Protein Folding

CMSC 423

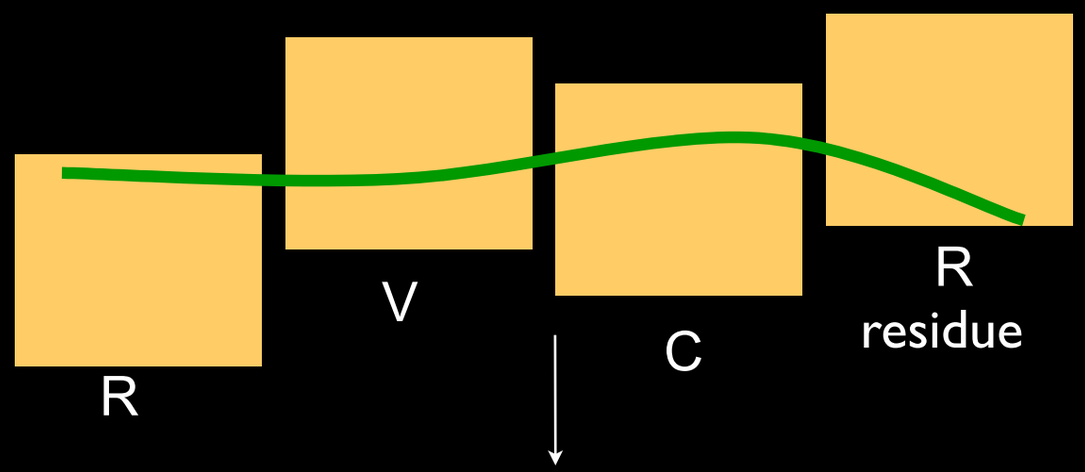
Proteins

mRNA
$\Sigma = \{A,C,G,U\}$

protein
$|\Sigma| = 20$ amino acids

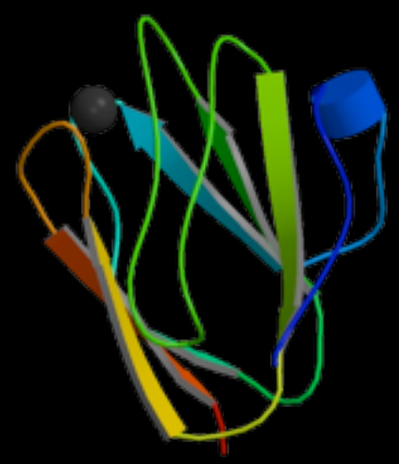
AGG GUC UGU CGA

R V C R

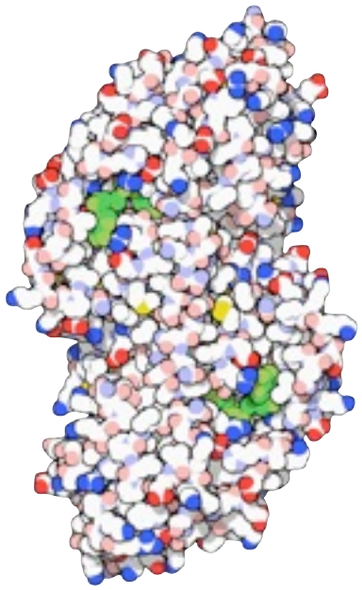


Amino acids with flexible side chains strung together on a backbone

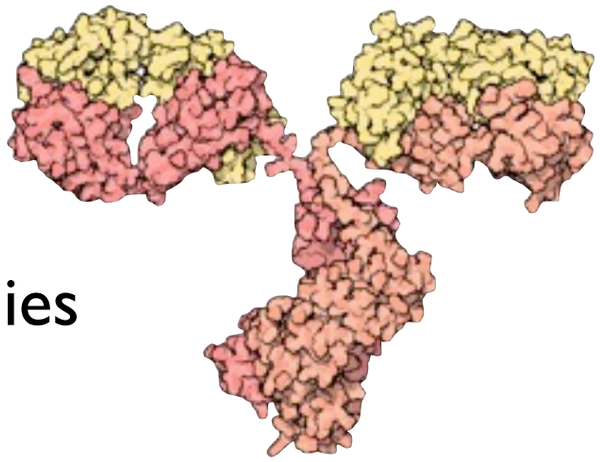
Function depends on 3D shape



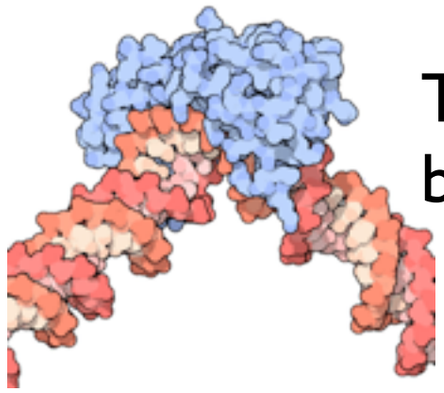
Examples of Proteins



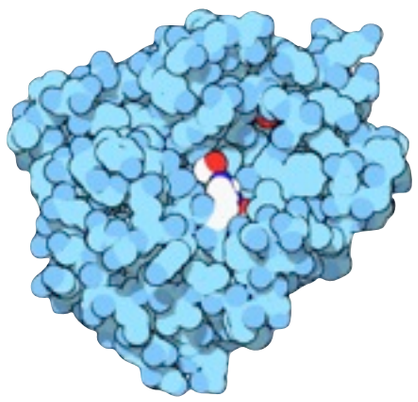
Alcohol dehydrogenase



Antibodies



TATA DNA binding protein



Trypsin: breaks down other proteins

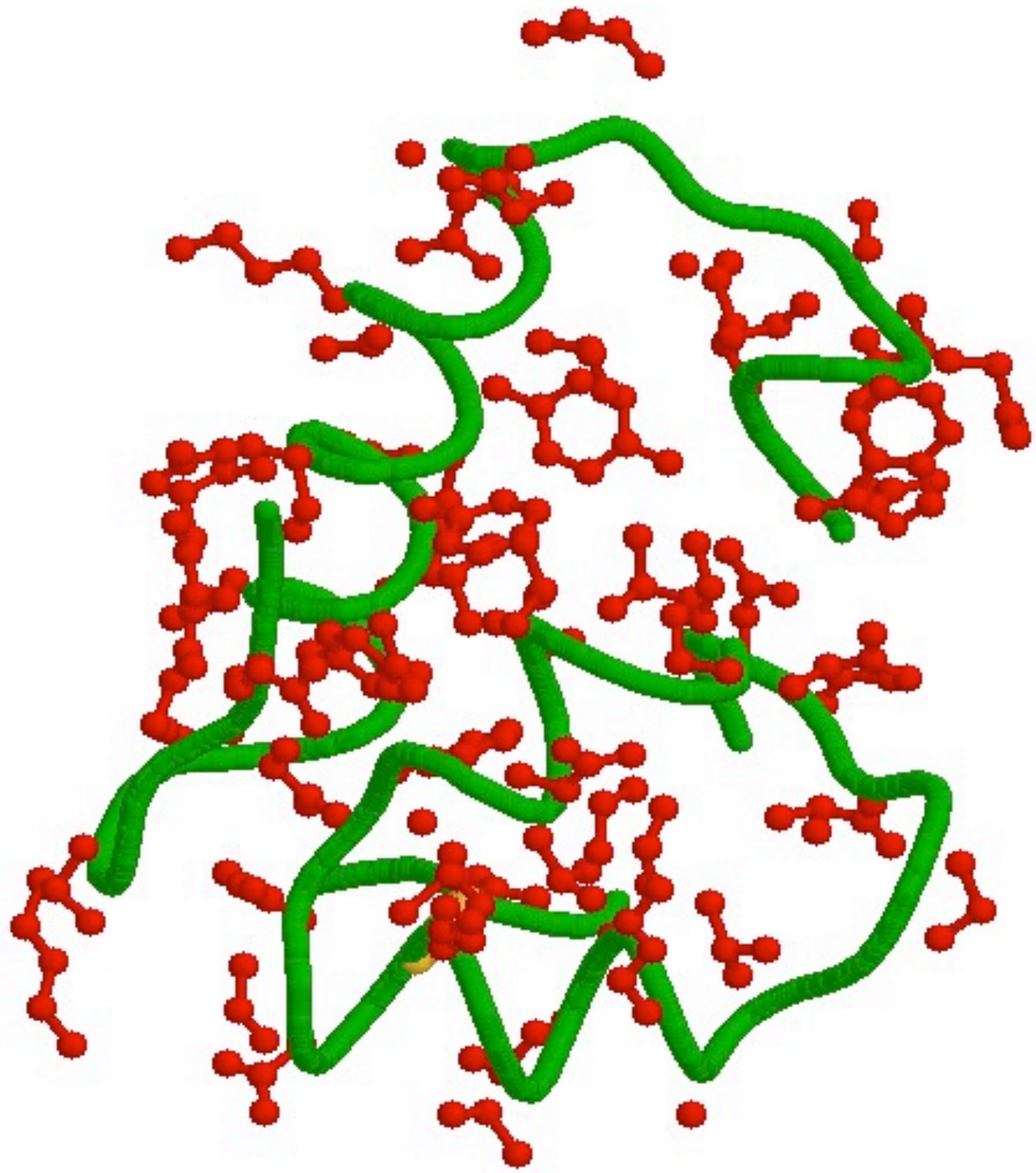
Collagen: forms tendons, bones, etc.

Examples of "Molecules of the Month" from the Protein Data Bank
http://www.rcsb.org/pdb/

Protein Structure

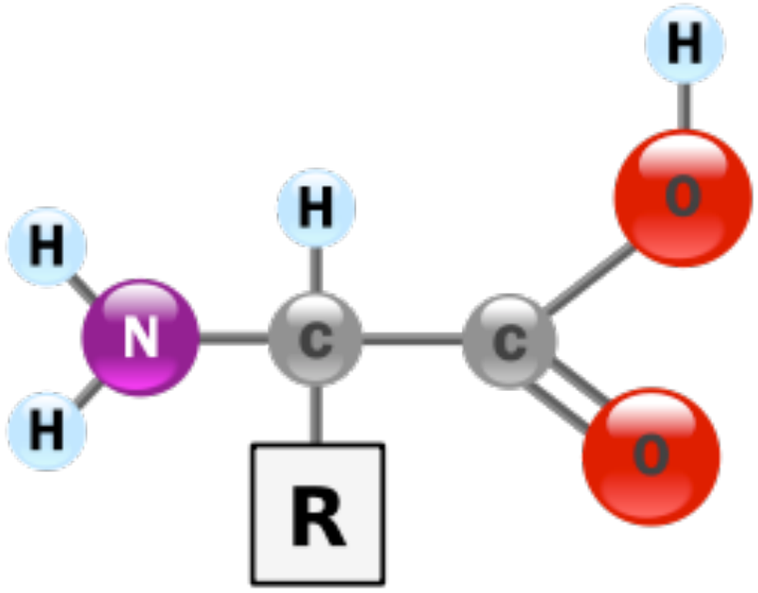
Backbone

Protein Structure



Backbone

Side-chains



Twenty standard Amino Acids

Nonpolar, aliphatic R groups

Glycine · Alanine · Valine · Leucine · Methionine · Isoleucine

Aromatic R groups

Phenylalanine · Tyrosine · Tryptophan

Positively charged R groups

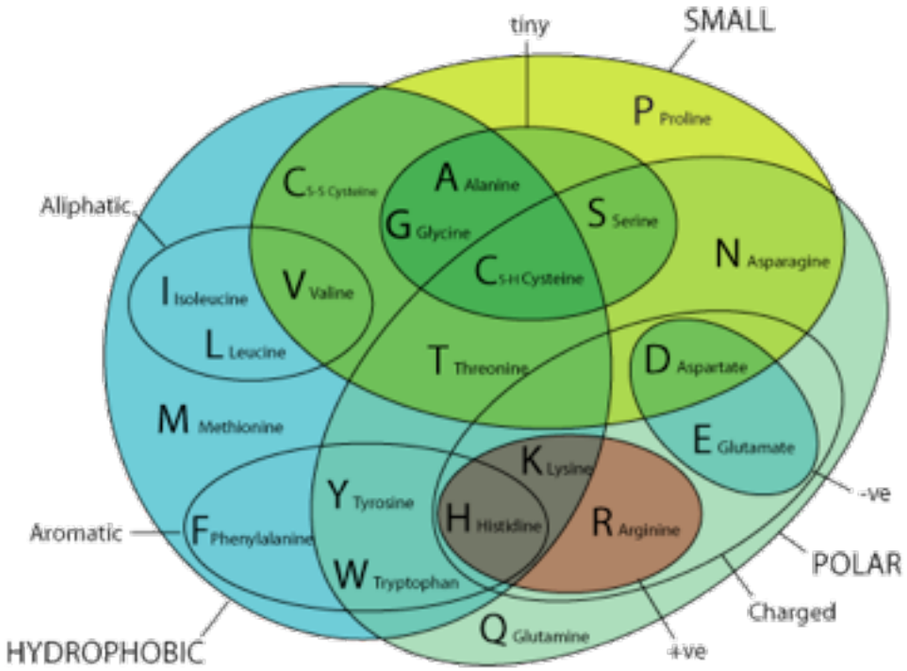
Lysine · Arginine · Histidine

Polar, uncharged R groups

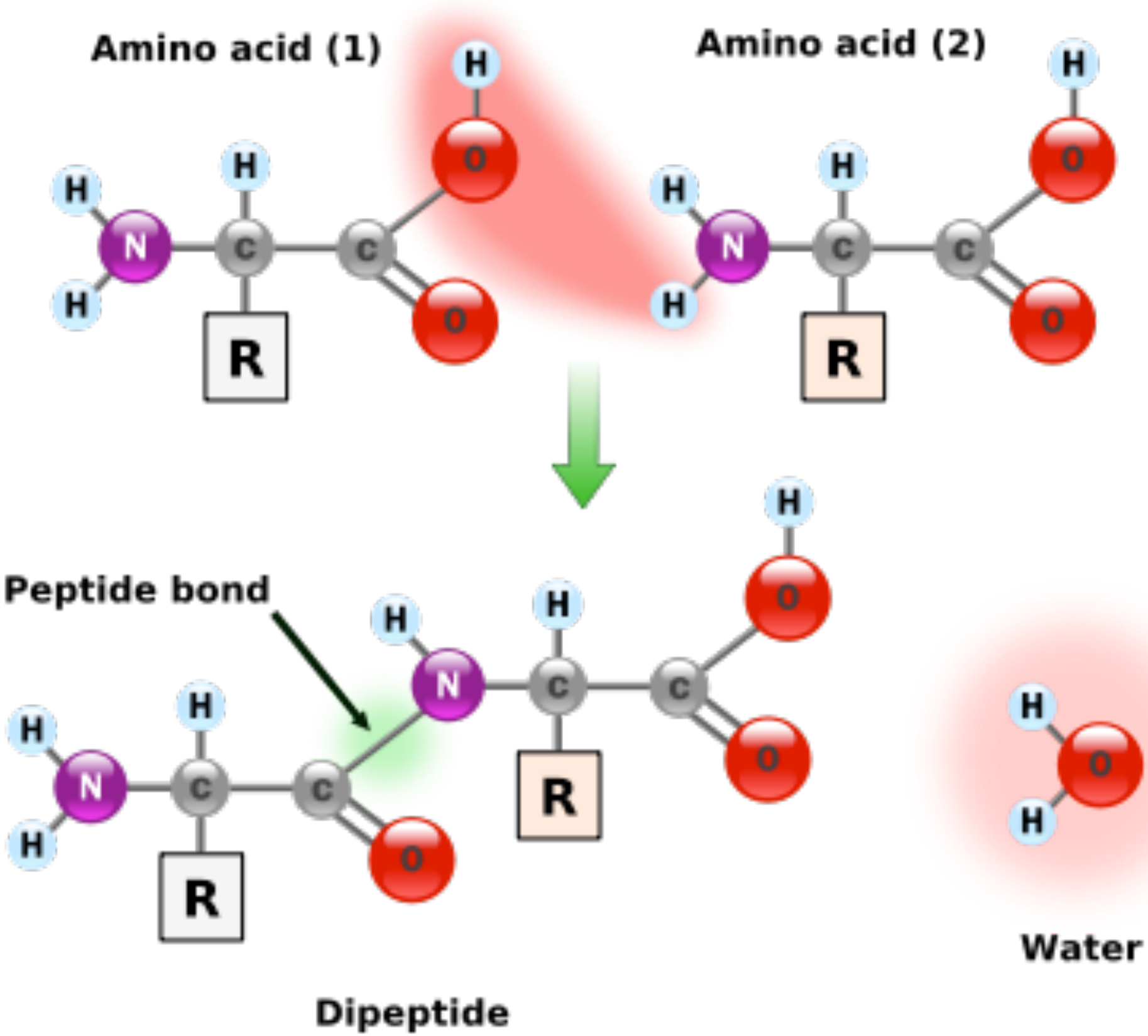
Serine · Threonine · Cysteine · Proline · Asparagine · Glutamine

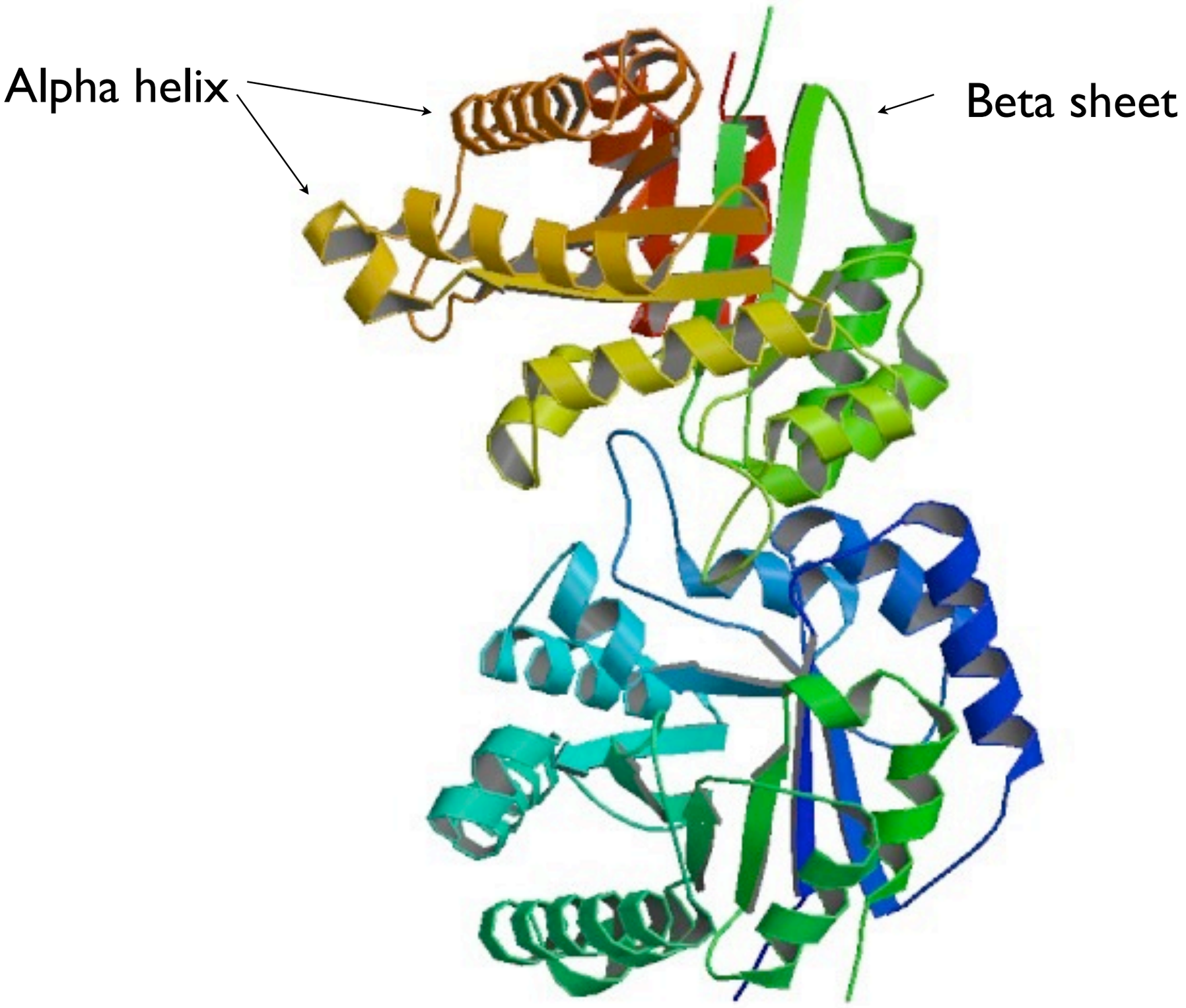
Negatively charged R groups

Aspartate · Glutamate



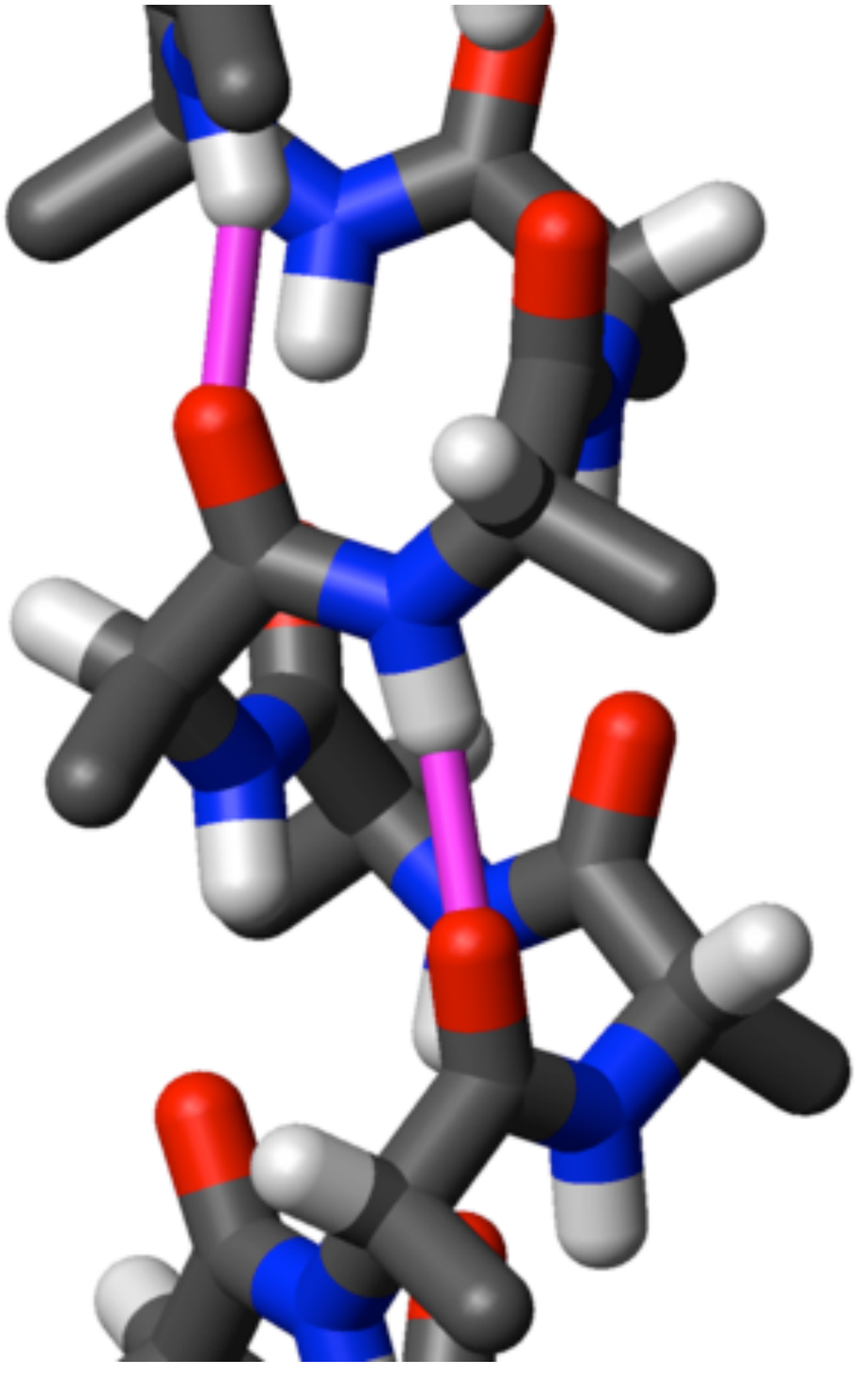
http://www.jalview.org/help/html/misc/properties.gif


Amino acid (1)
Amino acid (2)
Peptide bond
Dipeptide
Water
H
N
C
O
R



1tim

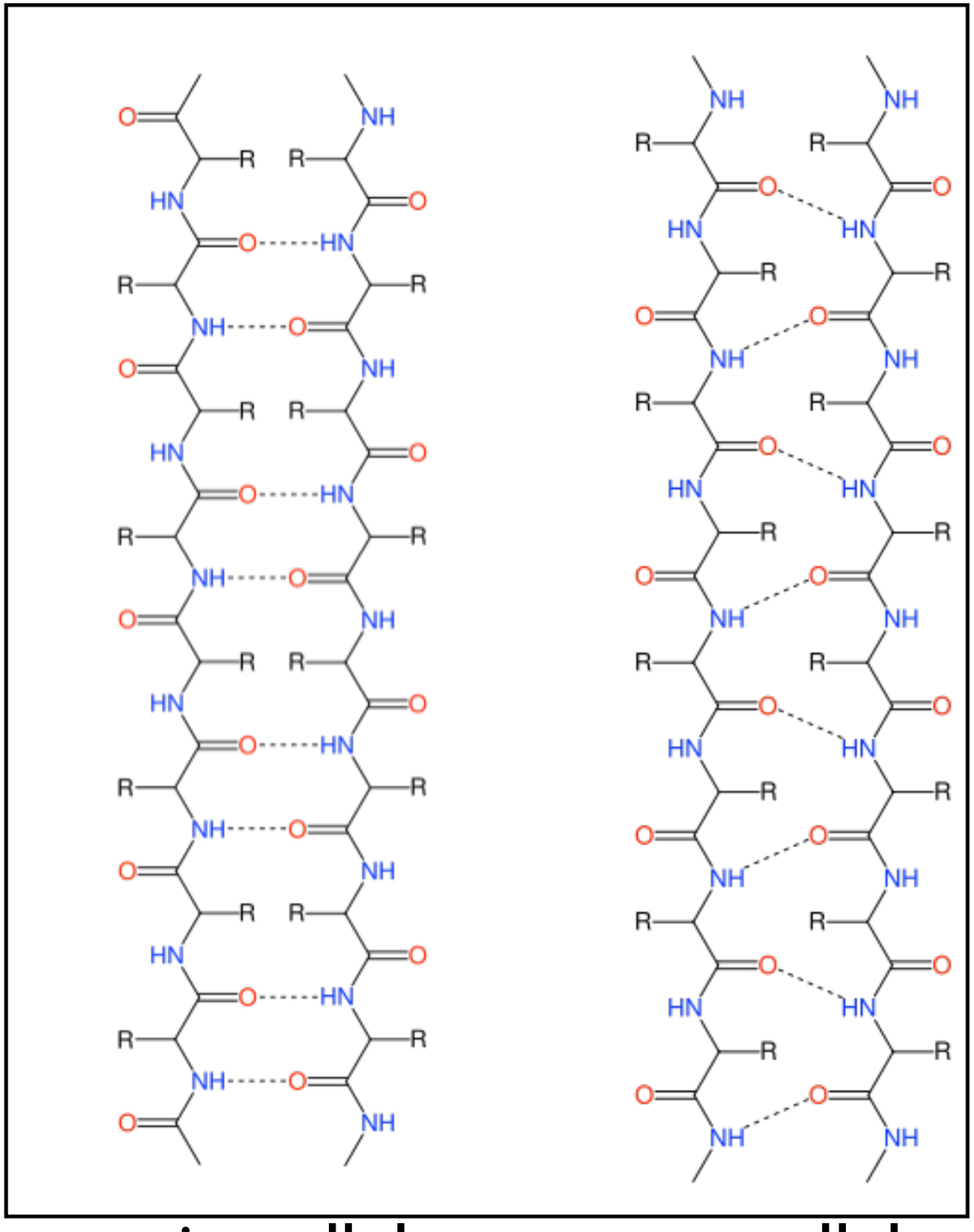
Alpha Helix



C'=O of residue *n* bonds to
NH of residue n + 4

Suggested from theoretical consideration
by Linus Pauling in 1951.

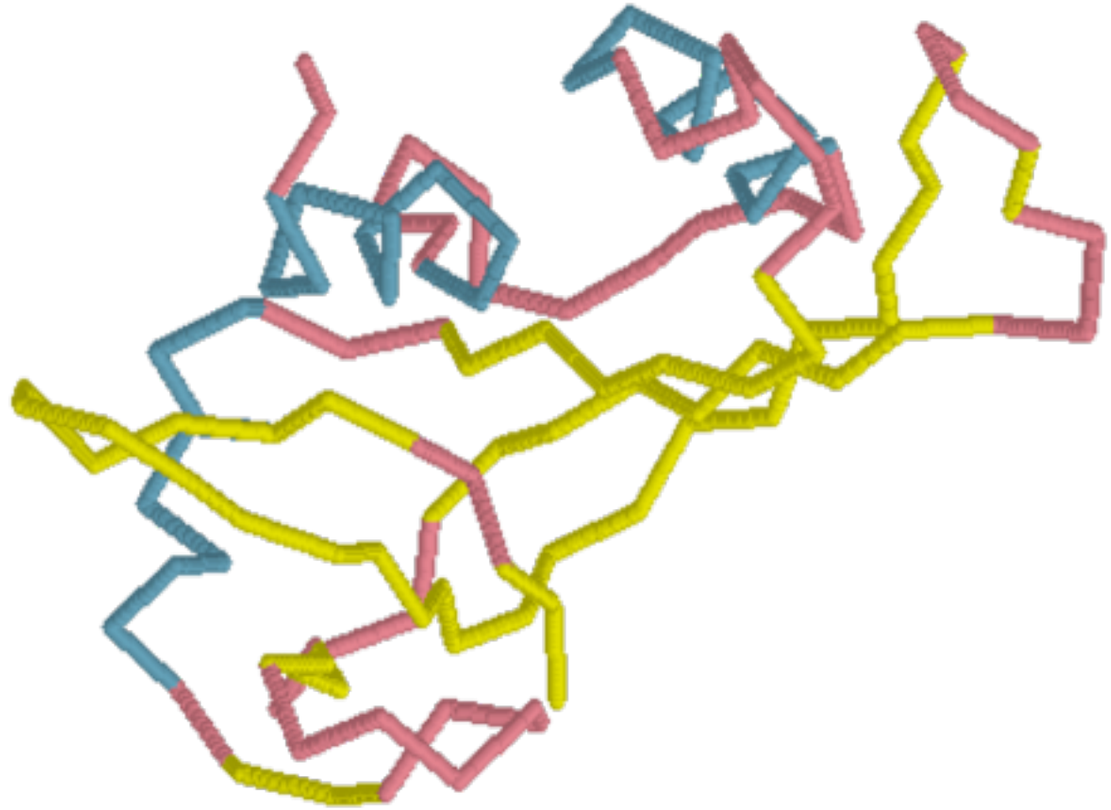
Beta Sheets



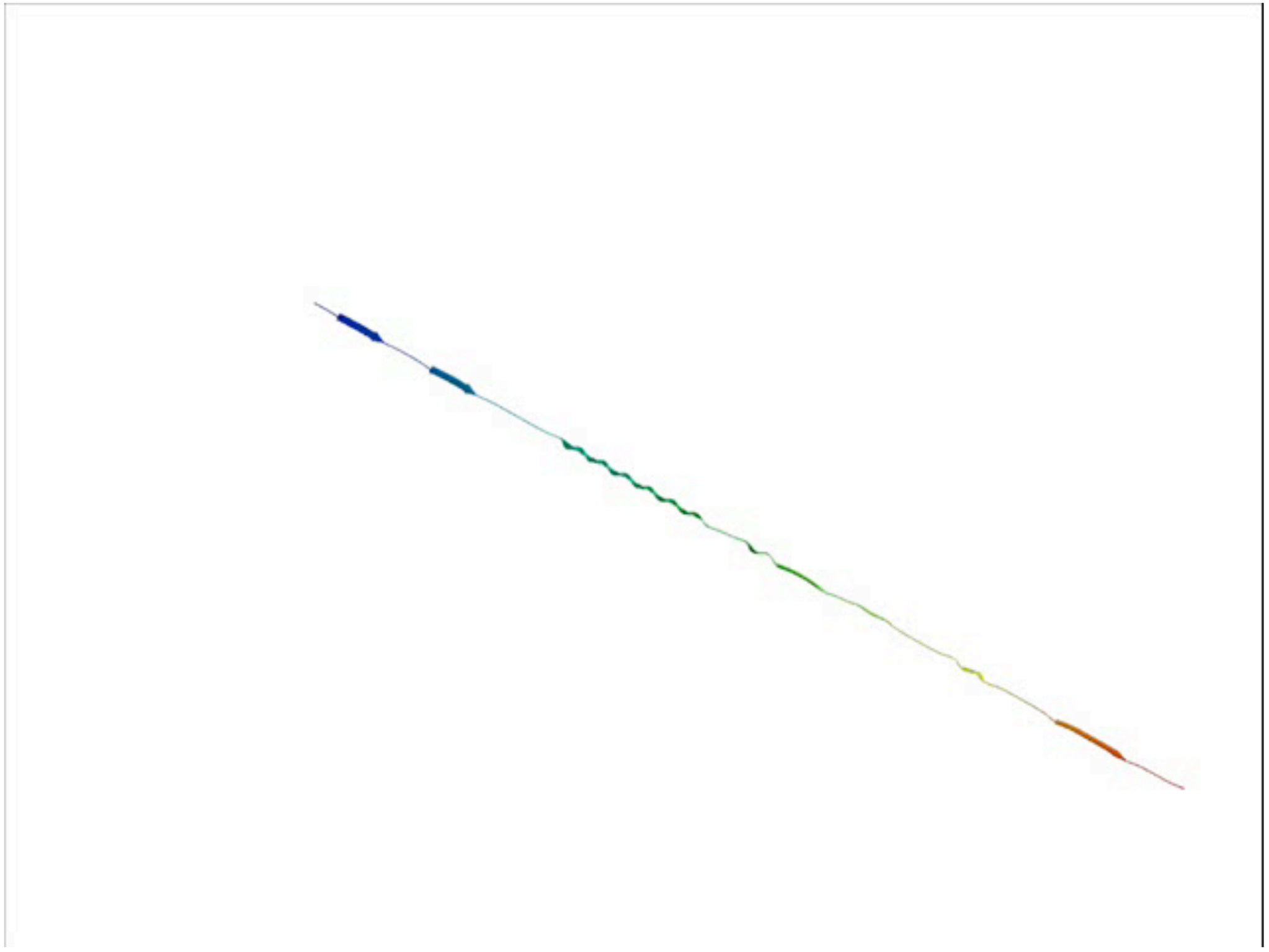
Structure Prediction

Given: KETAAAKFERQHMDSSTSAASSSN…

Determine:

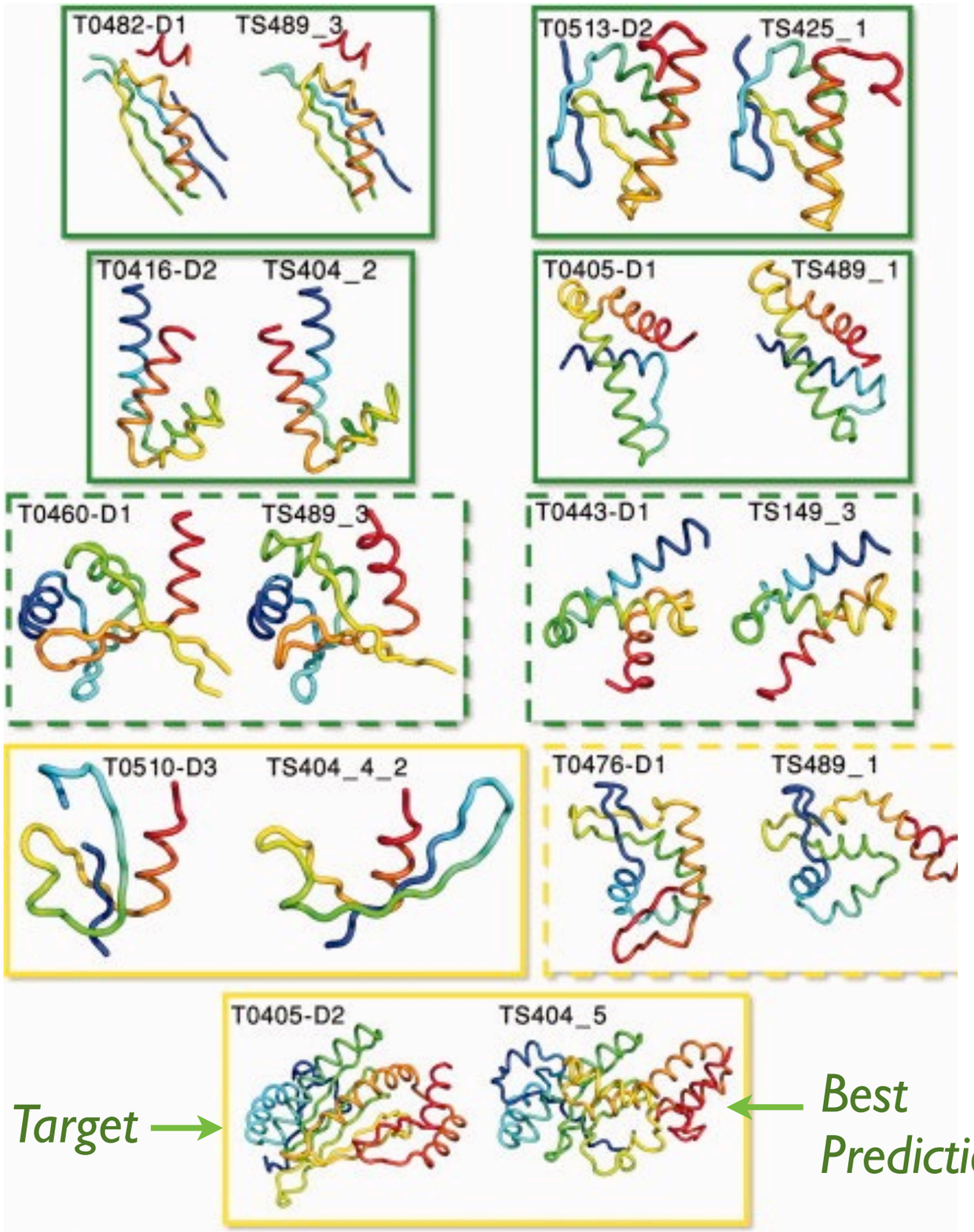


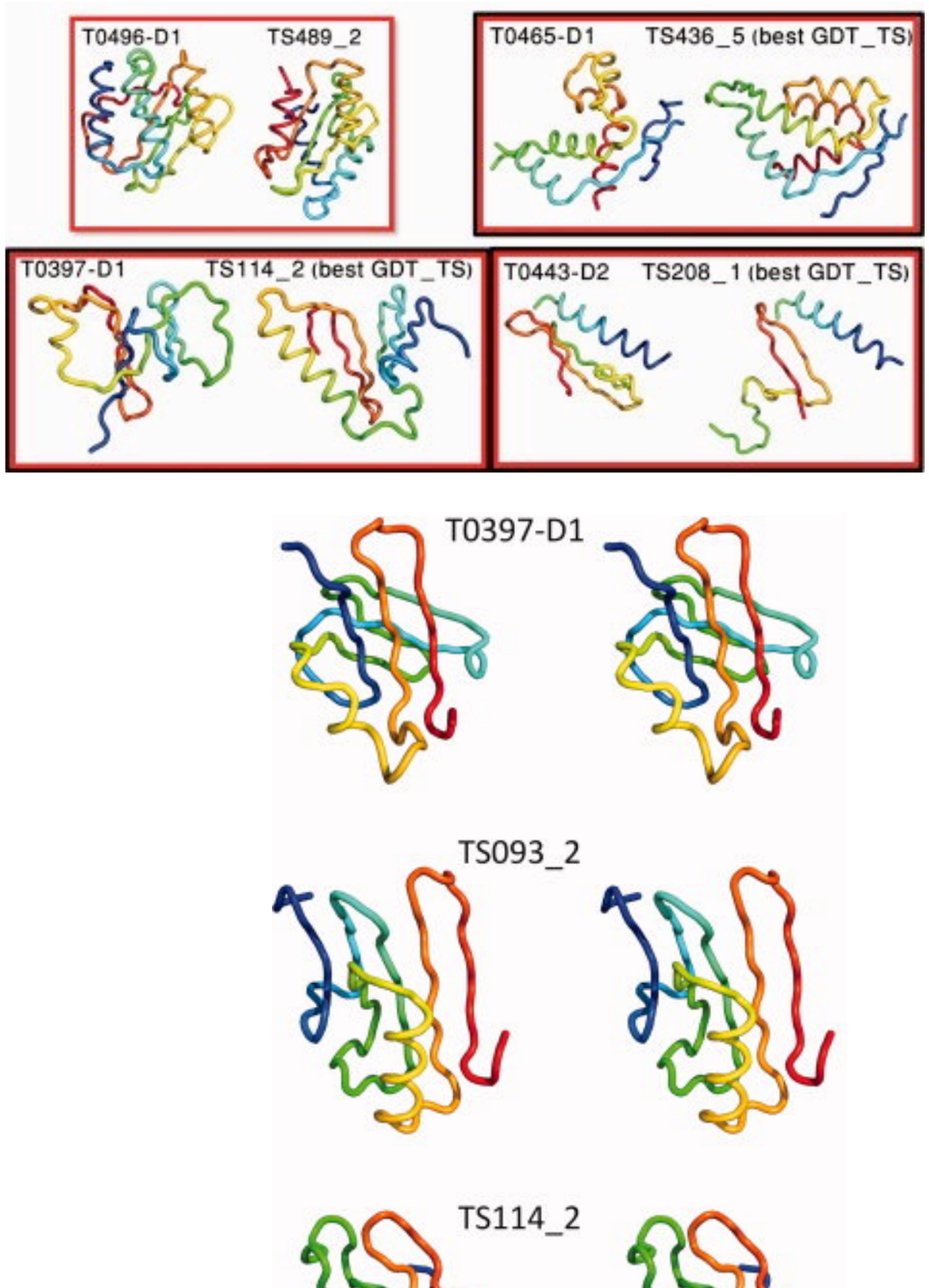
Folding Ubiquitin with Rosetta@Home



http://boinc.bakerlab.org/rah_about.php

CASP8

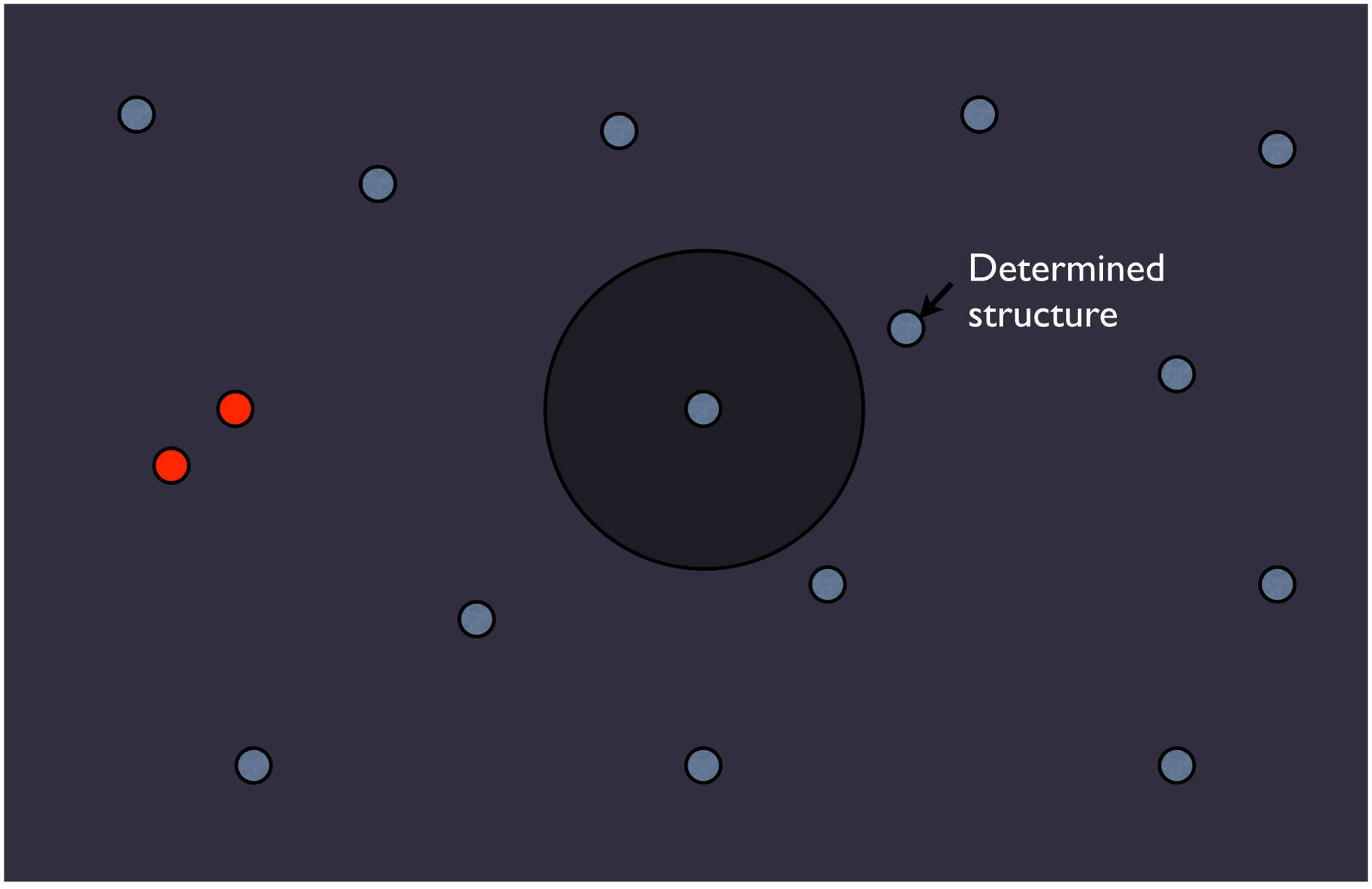




Ben-David et al, 2009

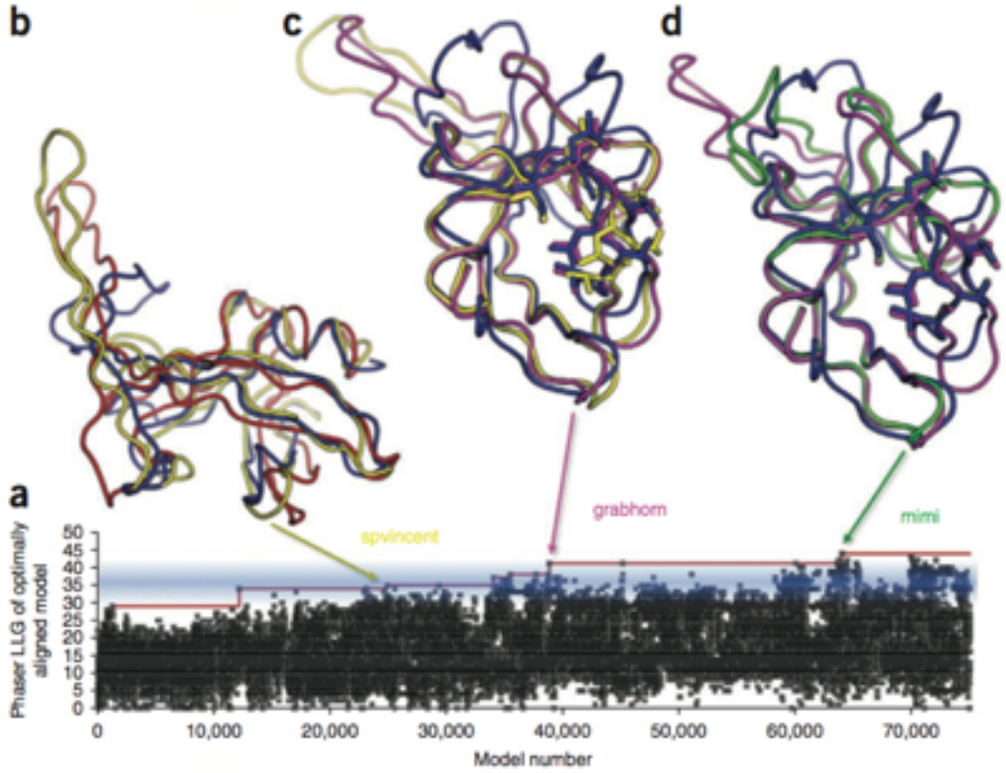
Critical Assessment of protein Structure Prediction

Structural Genomics

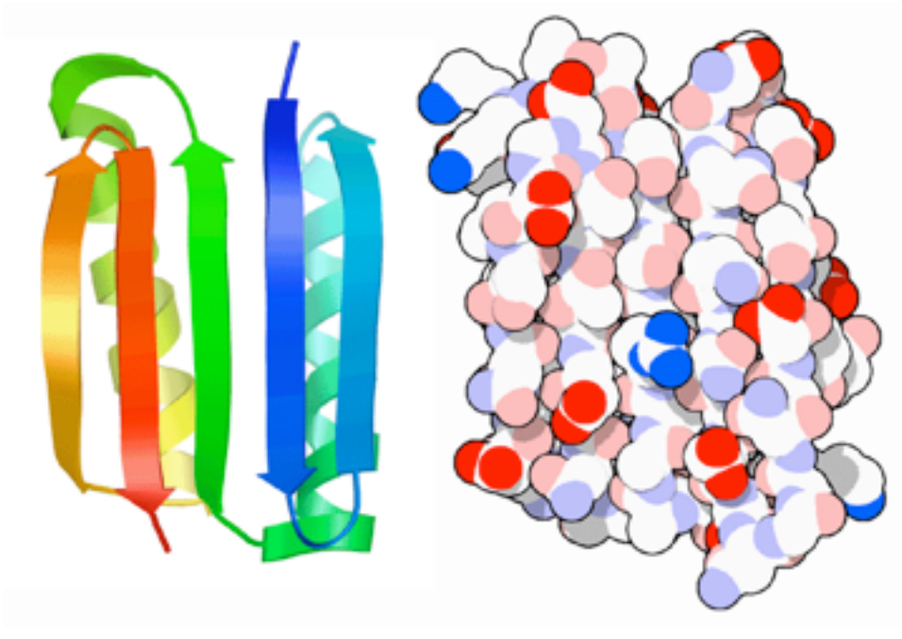


Space of all protein structures

Structure Prediction & Design Successes



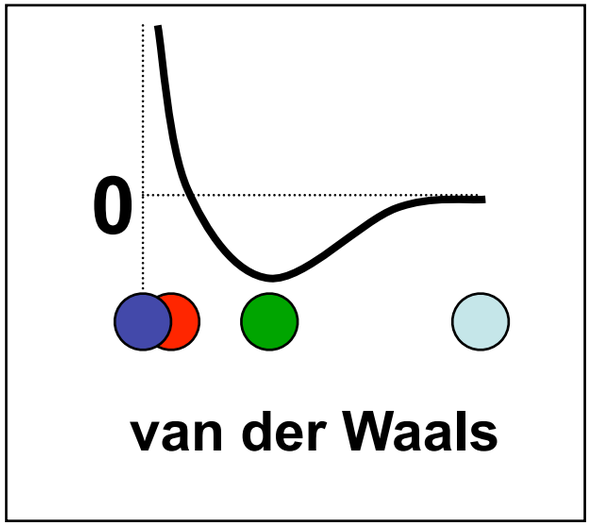
FoldIt players determination the structure of the retroviral protease of Mason-Pfizer monkey virus (causes AIDS-like disease in monkeys). [Khatib et al, 2011]

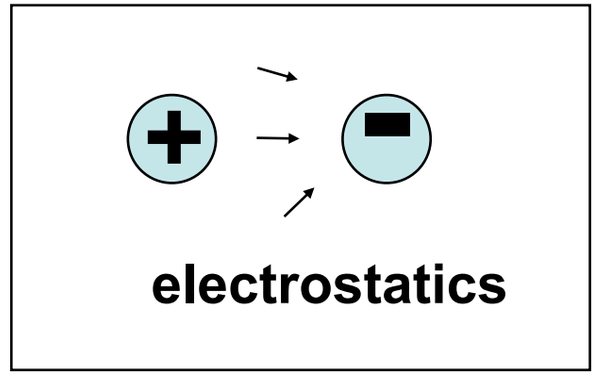


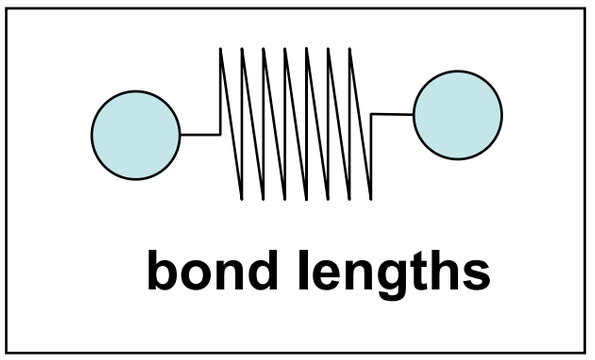
Top7: start with unnatural, novel fold at left, designed a sequence of amino acids that will fold into it. (Khulman et al, *Science*, 2003)

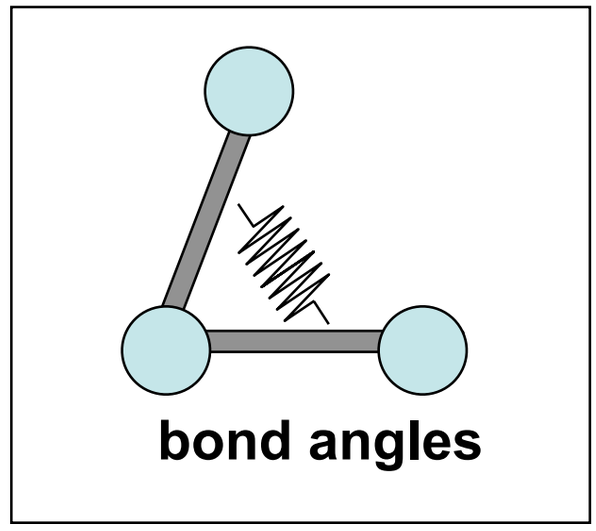
Determining the Energy

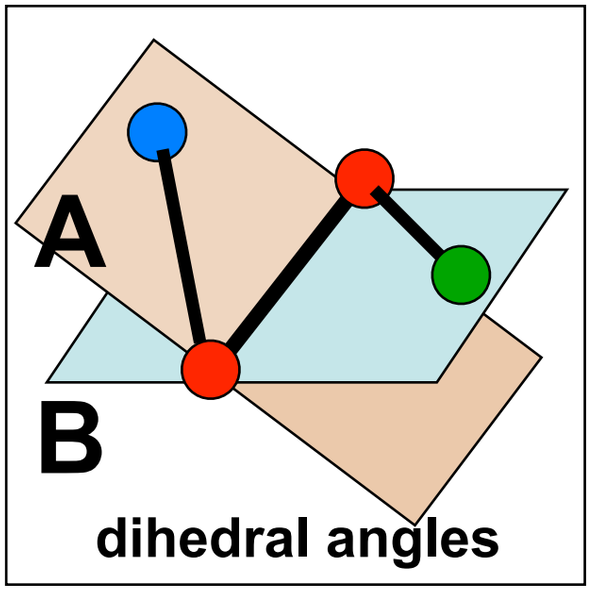
- Energy of a protein conformation is the sum of several energy terms.
- "Force Fields" such as CHARMM and AMBER give explicit approximations to each of these terms.

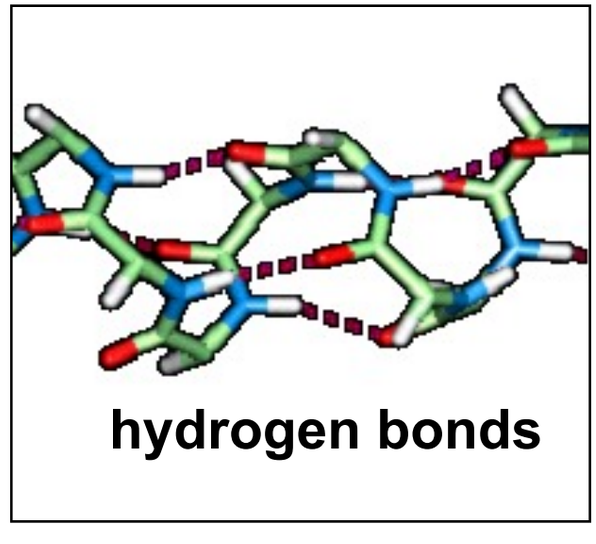












Energy Function (AMBER) Details

calculate the potential energy of a protein structure

$$V(r^N) = \sum_{\text{bonds}} \frac{1}{2} k_b (l - l_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2$$

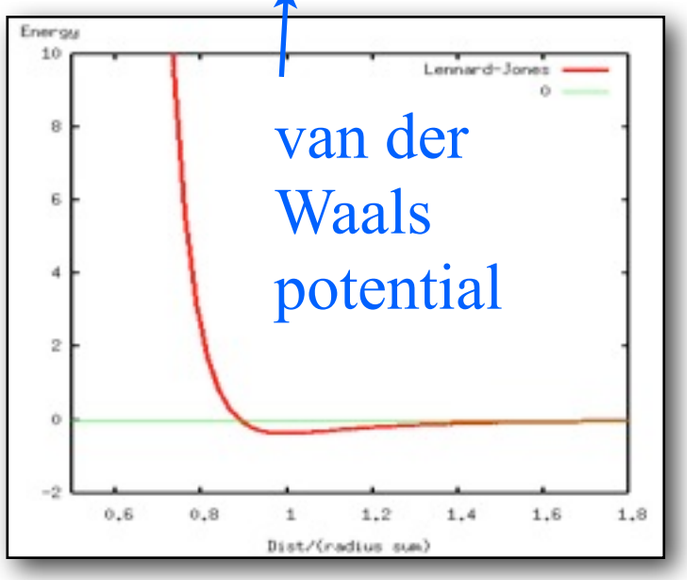
← Hook's law, spring of ideal length l_0 or θ_0 and tension k_b, k_a

$$+ \sum_{\text{torsions}} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$

← function dependent on how much a bond is twisted

$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

Sum over all pairs of atoms



van der Waals potential

electrostatic between particles of charge q_i and q_j: derived from Coulumb's law

equations typeset courtesy Wikipedia

Rosetta@Home Algorithm (High-level)

```
S = linear, unfolded chain
While some part of chain hasn't been moved a lot:
    Move part of S to get structure S'
    If energy(S') < energy(Best):
        Best = S'
    If energy(S') < energy(S):
        S = S'
Else with probability related to energy(S) - energy(S'):
    S = S'
```

Stage 1: uses big moves and a simple energy function

State 2: uses small moves and a complex energy function

Rosetta@Home Algorithm (High-level)

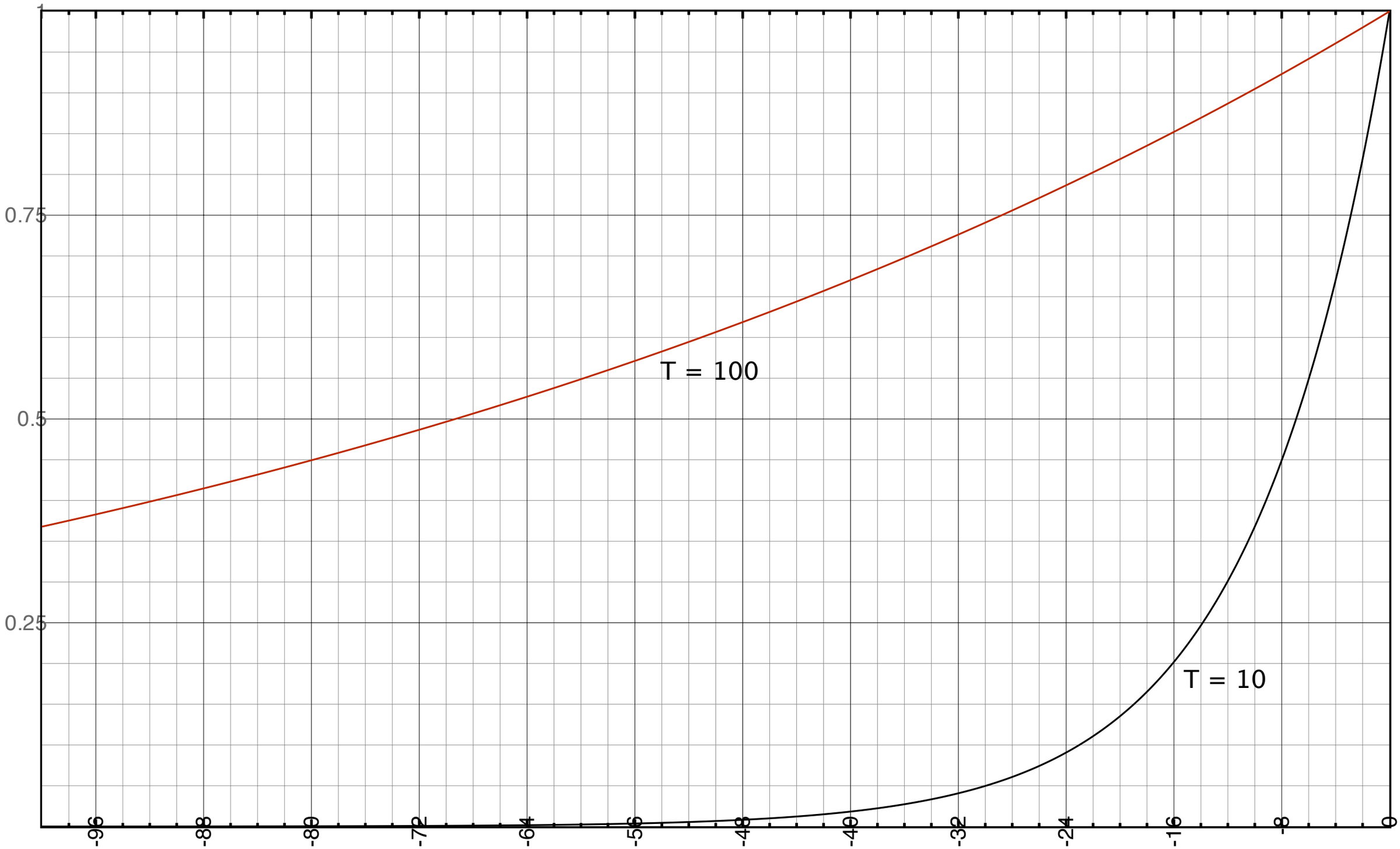
```
S = linear, unfolded chain
While some part of chain hasn't been moved a lot:
    Move part of S to get structure S'
    If energy(S') < energy(Best):
        Best = S'
    If energy(S') < energy(S):
        S = S'
    Else with probability related to energy(S) - energy(S'):
        S = S'
```

exp((energy(S) - energy(S'))/T)

Stage 1: uses big moves and a simple energy function

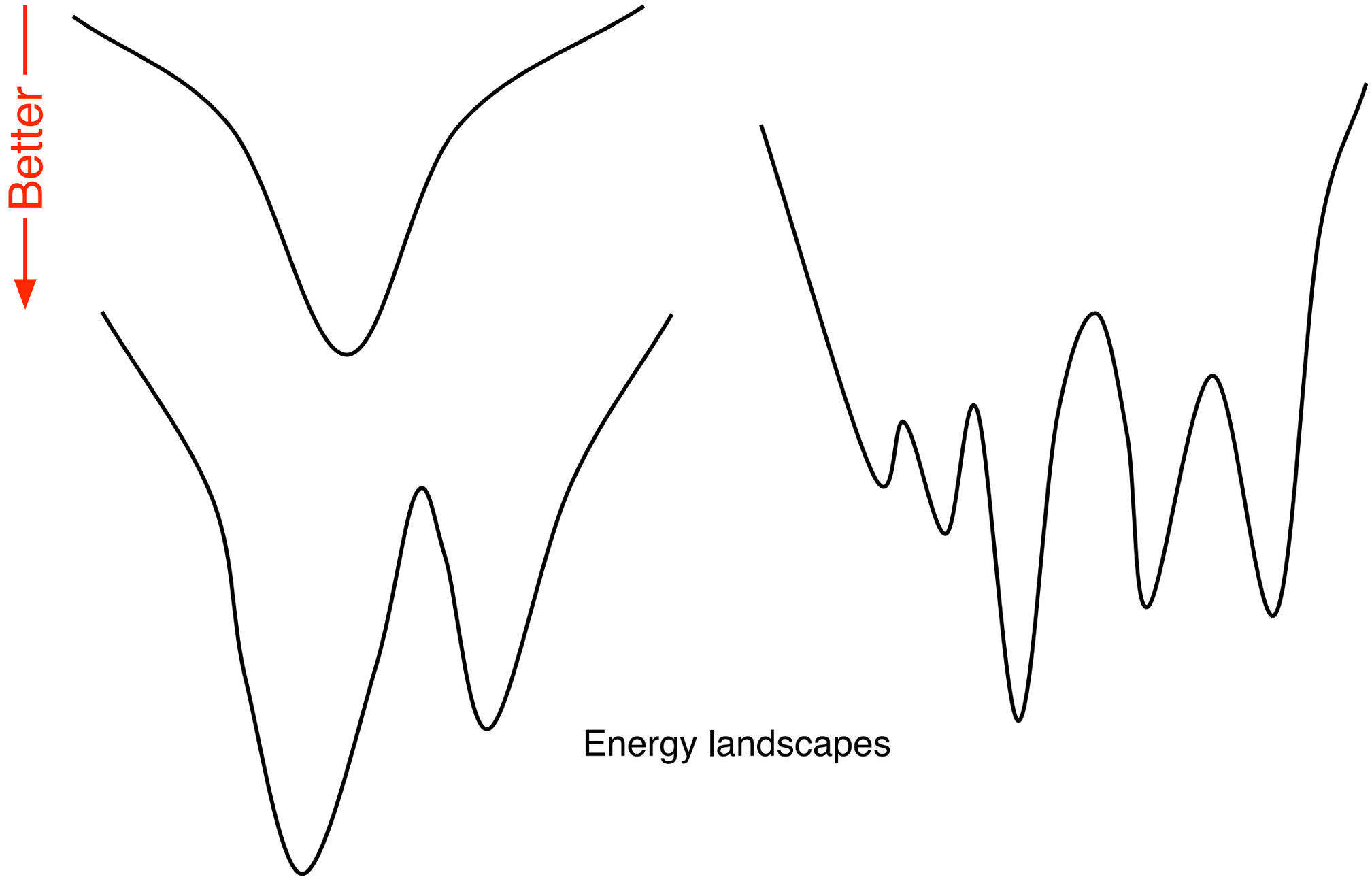
State 2: uses small moves and a complex energy function

exp(Δenergy)/T)



When T is large, more likely to accept a “bad” move.

Avoiding Local Minima



Energy landscapes

At low values of T, you will walk down towards a local minima.
At high values of T, you may jump out of a valley.

Simulated annealing idea: start with a high value of T and decrease over time (cooling schedule).

Summary

Protein structure vital in understanding protein function.

Prediction of protein structure is a very hard computational problem

- Some notable successes over the last ≈10 years
- Based on carefully constructed energy functions

Main algorithmic tool: simulated annealing-like randomized algorithms that efficiently explore the space of conformations