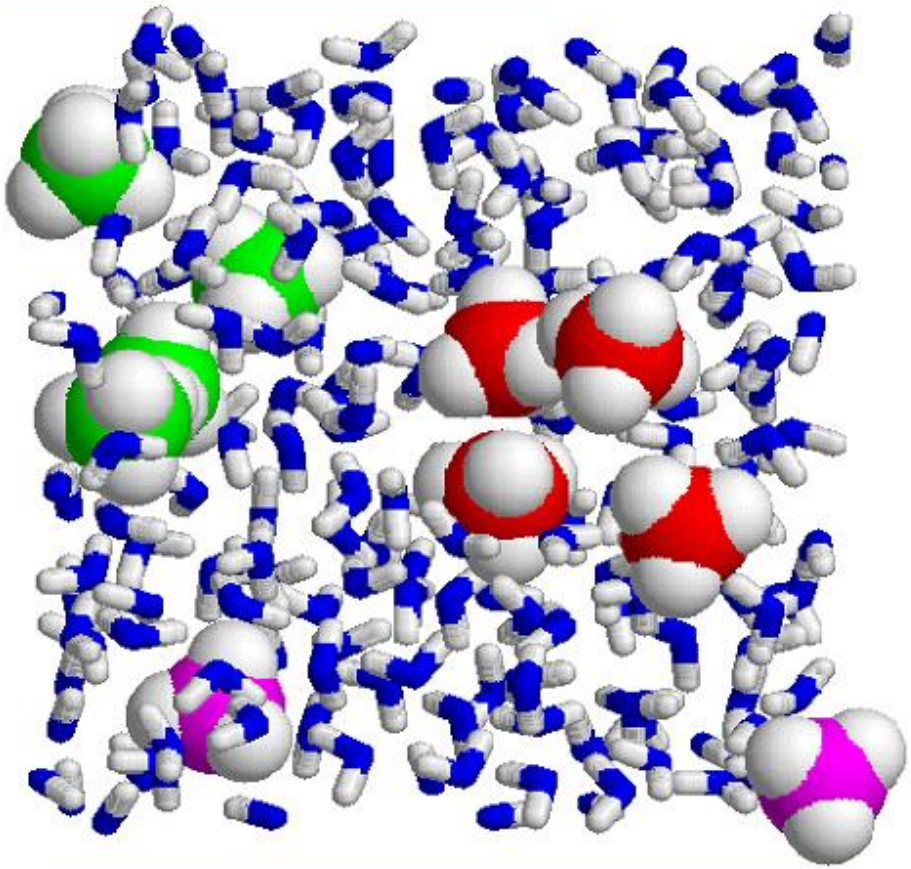
Time step, Δt, must be very small at 10^{-15} seconds or 0.001 ps.

Total energy ($U_{potential} + U_{kinetic}$) must not change with time

HYDROPHOBIC EFFECT

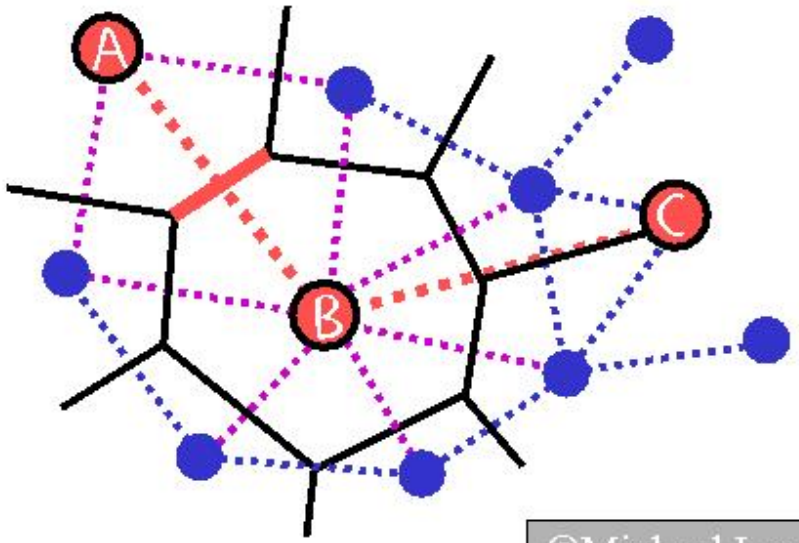
SIMULATING THE HYDROPHOBIC EFFECT

Tanya Raschke



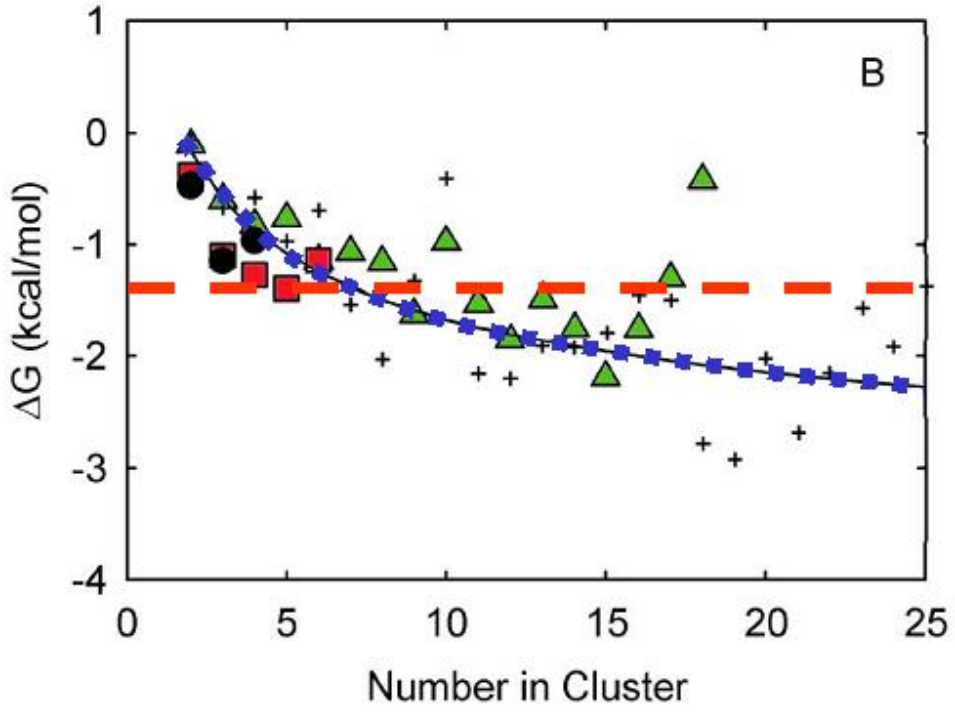
Box with periodic boundaries.

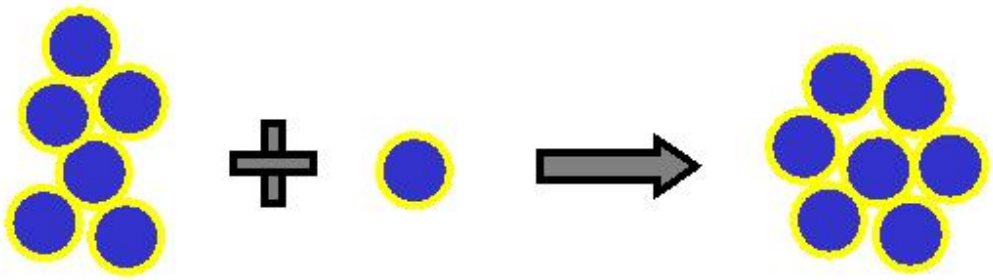
- 1 nanosecond MD simulations in periodic water boxes with from 30mM to 3 Molar hydrocarbon solution.
  Encad with F3C water (1996).
- Measure cluster formation by Voronoi.
  $d(AB) = d(BC)$, but only A, B touch.



MOVIE OF BENZENE MOLECULAR DYNAMICS IN WATER AT ROOM TEMPERATURE

HYDROPHOBIC ENERGY IS COOPERATIVE





N−1 + 1 → N

- $\Delta G_N = -kT \log[C_N/(C_{N-1}\, C_1)]$
- Assume clusters are close-packed spheres:

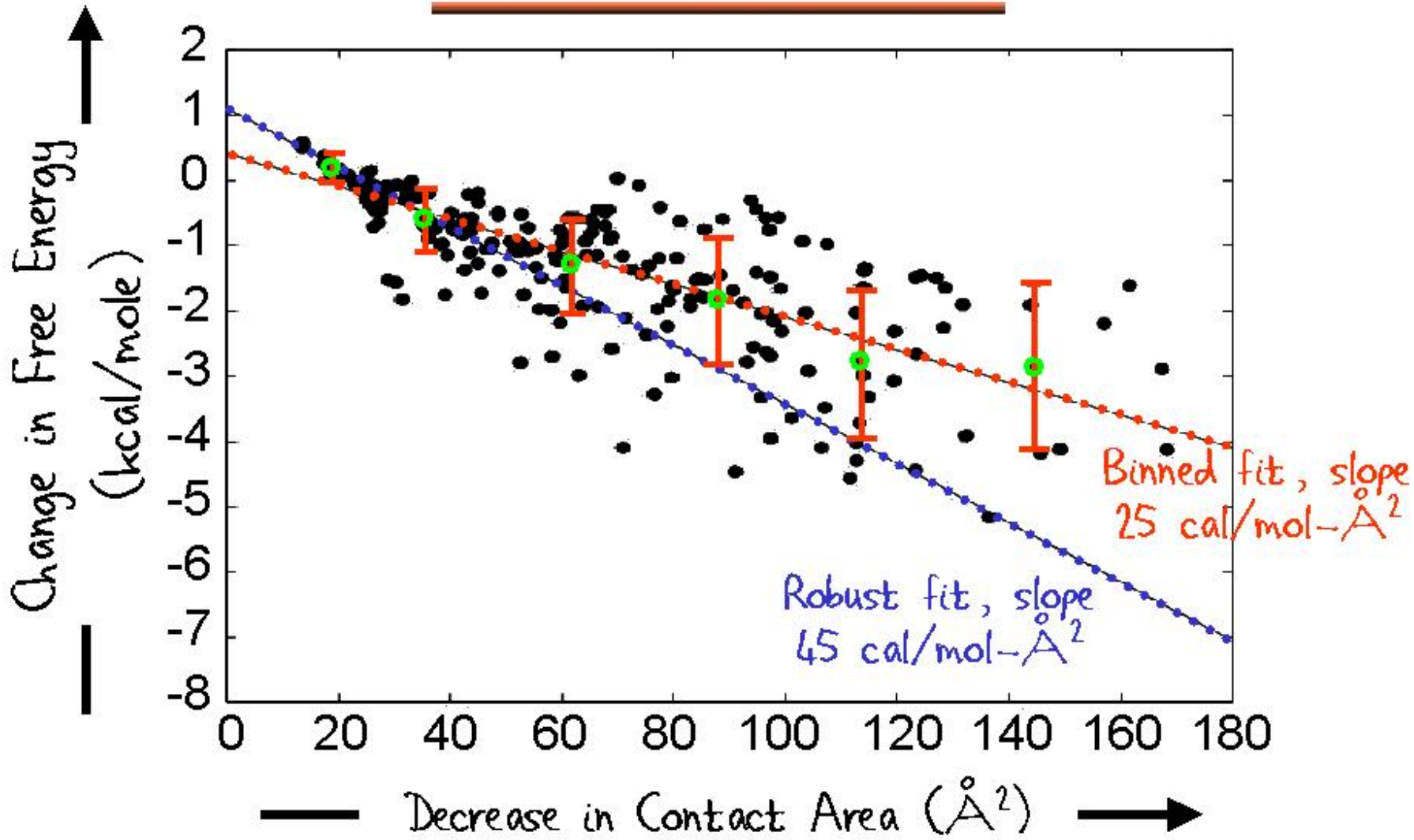
  $V_N = NV_1$

  $A_N = \alpha(V_N)^{2/3} = \beta(N)^{2/3}$

  $\Delta A_N = \beta[(N)^{2/3} - (N-1)^{2/3}]$
- If $\Delta G_N = \gamma \Delta A_N$, then

  $\Delta G_N = \gamma\beta[(N)^{2/3} - (N-1)^{2/3}]$
- Determine γ by fitting with $[(N)^{2/3} - (N-1)^{2/3}]$.

HYDROPHOBIC ENERGY DEPENDS ON BURIED SURFACE

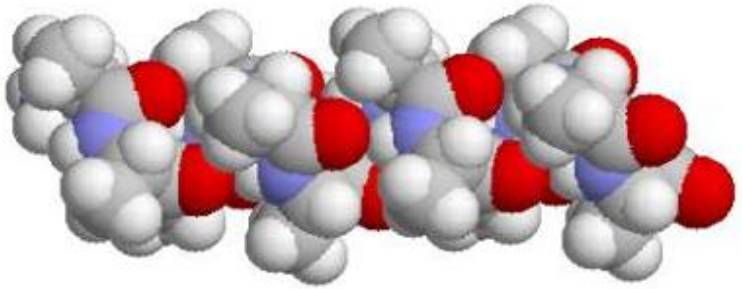


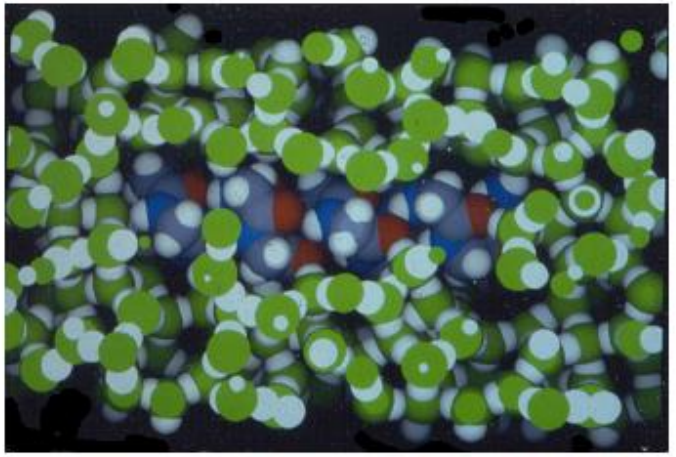
Constant of proportionality matches experiment.

SIMULATE UNFOLDING

UNFOLD THE α-HELIX

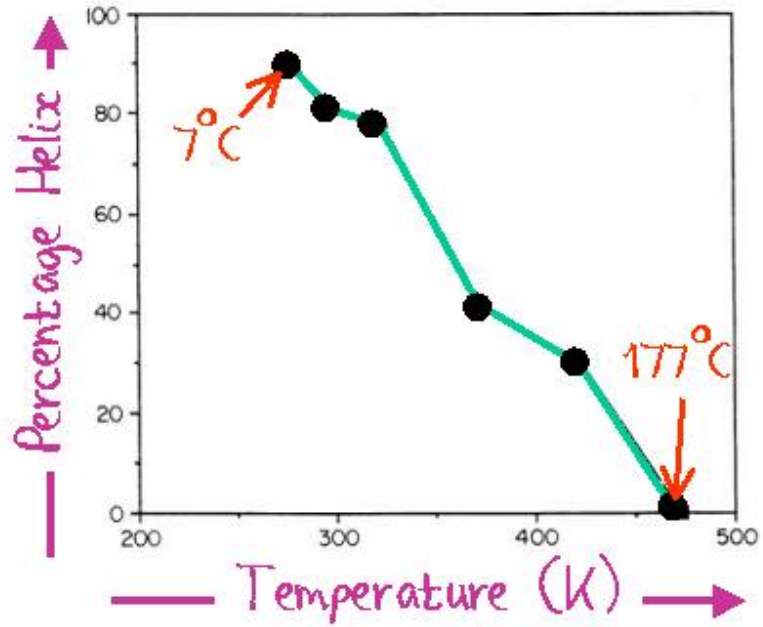
13 Alanine residues



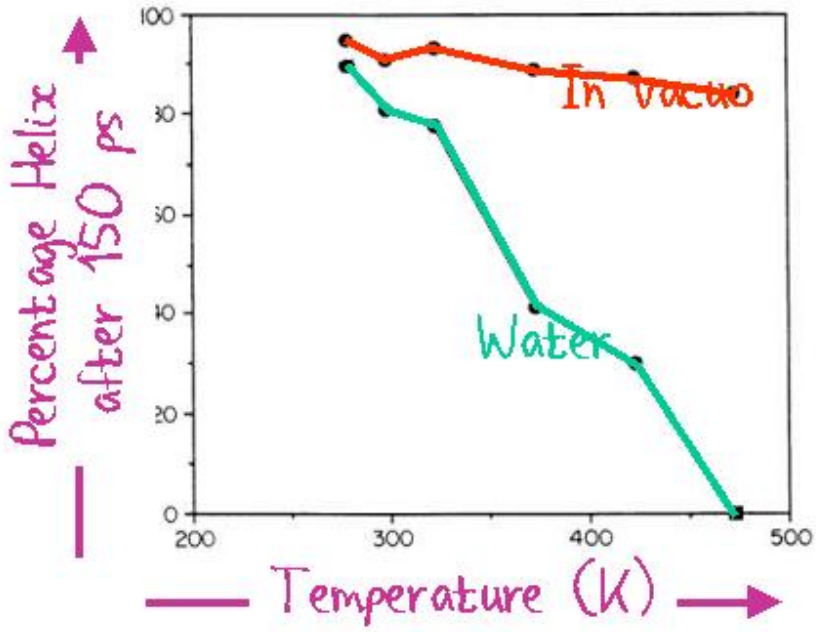


Put it in a box of water.

- Start as an ideal α-helix in a box of water.
- Run 200 ps (100,000 time steps) of molecular dynamics at six different temperatures.
- Record percentage α-helix formed for last 50 ps.
- See temperature-induced melting on picosecond time-scale.



α-HELIX LESS STABLE IN WATER

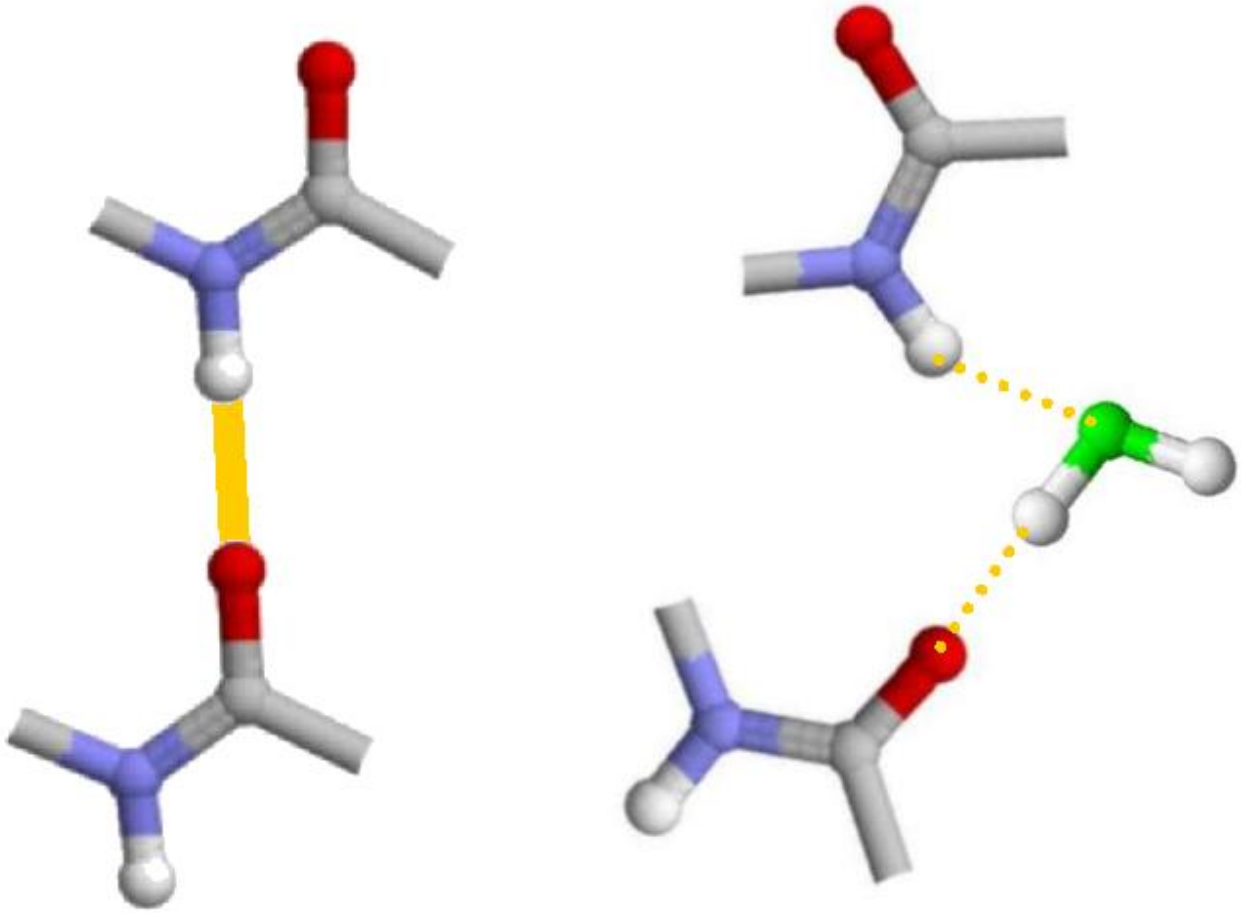


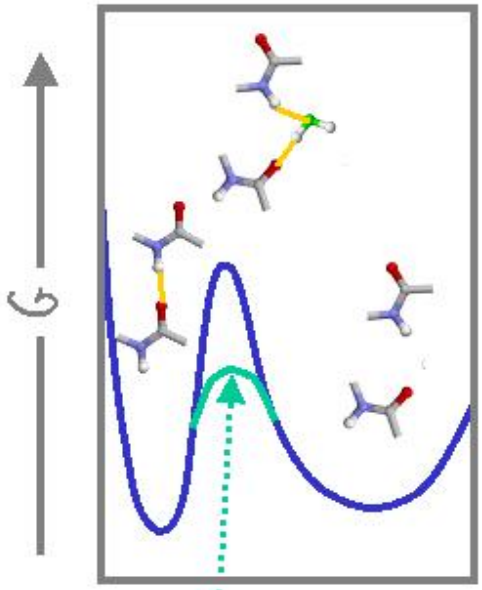
- In vacuo the helix is very stable even at high temperature.
- In water the helix is unstable at high temperature.
- The rate of melting depends on temperature.
- Happens because water molecules stabilize the transition state.

WATER ALLOWS HYDROGEN BONDS TO BREAK

Intact hydrogen bond in helix

Hydrogen bond is breaking



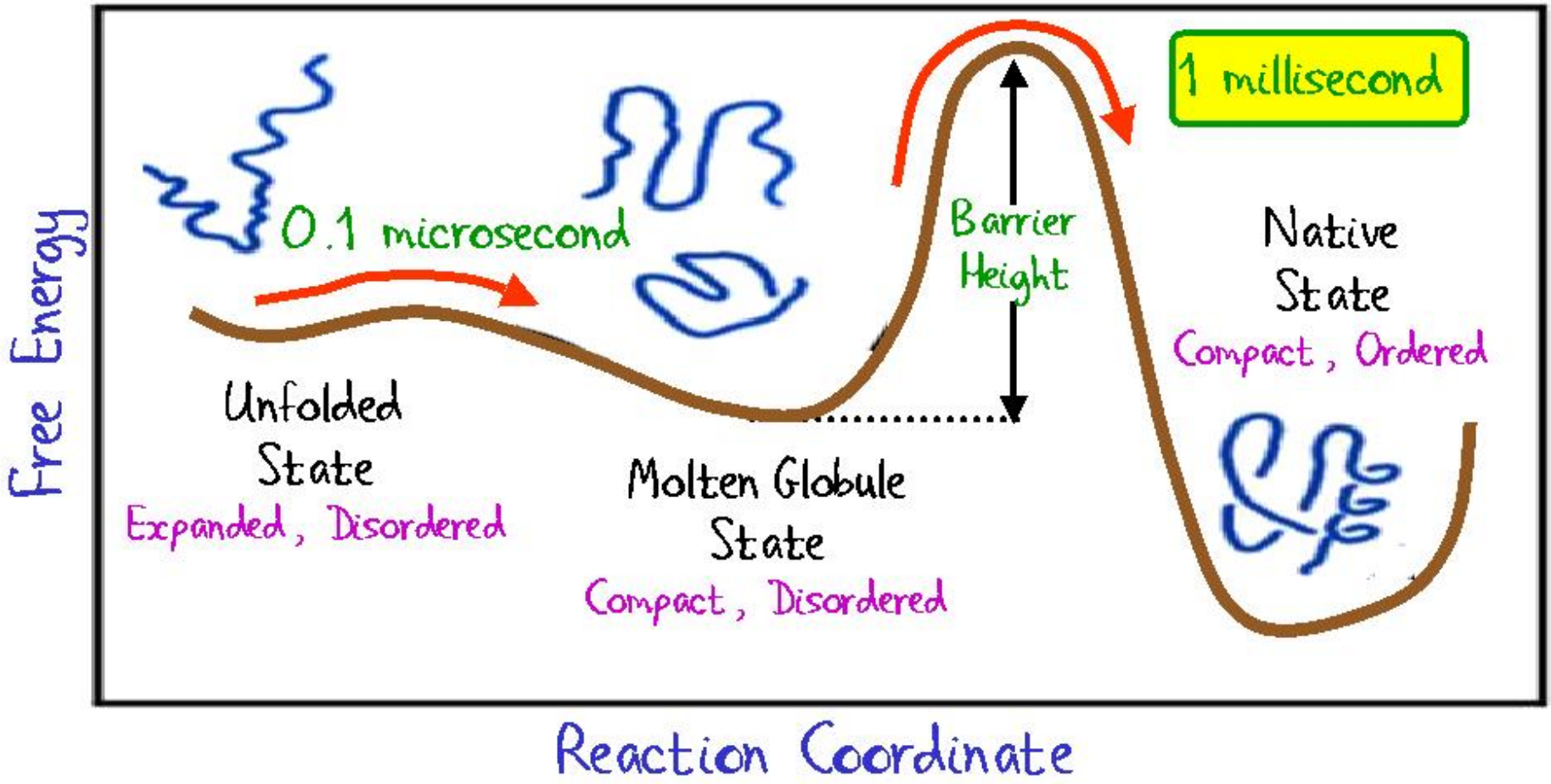


Free Energy barrier between states is much lower in water

- Water catalyzes the breakage of hydrogen bonds by stabilizing the transition state.

SIMULATE FOLDING

SIMULATING FOLDING IS DIFFICULT?



- Simulation of 1 millisecond requires 10,000 CPU years!
- Must get over high barriers & many degrees of freedom.

MASSIVE COMPUTATIONAL RESOURCES

- Empty Supercomputers.
- Blue Gene (IBM).
- Folding@home (Vijay Pande).

FOLDING@HOME

Folding@home
from genome to structure

http://www.stanford.edu/group/pandegroup/Cosm/

Using Folding@home

- Project Goals: solving the protein folding problem
- How you can help
- Download (New!! Version 1.33)
- How to install our software
- Frequently asked questions (FAQ)
- Contact Folding@home (Help Center)
- Folding@home discussion board

Join Folding@home by running our screen saver or client software

- Fold proteins on 20,000 computers using the program as a Screen Saver!

Like SETI@home

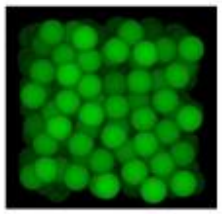
PANDE MOVIE

α-HELIX FOLDING

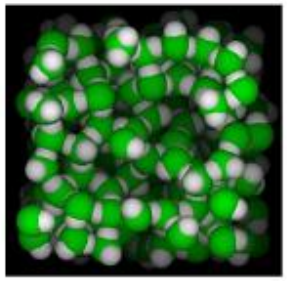
50 YEARS OF SIMULATION

- We have 10,000,000 times more resources.
- Systems have become larger (100 times).
- Runs have become longer (100,000 times).
- Energy functions have become simpler.
- Fit reality well.

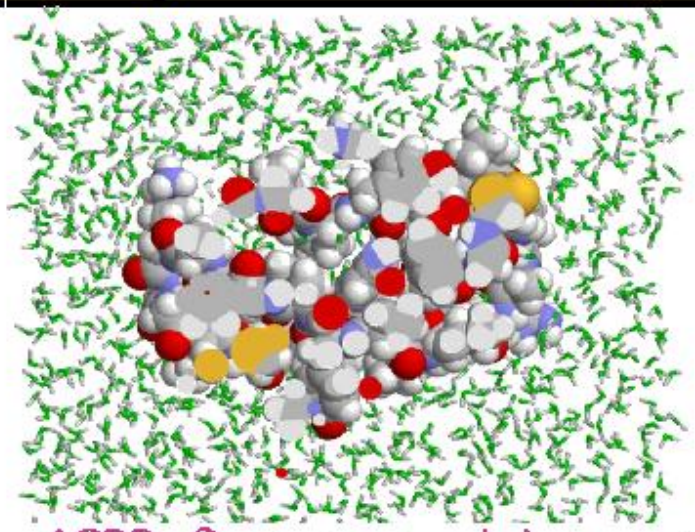
Nothing bad has happened!



1955 Argon



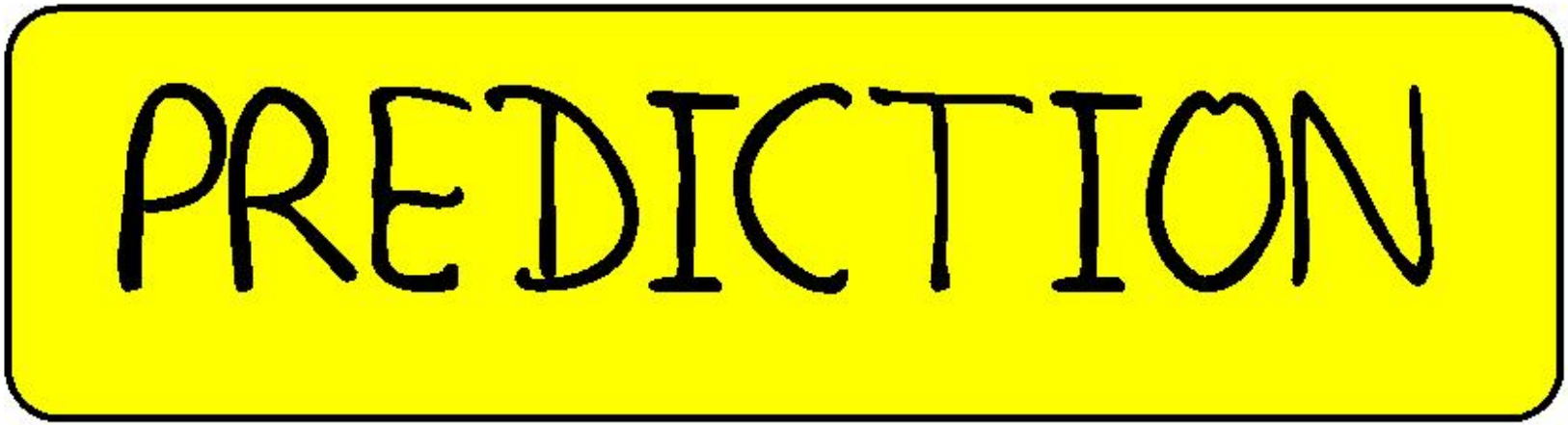
1970 Water



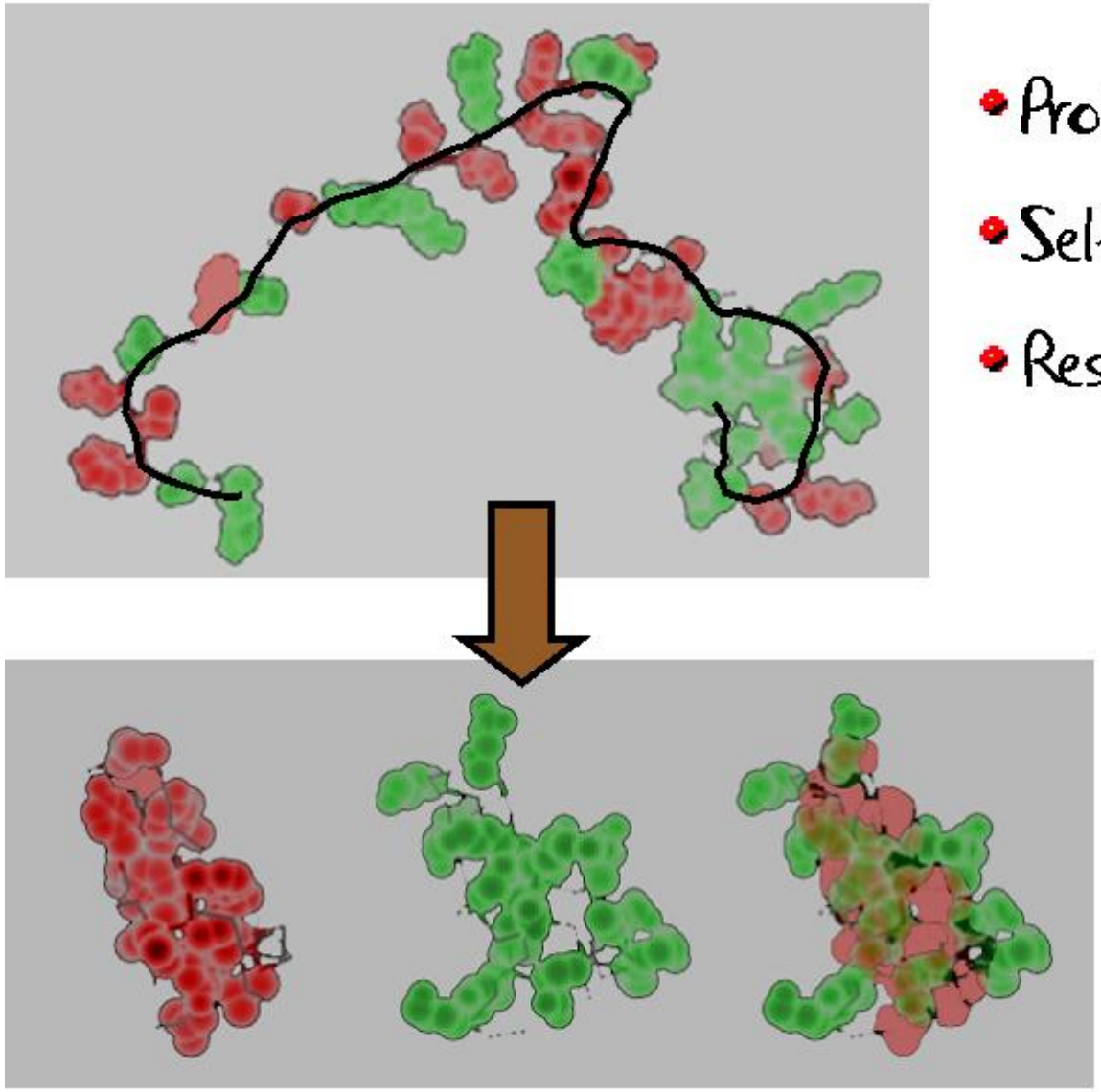
1988 Protein in Water

?

2005



WHAT DRIVES FOLDING?



- Protein is a chain.
- Self-avoiding and close packed.
- Residue preferences:
  - Inside/Outside
  - Specific Neighbors

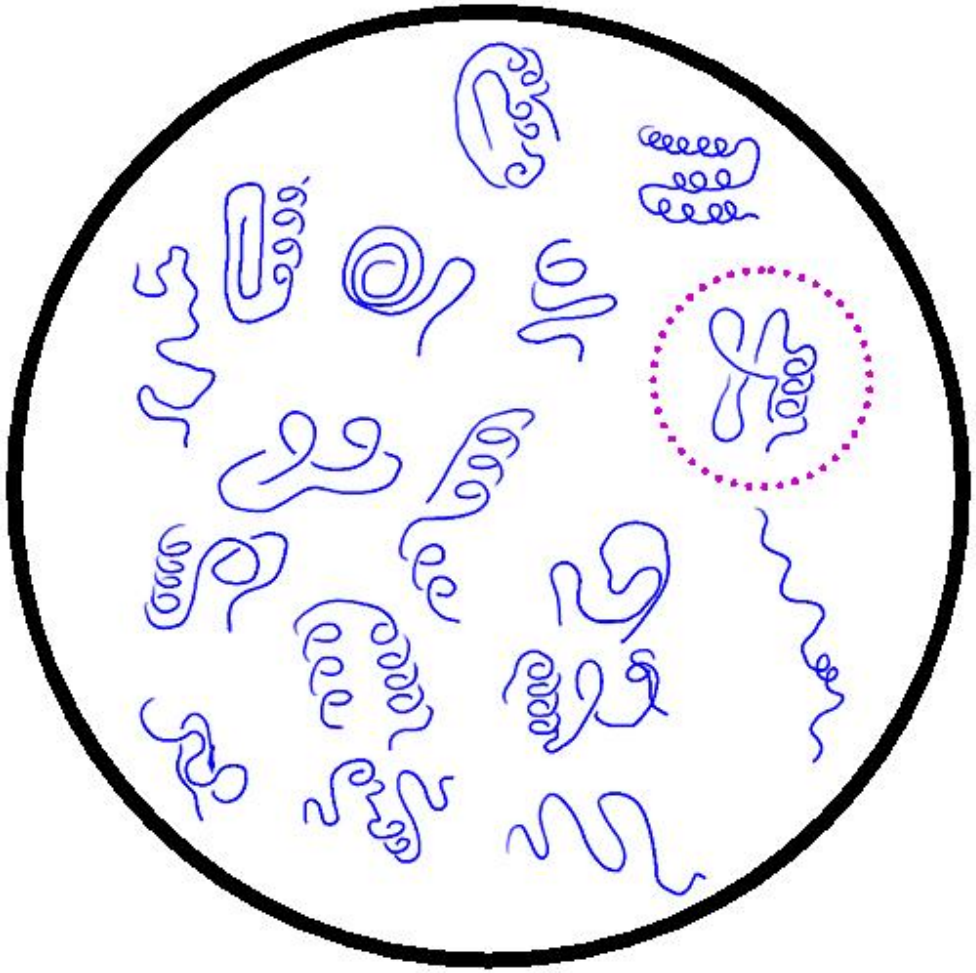
Hydrophobic Hydrophilic All Residues

Red are hydrophobic, like to be away from water

Green are hydrophilic, like contact with water

Discrimination Paradigm

A PARADIGM FOR PREDICTING STRUCTURE



Need a good energy function

DECOYS

- Construct a large number of possible folded shapes (Decoys).

DISCRIMINATION

- Select the correct, native fold.

THE CASP EFFECT

- Critical Assessment of Structure Prediction.
- Predict what no one knows.
- Predict what is about to be known.
- Carefully control evaluation and assessment. Competition?
- Meet to discuss what went wrong and what went right.
- Have had CASP1 ('94) through CASP4 ('00).

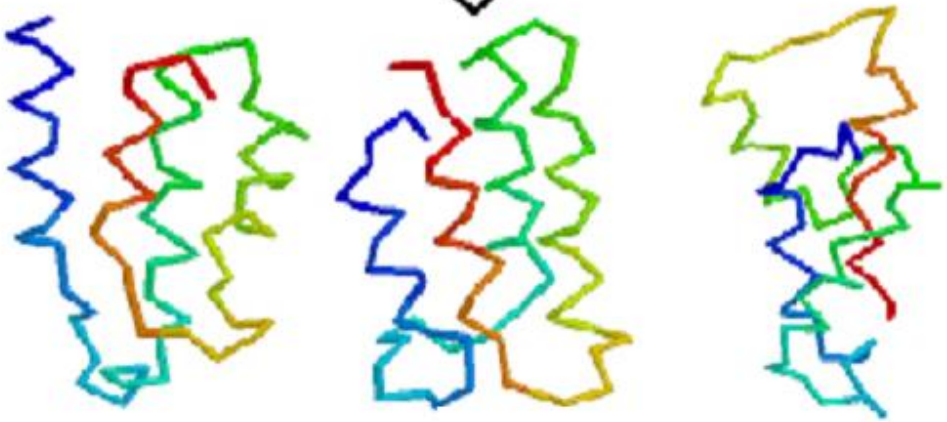
John Moult

Moult, Pedersen, Judson and Fidelis, A Large-Scale Experiment to Assess Protein Structure Prediction Methods. Proteins. 23: ii-v (1995)

HIEARCHICAL REDICTION
1998

HIERARCHICAL STRUCTURE PREDICTION

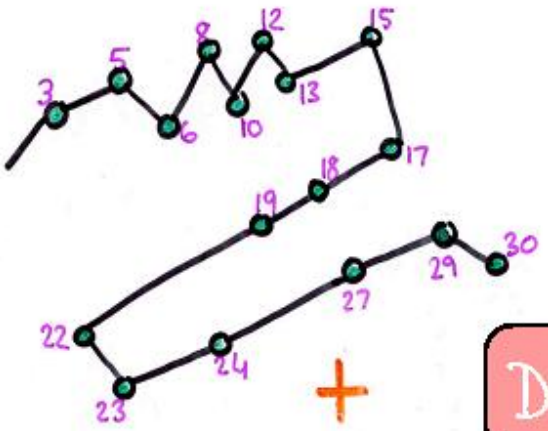
Use a $\sqrt{3}$-state lattice model.



Use a 4-state off-lattice model.

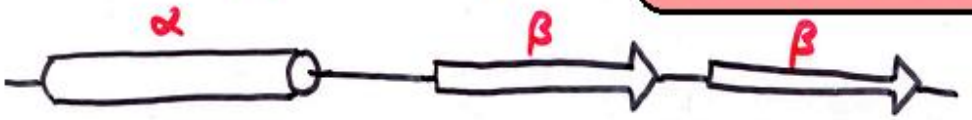
Yu Xia

Ram Samudrala



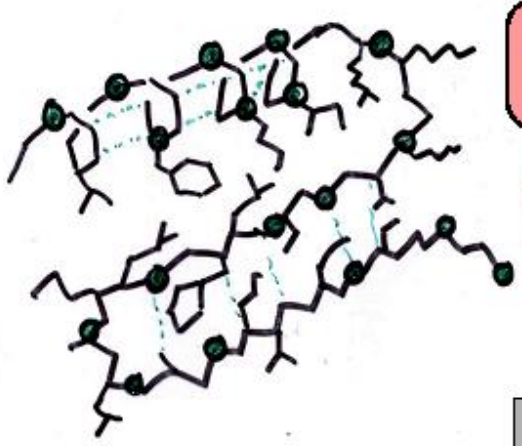
1 of 10,000 low-energy shapes.

David Hinds



Predicted secondary structure

Britt Park

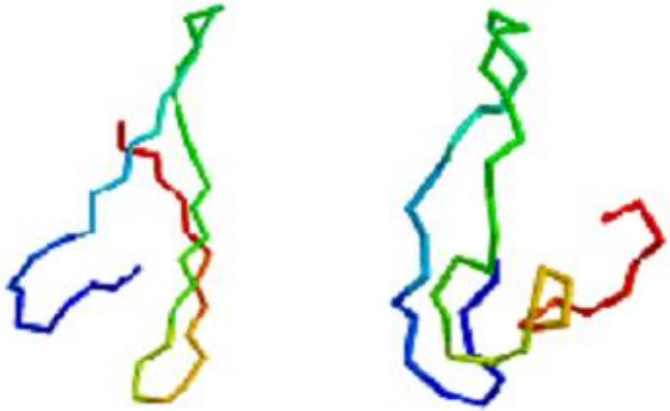


Add all atoms in full detail.

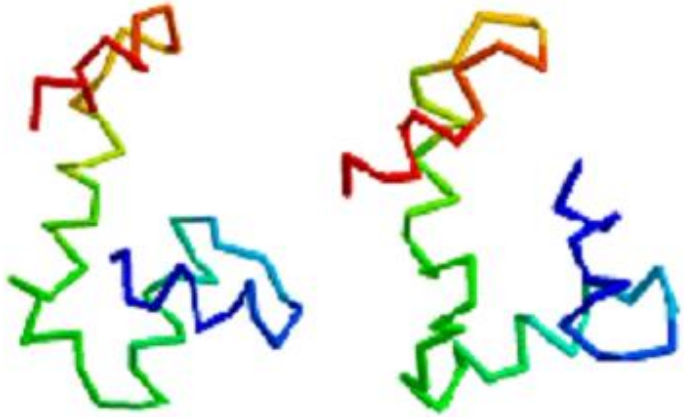
HIERARCHICAL PREDICTION DOES WELL

Does well at CASP3 (Critical Assessment of Structure Prediction, Asilomar 1998).

T46/adg 7.5 Å (49 residues; 66113)

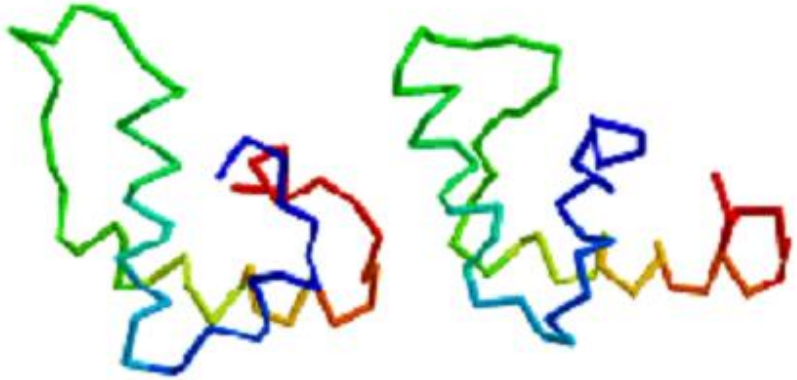


*** T56/dnab 6.8 Å (60 residues; 67126)**



**** T59/smd3 6.7 Å (46 residues; 3075)**

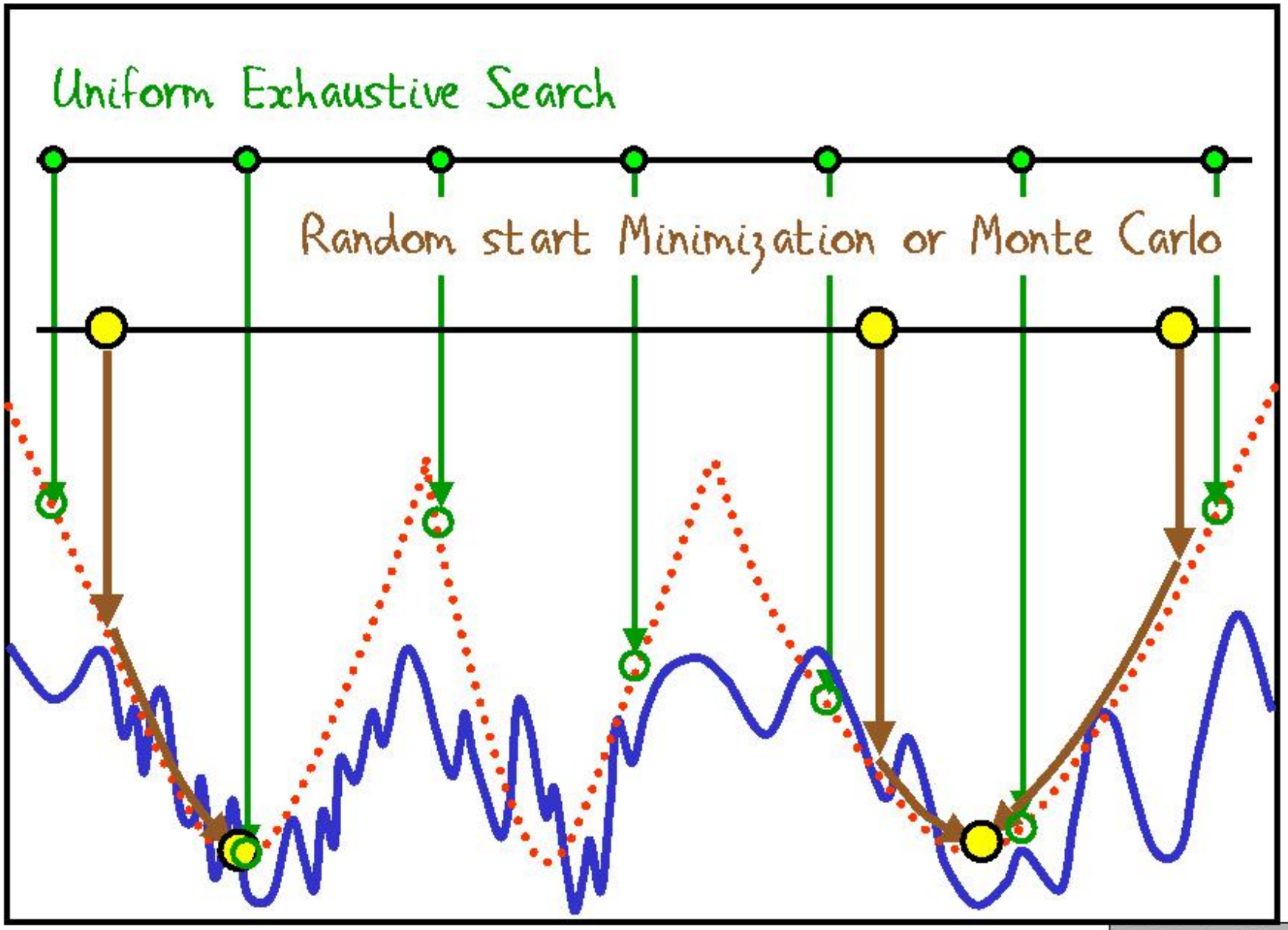
**** T61/hdea 7.4 Å (66 residues; 974)**



Samudrala, Xia, Huang & Levitt. Bona Fide Ab Initio Protein Structure Prediction Using a Combined Hierarchical Approach. Proteins, 37 (3S): 194–198 (1999).

SPECIAL POTENTIALS 2000

SAMPLING ANT LION TOWN POTENTIALS



Energy Minimization

ALL-ATOM ENERGY MINIMIZATION

- Minimize all-atom energy with respect to all torsion angles.
- Augment the normal potential energy function with:
  - Cooperative hydrogen bonds.
  - Cooperative hydrophobic interactions.
  - Forced exposure of charges.

Chen Keasar

POTENTIAL ENERGY IN TORSION SPACE

$$U = \sum K_\phi [1 - \cos(n\phi + \delta)]$$

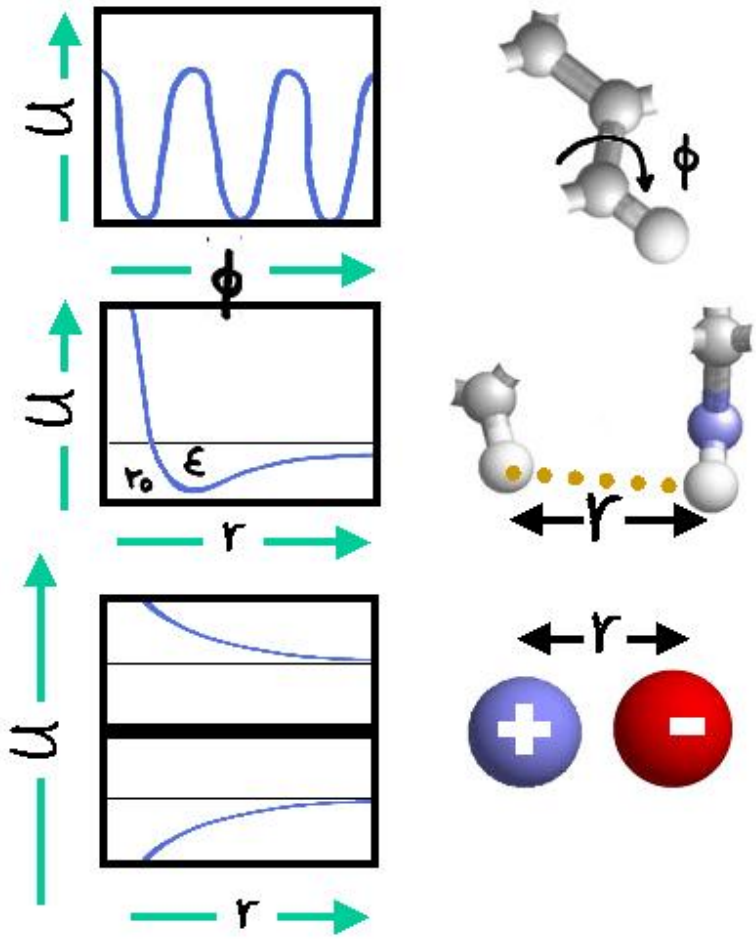
All Torsion Angles

$$+ \sum \epsilon \left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6} \right]$$

All nonbonded pairs

$$+ \sum 332 q_i q_j / r$$

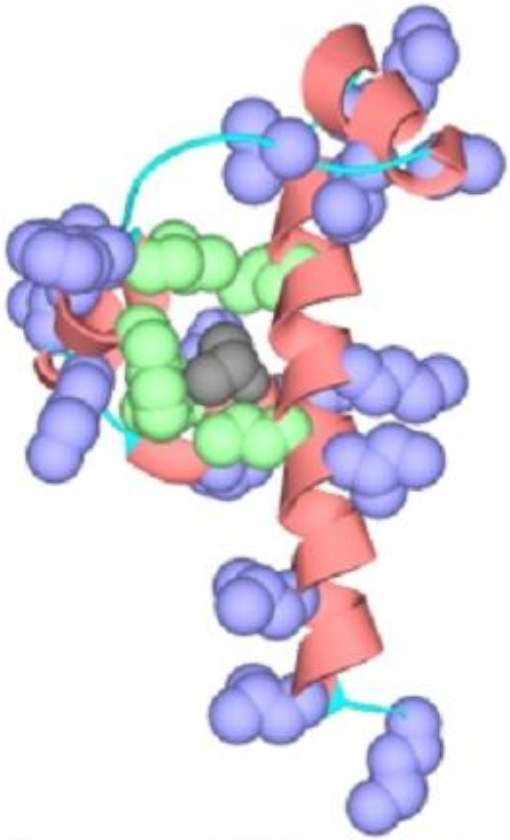
All partial charges



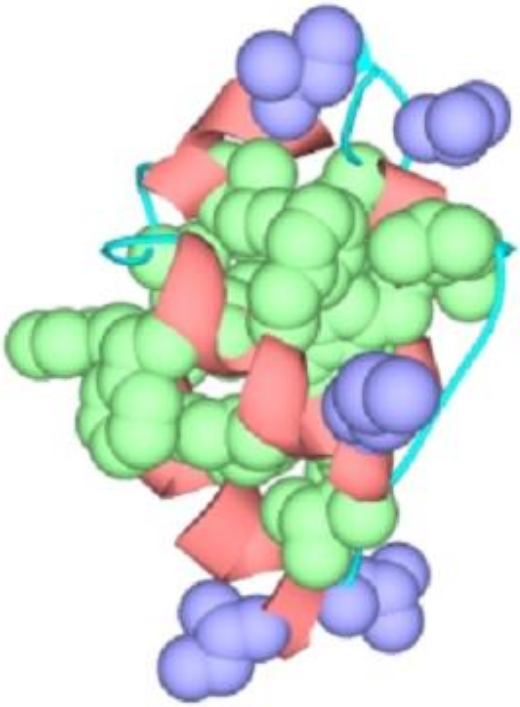
- A protein with N residues has about 4N (φ, ψ, χ) single bond torsion angles. The same protein has about 50N Cartesian coordinates (x, y, z).

COOPERATIVE HYDROPHOBIC PACKING

- Cooperative hydrophobic compaction makes a good core.



Original Potential

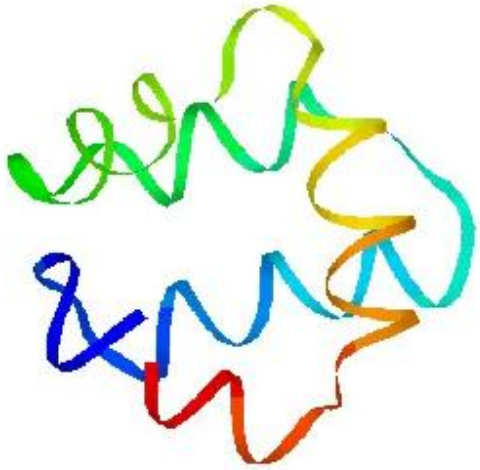


Modified Potential

STRUCTURE PREDICTION BY MINIMIZATION

- Minimize special energy function with respect to torsion angles (φ, ψ, χ).
- Add energy terms for cooperative hydrogen bonds and hydrophobic compaction.

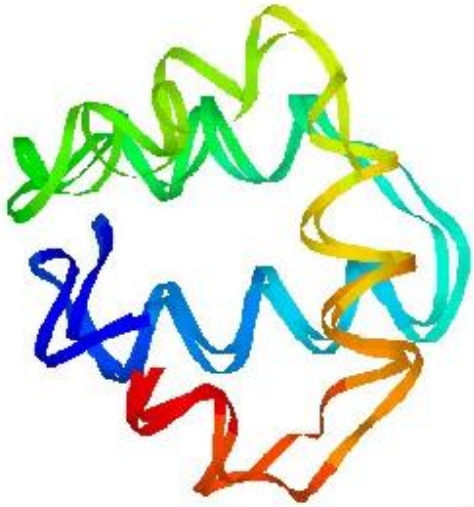
10Å = 1 nm



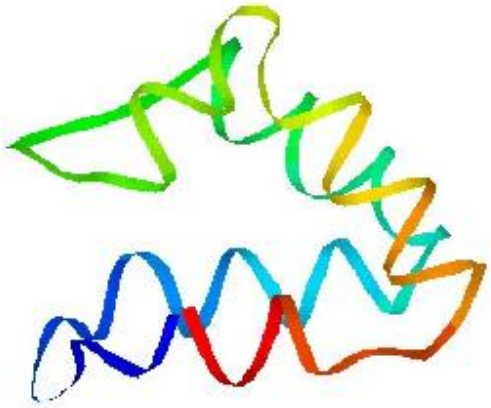
Structure of T102

This method did well at CASP4, 2000.

25Å



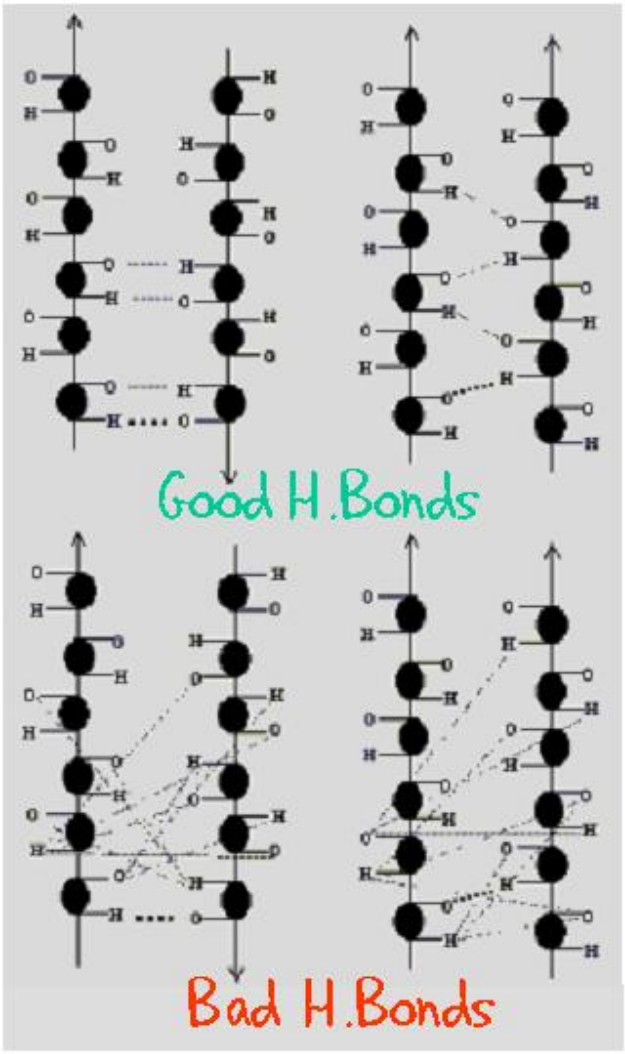
Best T102. RMS= 3.3Å

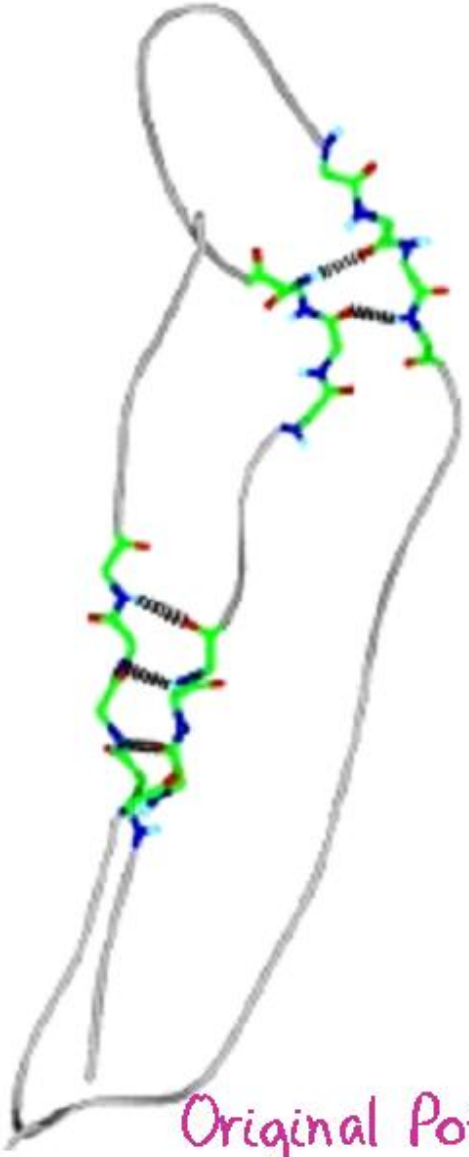


T102 Submitted. RMS=5.0Å

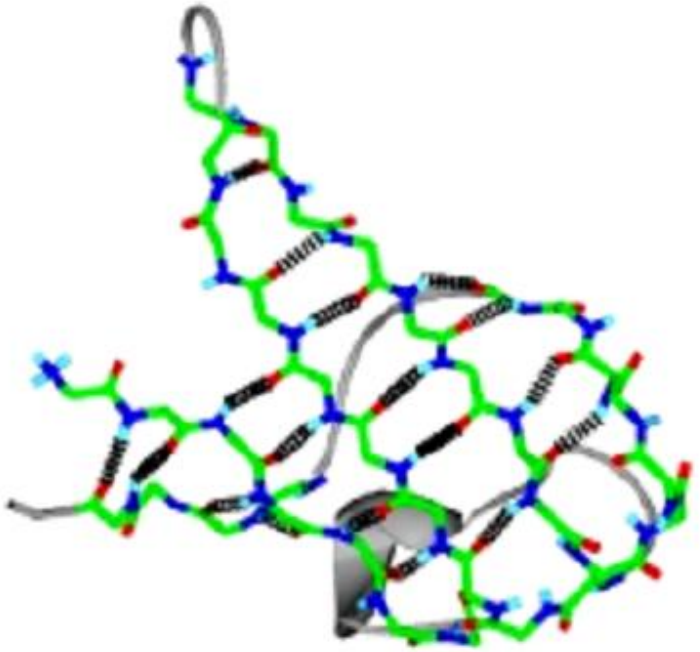
COOPERATIVE HYDROGEN BONDS

- Cooperative hydrogen bonds give rise to good secondary structure.





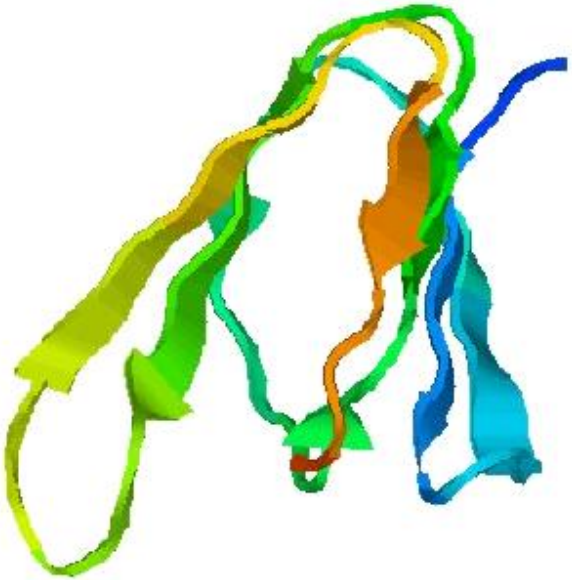
Original Potential



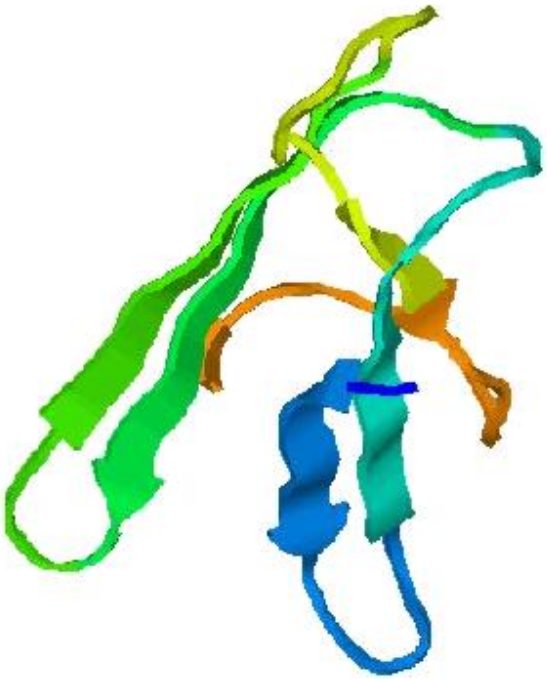
Modified Potential

ALL-β PREDICTION SUCCESS

- All-β sheet proteins are the hardest to predict.
- Torsion minimization does well on T114, an all β-protein.



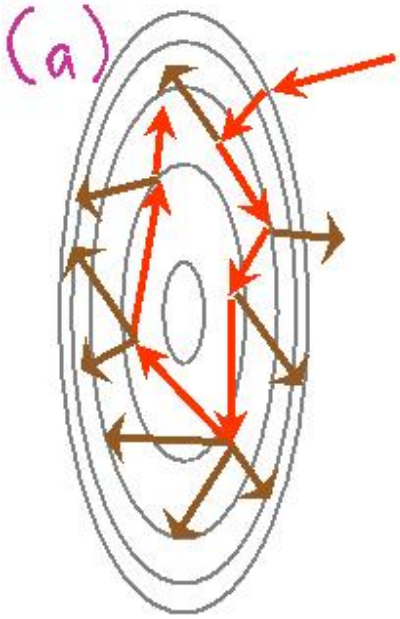
Native Structure.

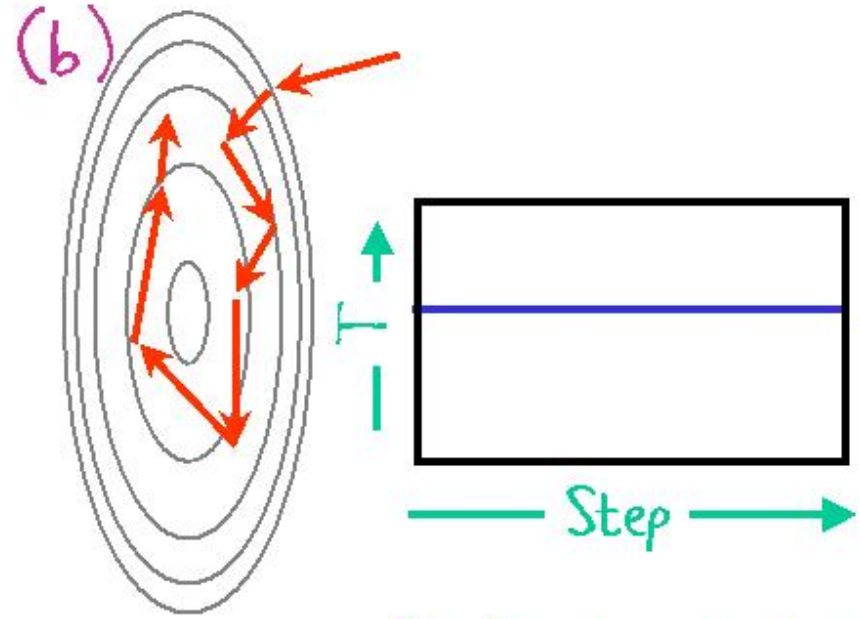


Prediction is somewhat similar.

Segment Monte Carlo

MONTE CARLO METHODS





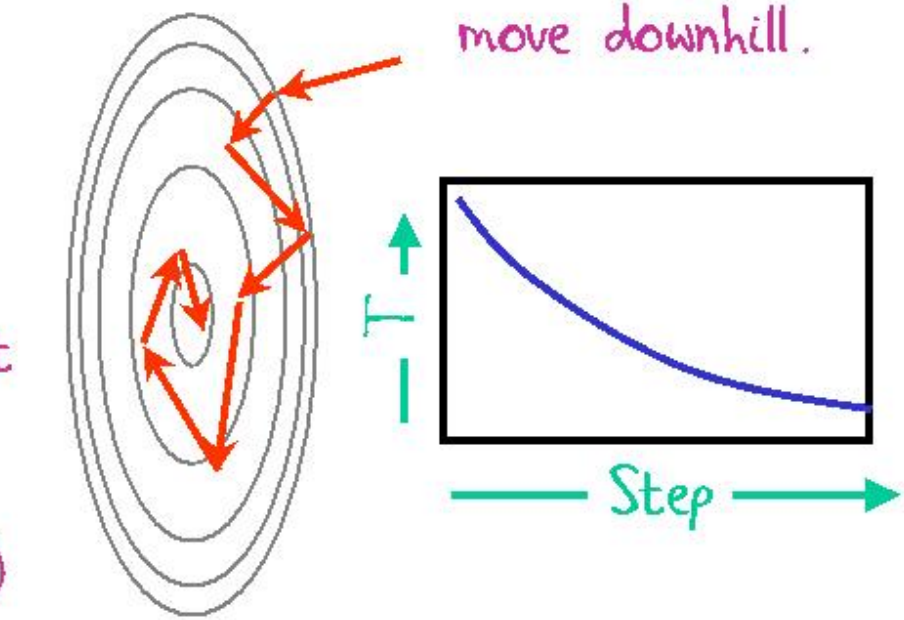
- Normal Monte Carlo:

  Make random moves and accept some of them (Metropolis).

(a) At each step, attempt many moves.

(b) Accept the first move that obeys: Random number, $R_N < \exp(-\Delta U/kT)$

At low temperature move downhill.



- Simulated Annealing:

  Reduce T, the temperature, as the run proceeds.

FRAGMENT MONTE CARLO

Fragment Library

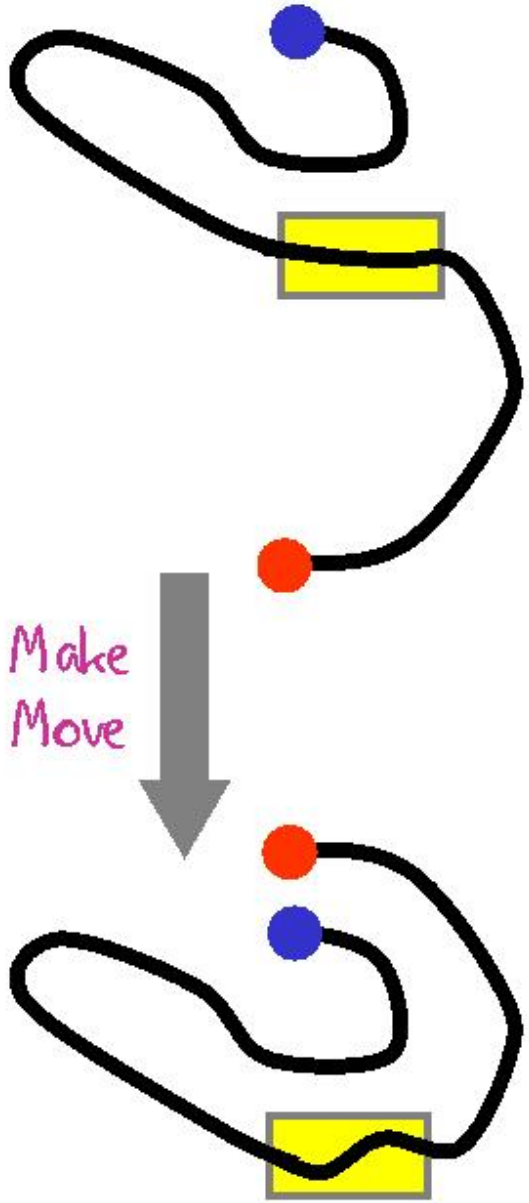

RVL


RFL


EVL

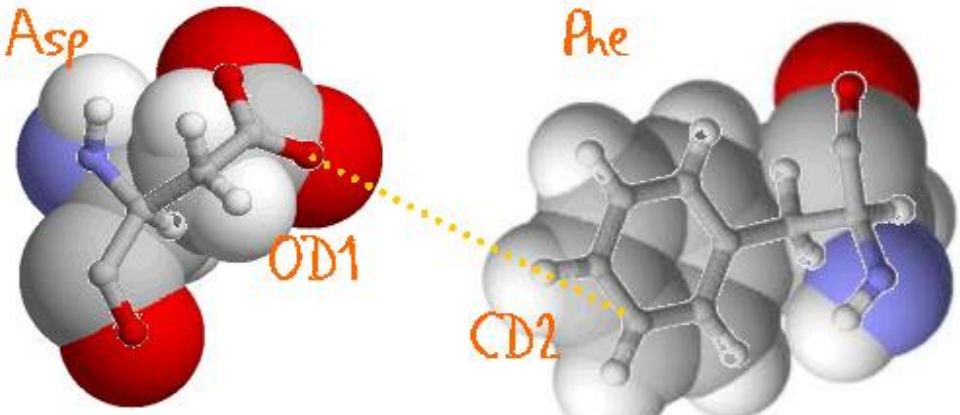

KVI

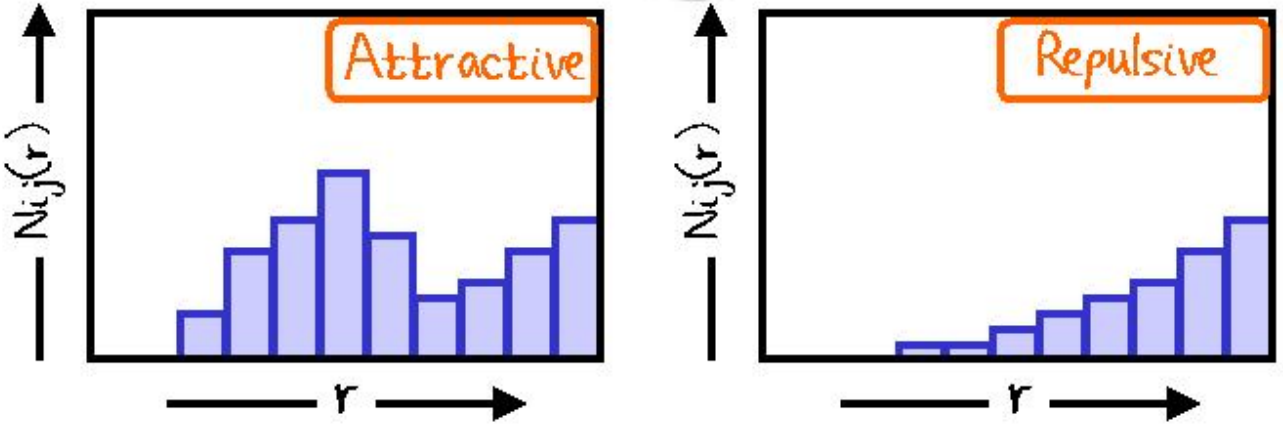

KVY

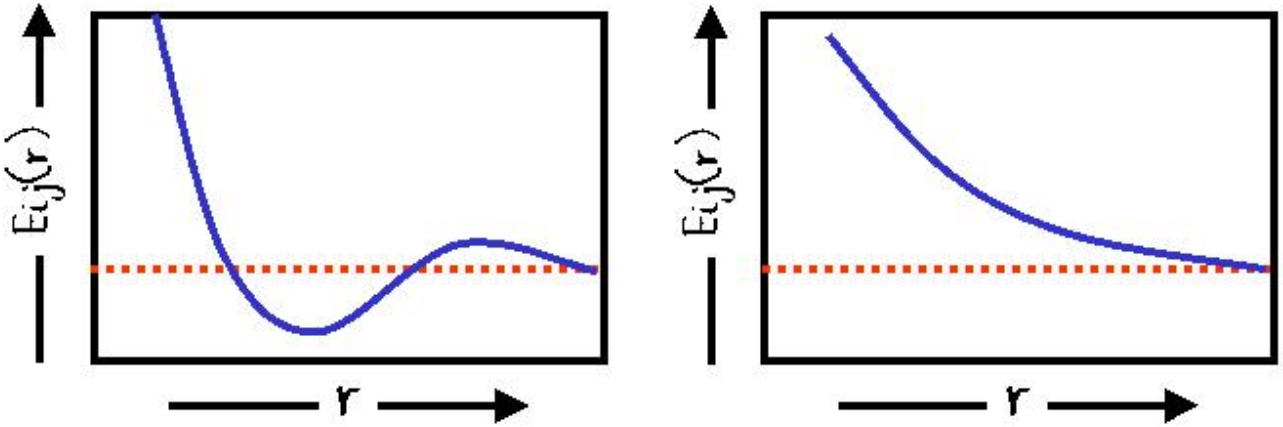


- Make a library of small fragments of similar sequence.
- Swap in a new fragment by setting six (φ, ψ) torsion angles.
- Accept move by Monte Carlo and anneal.

KNOWLEDGE-BASED ENERGIES







- Get distribution of distances between pairs of atom centers of a particular type, e.g. D-OD1...F-CD2.

- Normalize and take log to get Energy score:

$$E_{ij}(r) = \log[N_{ij}(r)/M_{ij}(r)]$$

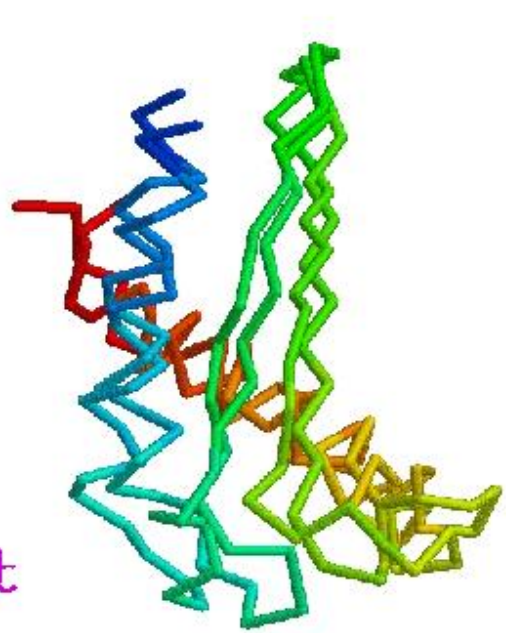
SEGMENT FOLDING PREDICTION

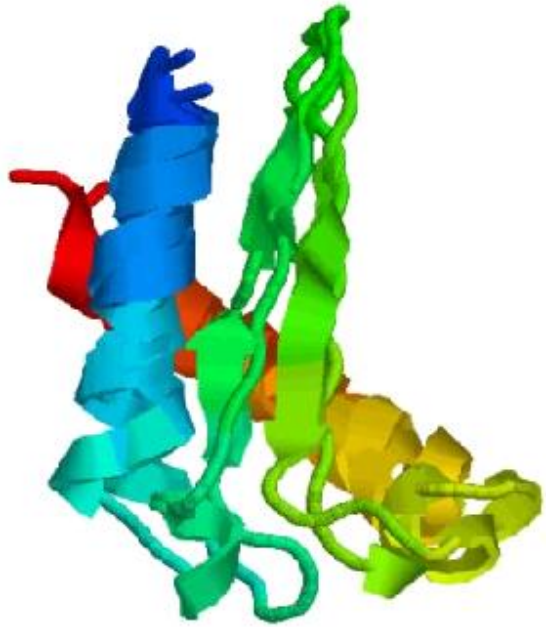
- Do Monte Carlo moves with respect to (phi, psi) torsion angles. Simulated annealing.

Ram Samudrala

- Use all-atom Knowledge-Based energy function.
  Add terms to enforce compaction.

- Get reasonable (phi, psi) angles from real protein fragments.





- This method does well at CASP4, Asilomar '00.

T0110 Fit 80 residues to 4.0Å

HARD PROBLEMS

WHY IS FOLDING SO HARD?

- Many different specific interactions.
- Cooperativity of the underlying interactions.
- Three-dimensional with very many possible spatial arrangements.
- Violates Crick's Law of Hard Problems.

WHY ARE WE GETTING BETTER AT FOLDING?

- Peer pressure (CASP)?
- Faster computers?
- Many more sequences?
- More structures?

INFORMATION + PHYSICS = LIFE

DNA Sequence → RNA Sequence → Protein Sequence → Folded Protein

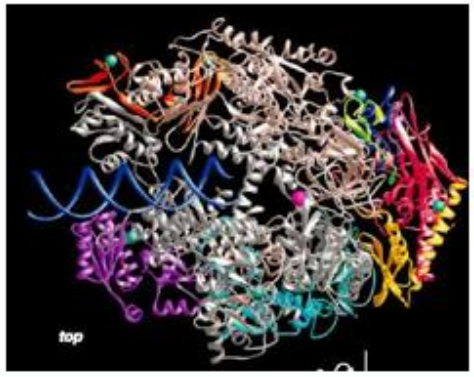
- *in silico*

Easy: Change T to U

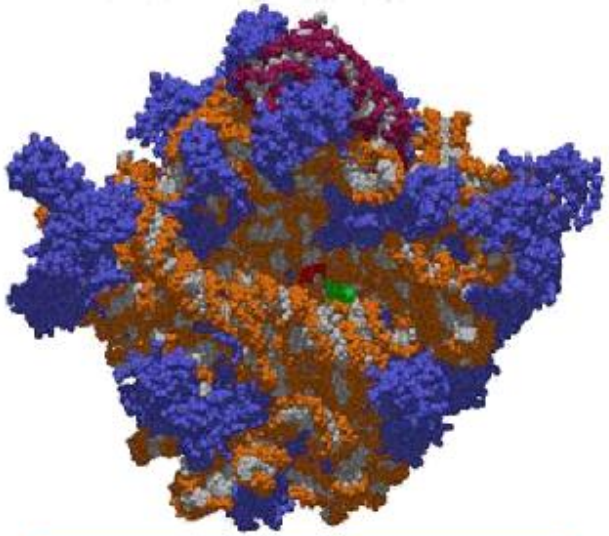
Easy: Triplet Code

Hard: Folding is many body simulation

- *in vivo*



Hard: Transcription Polymerase



Hard: Translation Ribosome

Easy: Folding is free by laws of physics

ACKNOWLEDGEMENTS

PEOPLE

- Tanya Raschke
- Erik Sandelin
- Boris Fain
- Patrice Koehl
- Michael Sykes
- Yu Xia
- Rachel Kolodny
- Chen Keasar
- Nizar Batada
- Chris Summa

SUPPORT

- NIH (NIGMS); DOE (SBI); NSF (ITR)

WEB

- http://csb.stanford.edu/levitt
  - Papers
  - Lecture
  - Thesis
- http://biospace.stanford.edu
- http://dd.stanford.edu
- http://astral.stanford.edu

THE END

HISTORICAL RECORD OF BEST PREDICTIONS AT CASP

CASP & YEAR	NUMBER TARGETS	BEST RESULT <Q3>	GROUP
CASP1 1994	6	63	Rost & Sander
CASP2 1996	24	70	Rost
CASP3 1998	18	75	Jones
CASP4 2000	28	80	Jones

- Steady improvement of about 5% per CASP (every two years)

NOTES

- PDF files on home page

- Searching www.google.com