Accurate prediction of protein structures - How to find the exact ground state

Frank Dressel
S. Kobe, M. Schroeder
Biotec, TU Dresden

29. Nov. 2007

Biotechnology Center

From DNA to proteins

- Information is stored in DNA
- DNA is transcribed in nucleus
- RNA is translated in ribosomes into a chain of amino acids

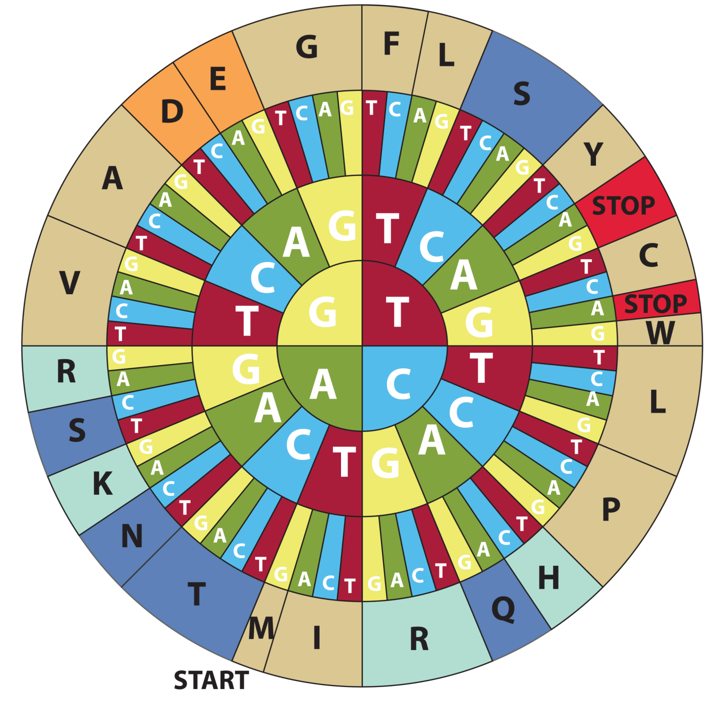
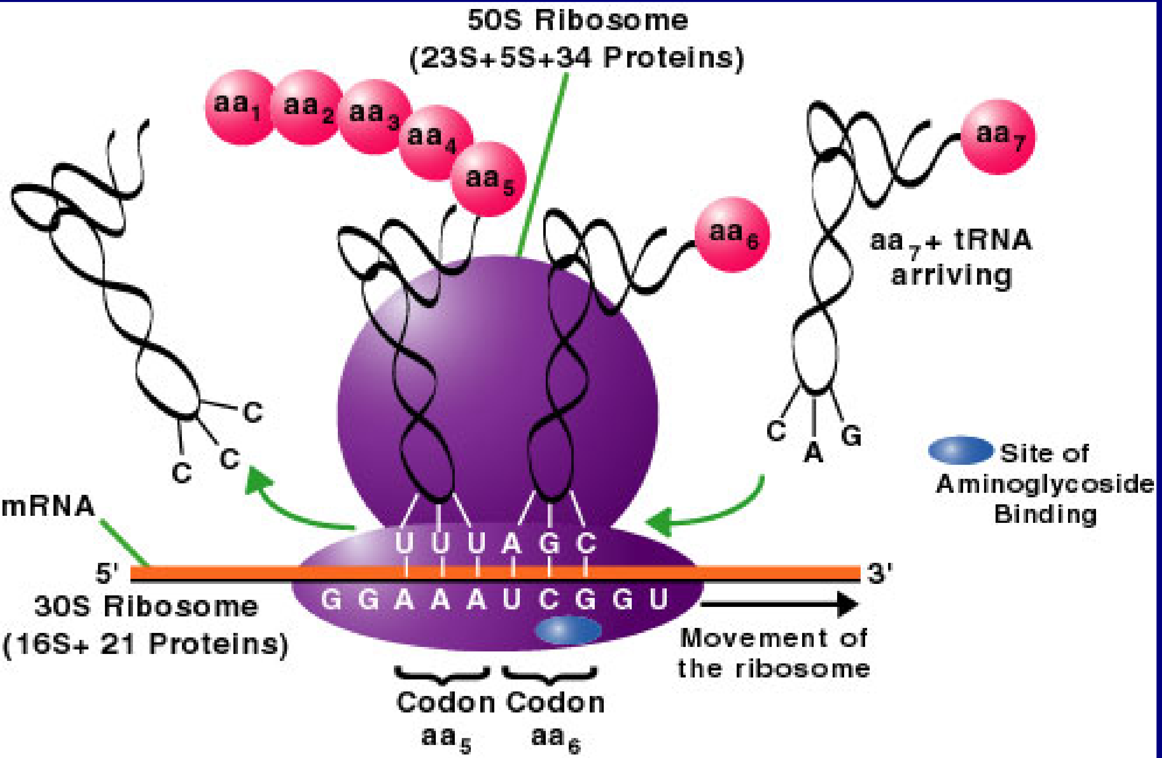
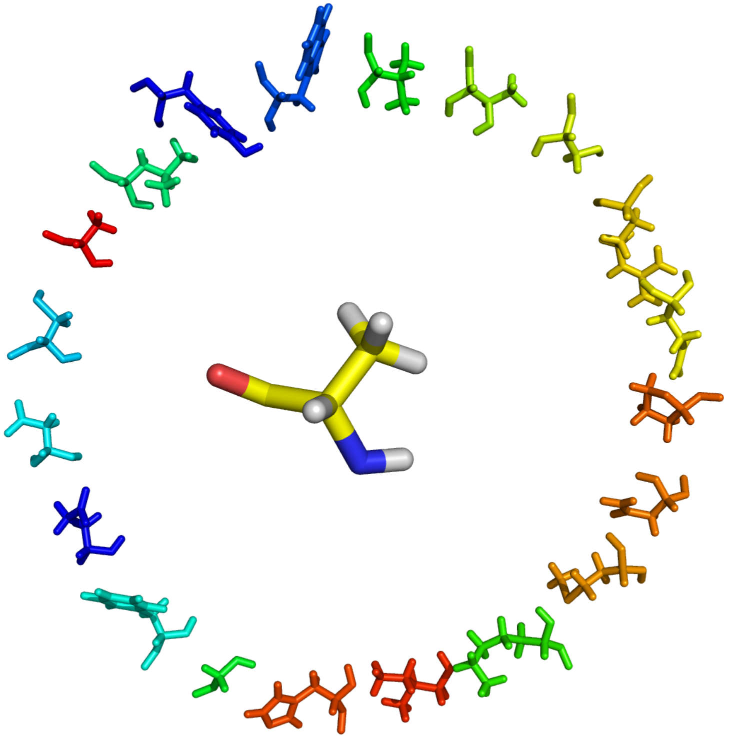


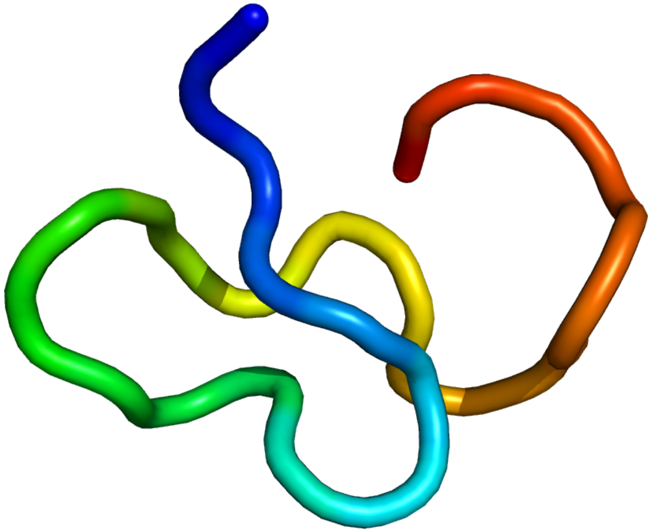
Image: C.Brooksbank, European Bioinformatics Institute



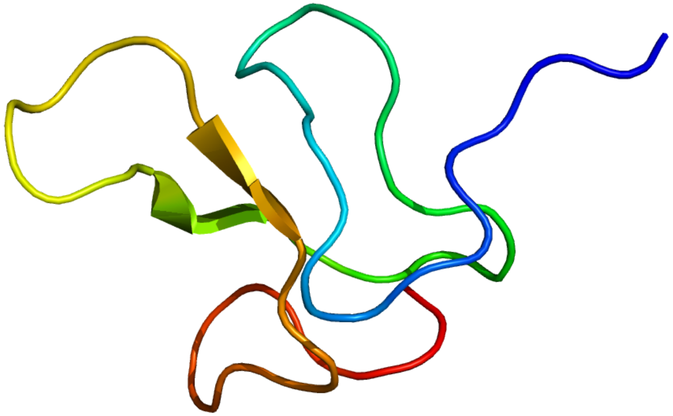


Protein folding

- Chain folds into stable 3d structure with a wide range of structures from disordered to ordered

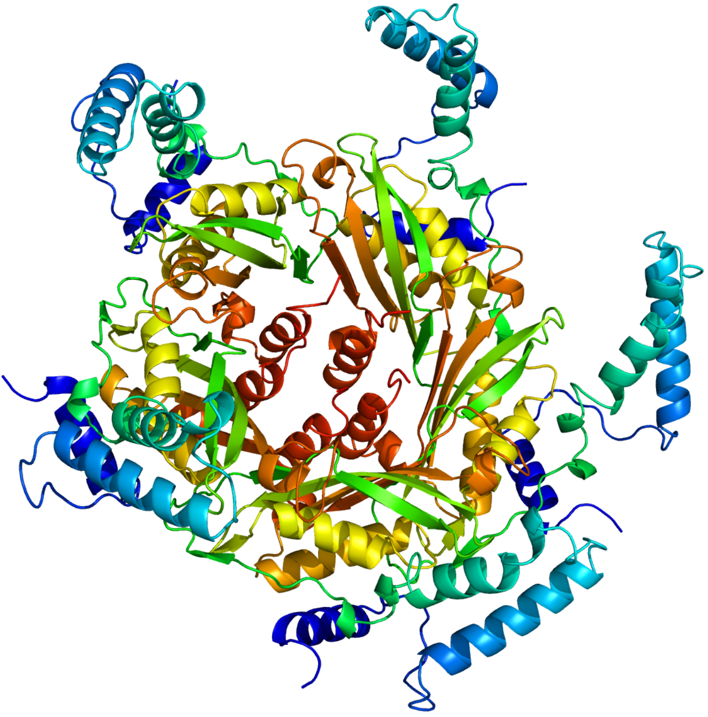


disordered



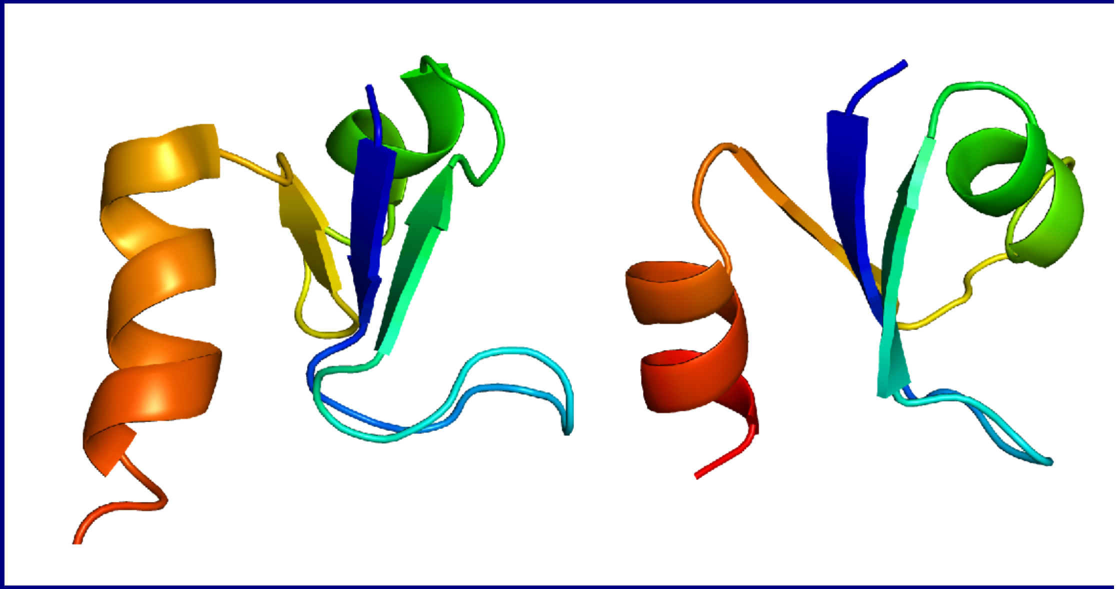
partly ordered

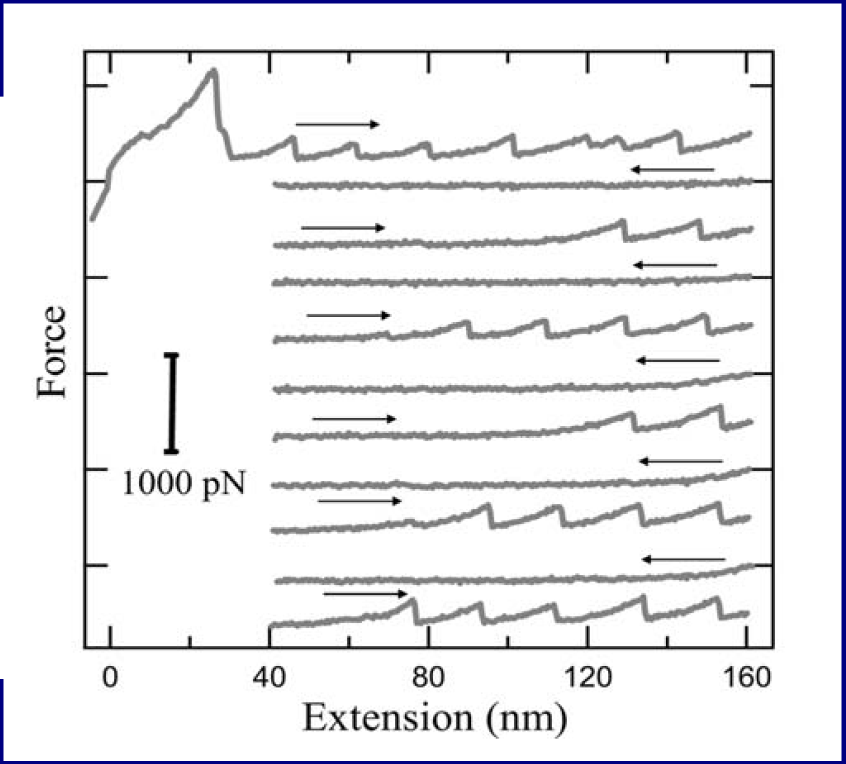
ordered



Protein folding

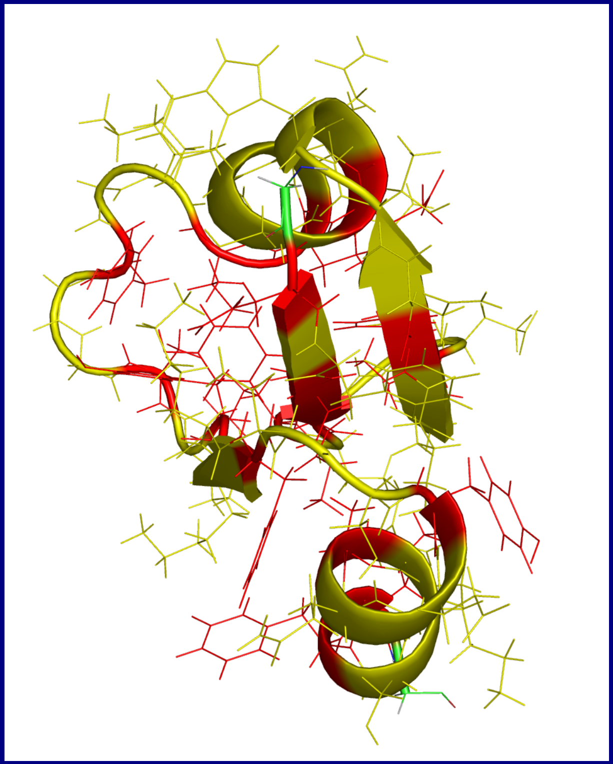
- Is the 3d structure encoded in the sequence?
- Unfolding experiments shows fast refolding
- Different sequences folds into the same structure
    - ~ 1000 different folds





Interaction scheme

- Define inside/outside with respect to C_β orientation
- Define an energy based on this scheme



$$e_i = -\frac{1}{\alpha_i}\ln\left(\frac{n_{inside,i}}{n_{outside,i}}\right)$$

i : Type of amino acid

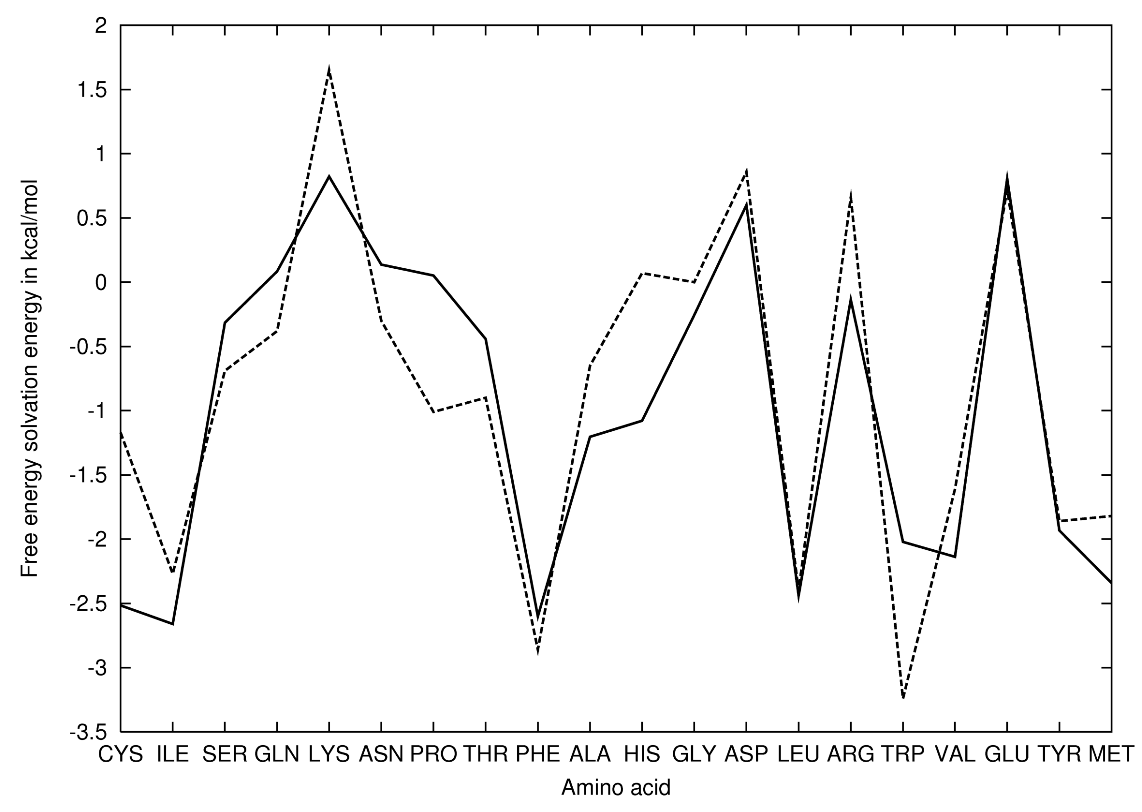
$$e_{ij} = -\ln\left(\frac{n_{ij,observed}}{n_{ij,expected}}\right)$$

Interaction scheme

- **Pairwise interaction scheme:**

$$e_{kl}=\left(e_{i(k)j(l)}+e_{i(k)}+e_{j(l)}\right)\Theta\left(8-\left|\vec{r}_k-\vec{r}_l\right|\right)$$

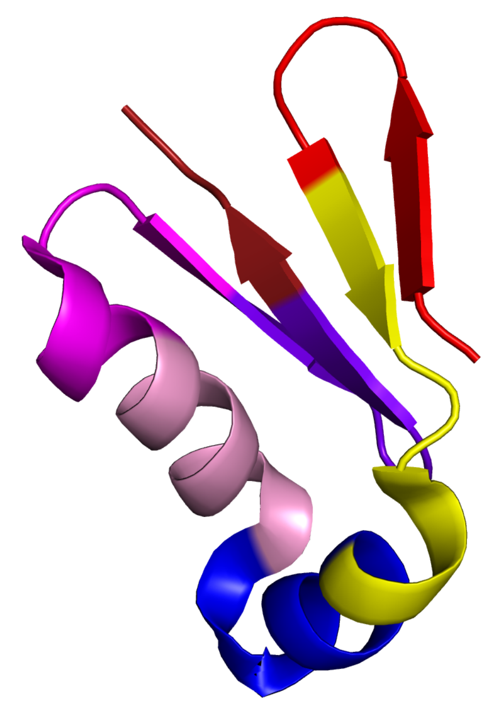
$$k,l: Amino\ \ acids\ \in\ the\ \ sequence$$



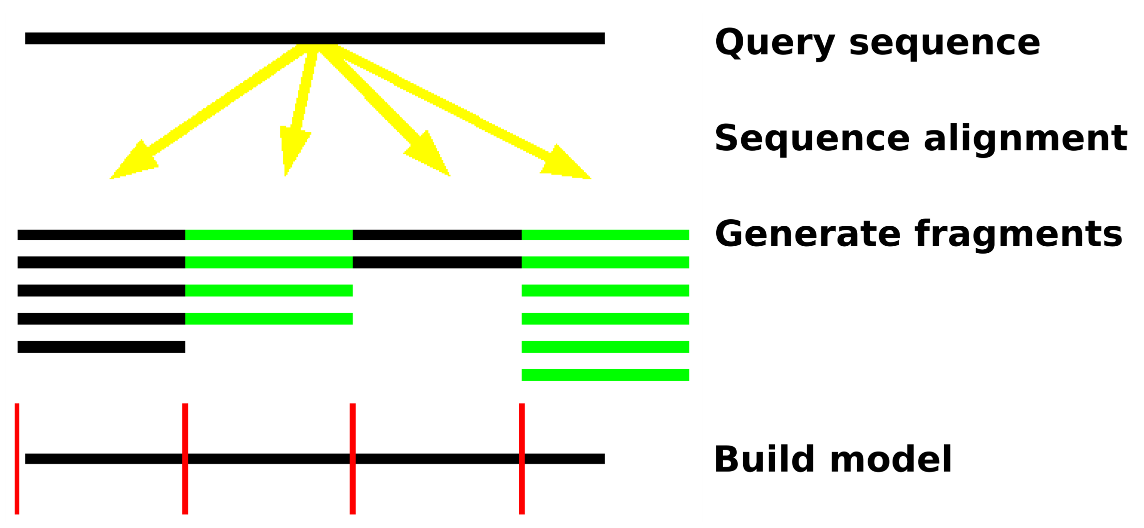
Comparison of experimental (solid) and theoretical (dashed) solvation energies.

Coarse graining the structure

- Cut the query sequence
- Perform sequence similarity search against PDB
- Take real structures from hits in the PDB
- Cluster them in a canonical coordinate system
- Use these fragments to build the protein model



Build the protein model by consecutive fragments



Calculating the ground state - Exact optimization

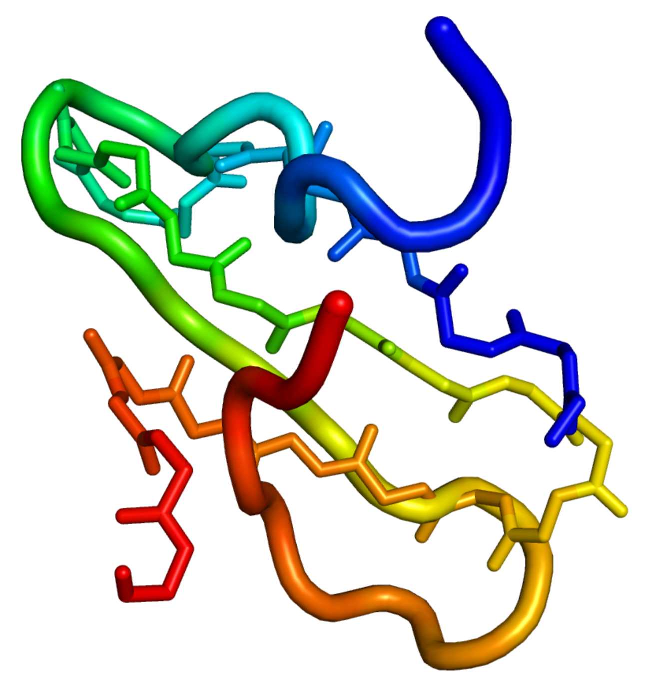
- Use branch-and-bound

$$E\left(\mathrm{subset}_i\right) \leq E\left(\mathrm{subset}_{i+1}\right)$$

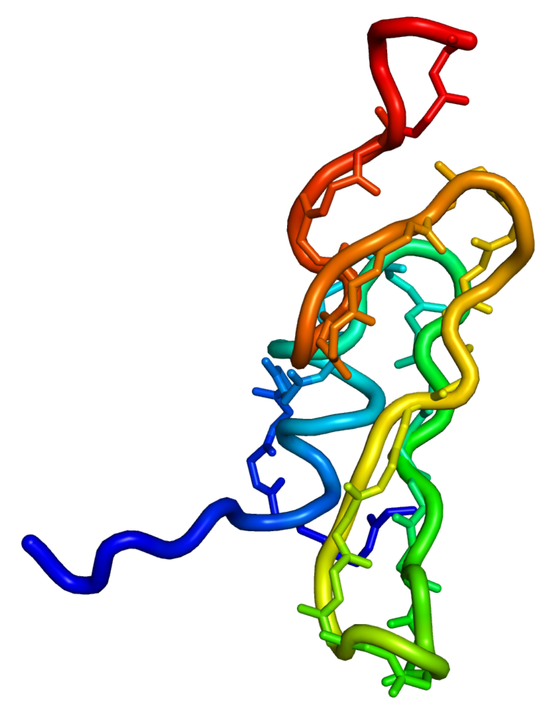
$$\mathrm{subset}_i \in \mathrm{subset}_{i+1}$$

- Give you the optimal solution and prove it
- Calculate all states below threshold

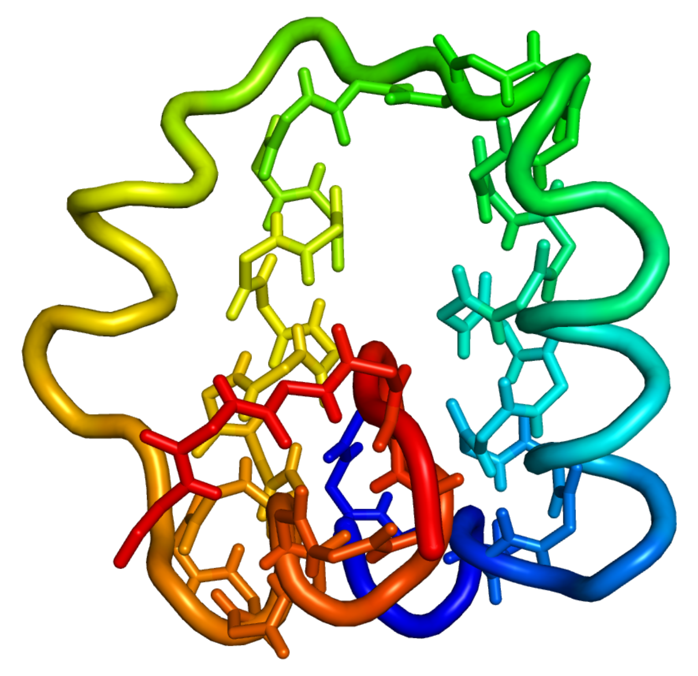
Model ground states vs. PDB



1E0N

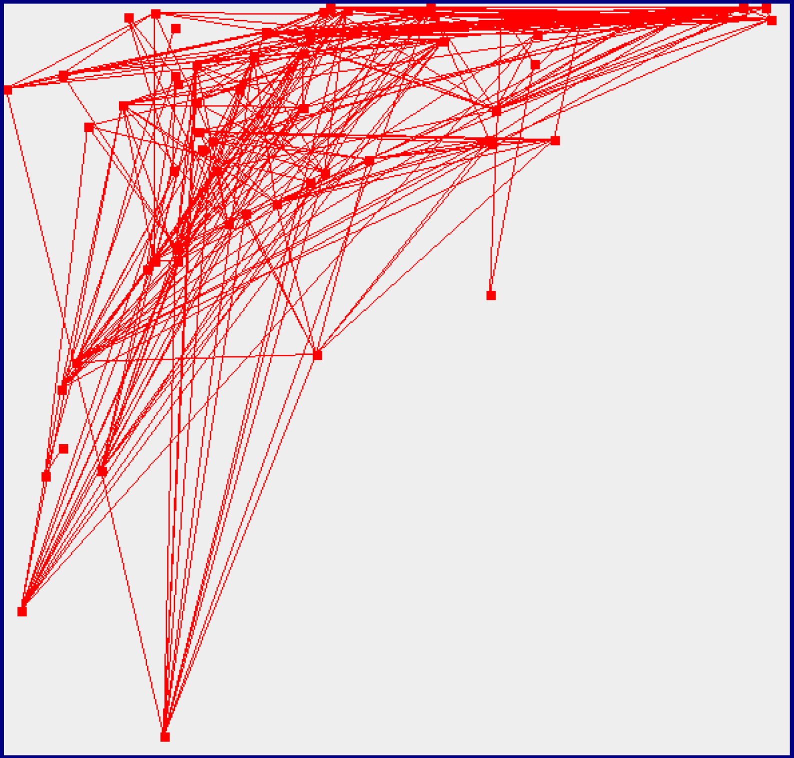


1JL9



1E0L

How many proteins are in ground state?



- **Prepare a set of small proteins from different Scop classes:**
  - **150 for “Small proteins”**
  - **18 for “All α”**
  - **5 for “All β”**
- **Calculate the low lying energy landscape**
- **Search the ground states**

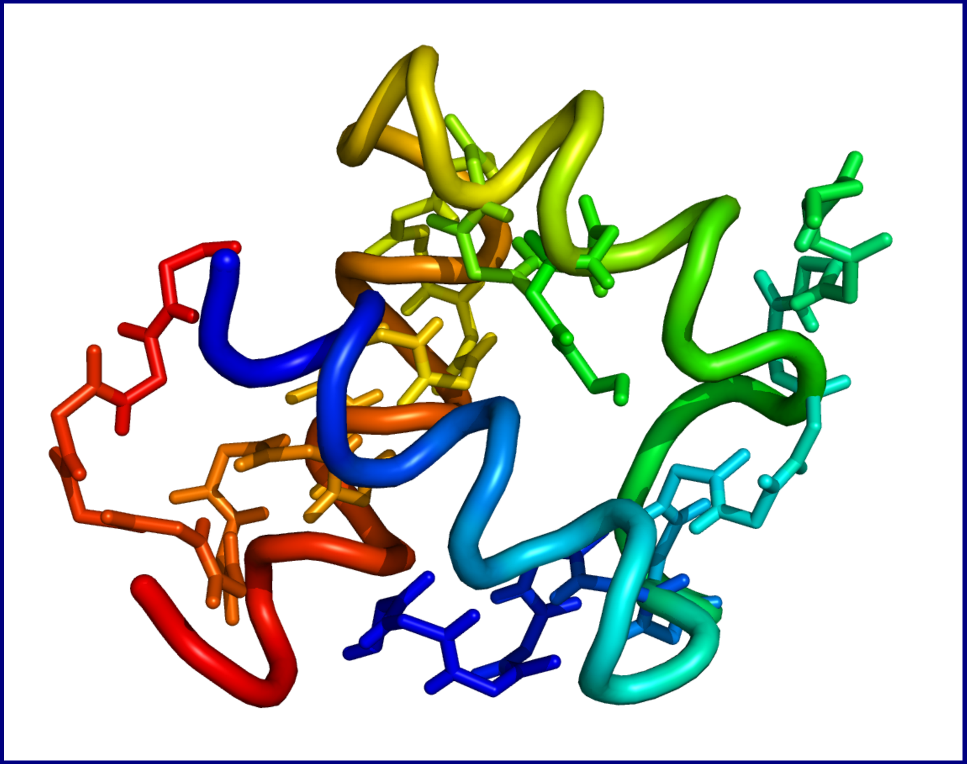
Small proteins	
Ground state < 6.5 A	106
Structure with < 6.5 in the landscape	131

All α	
Ground state < 6.5 A	14
Structure with < 6.5 in the landscape	18

All β	
Ground state < 6.5 A	3
Structure with < 6.5 in the landscape	5

What are the reasons for “Non ground-state-ness”?

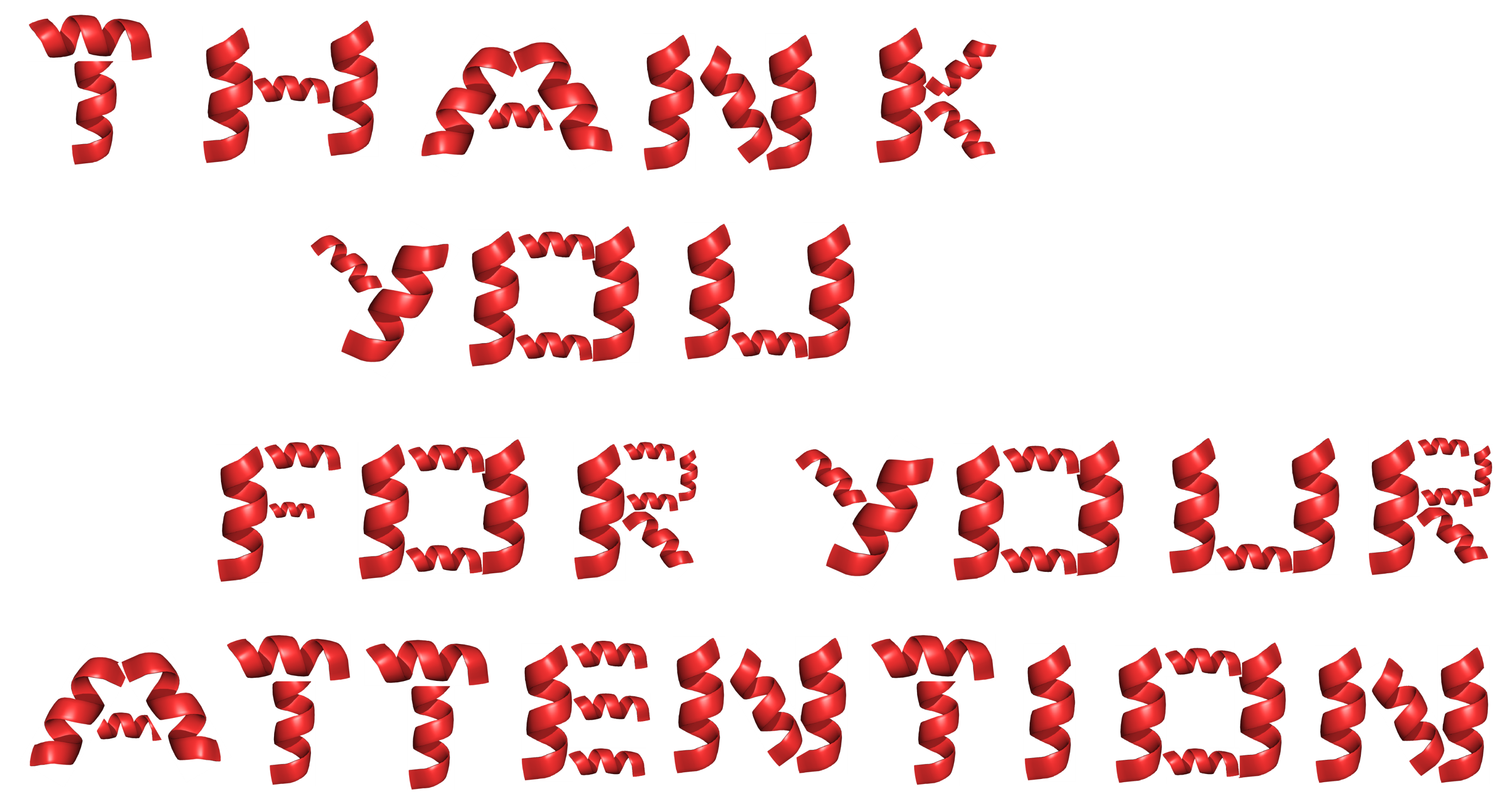
- Helices are in wrong orientation
- Other secondary structure elements are in wrong orientation
- Protein can not be modelled
- Native state ≠ Ground state



1JJS: Large structural changes upon ligand binding (Lin et al., 2001)

Discussion

- In most cases, the ground state corresponds to native state
- Nevertheless: There are some proteins (mostly in “Small proteins”), where native state doesn't correspond to ground state
- Protein should be able to perform structural changes
    - Small energy barriers between native and non-native state


THANK
YOU
FOR YOUR
ATTENTION