JÜLICH
FORSCHUNGSZENTRUM

Statistical Modeling of Protein Folding

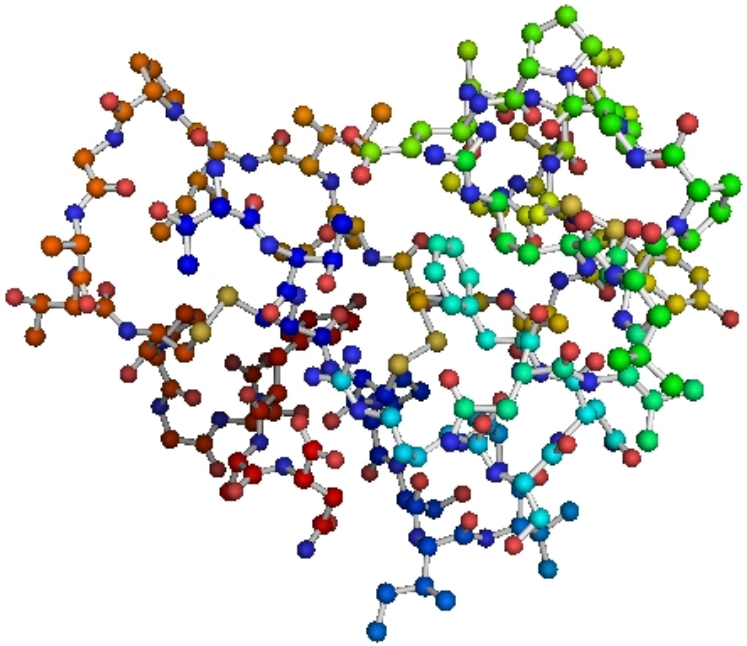
Guest Student Programme 2010

October 20, 2010 | Julie Krainau

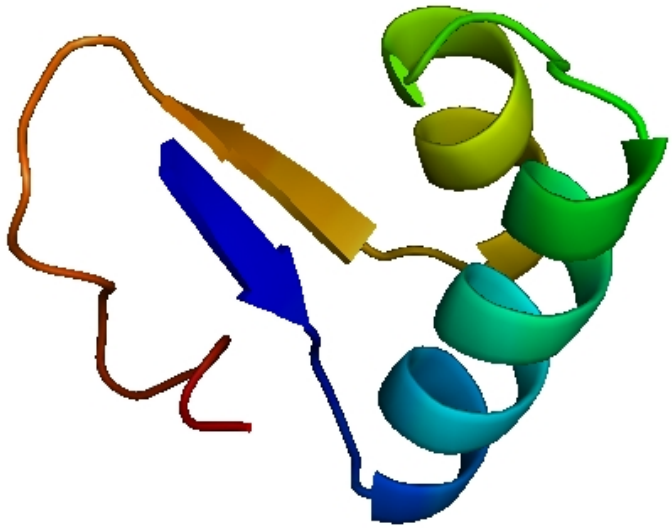
What are Proteins Made of?

- Consist of a chain of amino acids
- 20 different amino acids
- Amino acids of a protein are named residues

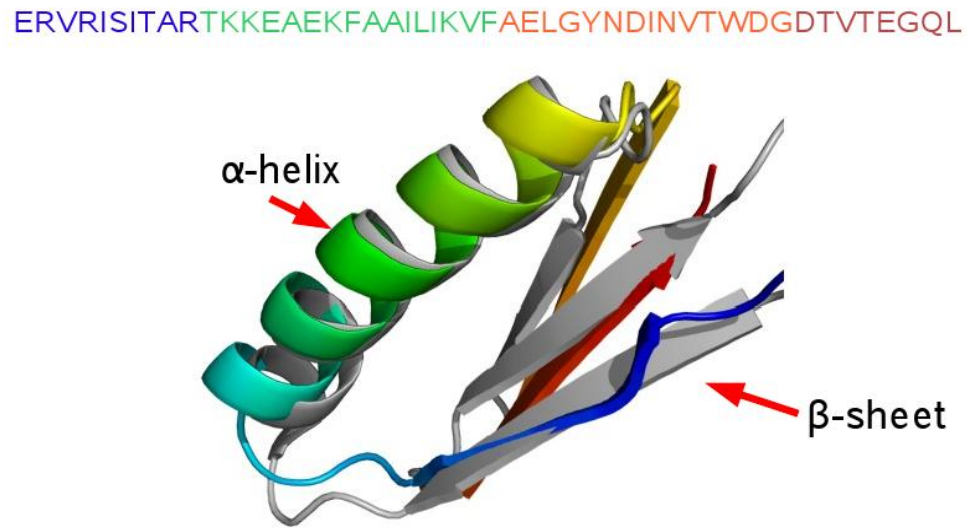
How Do They Look Like?



How Do They Look Like?



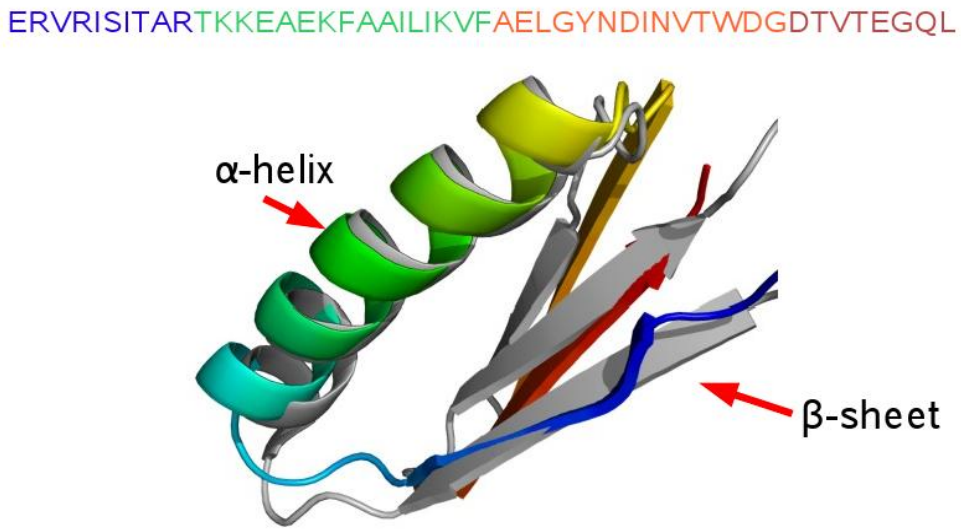
How Do They Look Like?



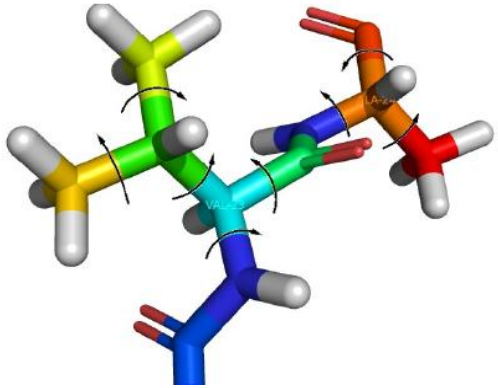
- *Primary Structure:* Sequence of residues
- *Secondary Structure:* Structural elements
- *Tertiary Structure:* 3d Struture of chain

Secondary Structure Elements

α-helices and β-sheets are based on a defined arrangement of hydrogen bonds caused by folding



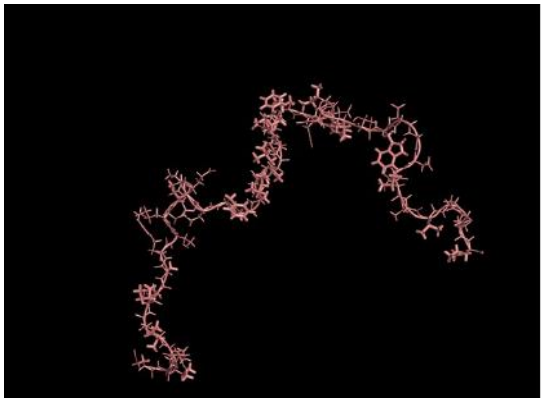
Protein Structure



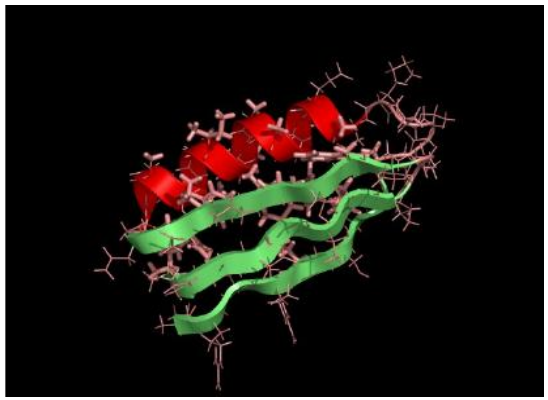
At room temperature the most relevant degrees of freedom are some torsion angles

Protein Folding

- Degrees of freedom of a tiny protein in the order of 100
- Even a tiny protein chain with a sequence length of 20-30 has an astronomical number of possible structures



Protein Folding



- But in the right environment these molecules can fold quickly in well defined 3d-structures
- Timescale is varying from μs - s

Why is Protein Folding Important ?

- Its folded shape is highly related to its function
- Misfolding of proteins currently thought to be related to neurodegenerative diseases

Why is Protein Folding Important ?

- Its folded shape is highly related to its function
- Misfolding of proteins currently thought to be related to neurodegenerative diseases
- Big economic problem

BBC Mobile | News | Sport | Weather | Travel | TV | Radio | More | Search the BBC

NEWS HEALTH

Home | UK | Africa | Asia-Pac | Europe | Latin America | Mid-East | South Asia | US & Canada | Business | Health | Sci/Environment | Tech | Entertainment | Video

21 September 2010 Last updated at 13:07 GMT

What's the true cost of dementia?

By Michelle Roberts
Health reporter, BBC News

Experts are warning that dementia is the greatest health and social crisis of the century as its global financial burden continues to escalate.

Family support is often heavily relied on

The World Alzheimer Report says dementia costs will amount to more than 1% of the world's gross domestic product this year at $604bn (£388bn).

To put this massive sum into context, if dementia were a country it would be the world's 18th largest economy.

But where exactly does this colossal cost come

Top Stories

NZ adds to India's Games pressure

France faces new wave of strikes

Wen hits out over yuan criticism

New horned dinos found in Utah

Features & Analysis

Death row
Why is a woman being executed in the US such big news?

Foul play
Sumo fights to restore image after Japan mafia links

Arctic ambition
Canada stakes claim to Northwest Passage and mineral wealth

How Can Folding Be Simulated?

Key Concept :

Native structure is associated with the free energy minimum of the system

- Formulate an interaction potential
- Find the free energy minimum

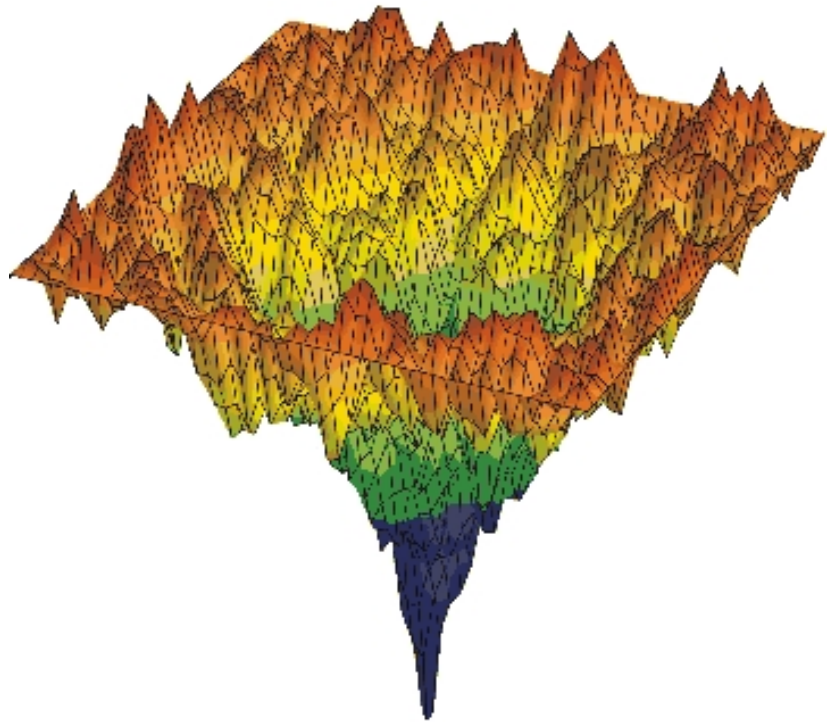
Creating the Interaction Potential

All interactions between the atoms have to be taken into account:

$$E_{tot} = E_{ExV} + E_{El} + E_{HB} + E_{HP}$$

- Excluded volume
- Hydrogen bonds
- Hydrophobicity
- Electrostatic interactions

Energy Landscape



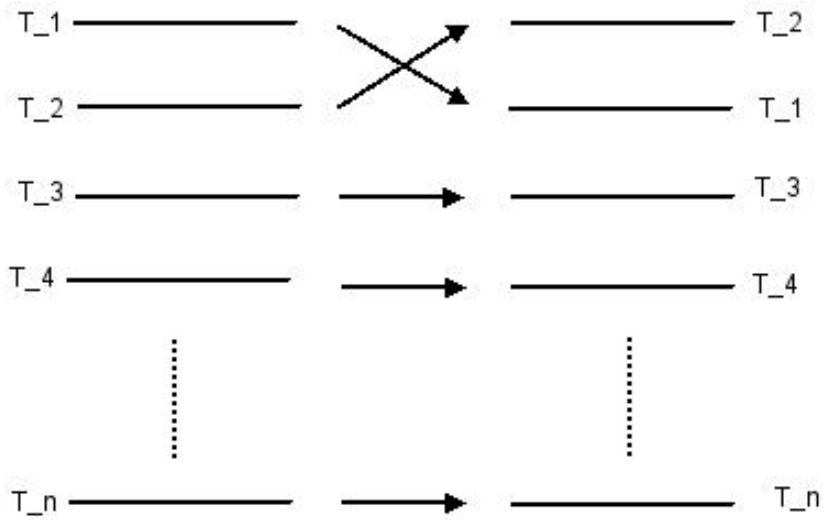
Finding the Minimum Using Monte Carlo Method

1. Algorithm starts off with an initial conformation with energy E_{old}
2. One degree of freedom is changed randomly and a new conformation is created with energy E_{new}
3. $\Delta E = E_{old} - E_{new}$
4. New state is accepted with a probability of $e^{-\frac{\Delta E}{kT}}$
   - If the new state is accepted, set $E_{old} = E_{new}$
   - If the new state is rejected, reverse E_{new}
5. Continue with step 2

Pitfall

- If the temperature is too low, it is very likely to get trapped in a local minimum
- If the temperature is too high, it is very likely to leap over the global minimum

Parallel Tempering



- N replica of the target protein
- Start independent MC run for each replica at different temperatures T_i
- After a defined number of iterations the temperatures of a pair of different runs is exchanged with a certain probability

ProFASi: The Protein Folding and Aggregation Simulator

- Open source C++ package for Monte Carlo simulation of protein folding and aggregation
- Developed by Sandipan Mohanty and Anders Irbäck, developed further by Simulation Laboratory Biology at JSC
- Implicit water
- Intermediate resolution : All atom protein model with fixed bond length and bond angles
- Physics based potential
- It is able to simulate both: Folding of α-helices and β-sheets

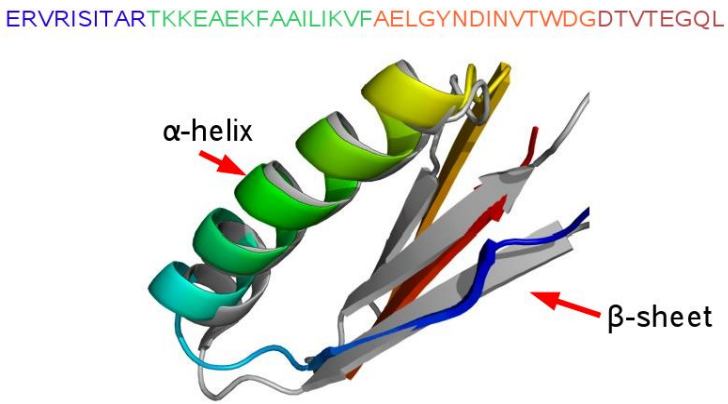
ProFASi: The Protein Folding and Aggregation Simulator

Tricks used to improve performance

- Finite range interactions: use cut-off and cell list method to avoid unnecessary calculations
- Monte Carlo: calculate only what has changed

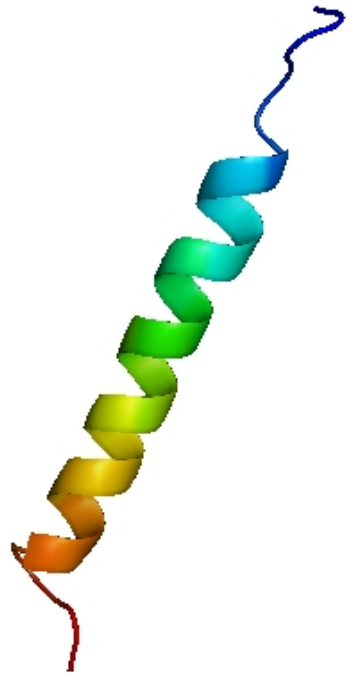
Similarity Measures

- Comparison by root-mean-square deviation (rmsd) of simulated current state and experimental native state

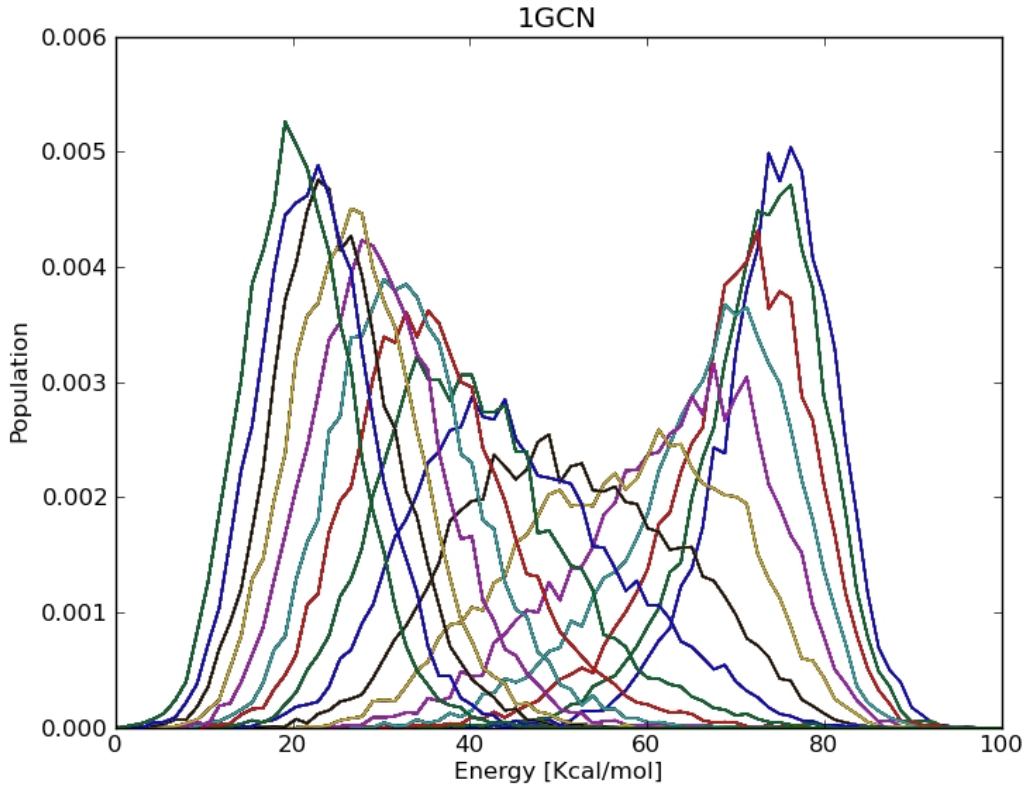


1GCN

- 29 residues, 471 atoms, 135 degrees of freedom
- 16 temperatures : 270 - 370 Kelvin
- 16 processors
- 2 days, 10.000.000 MC sweeps

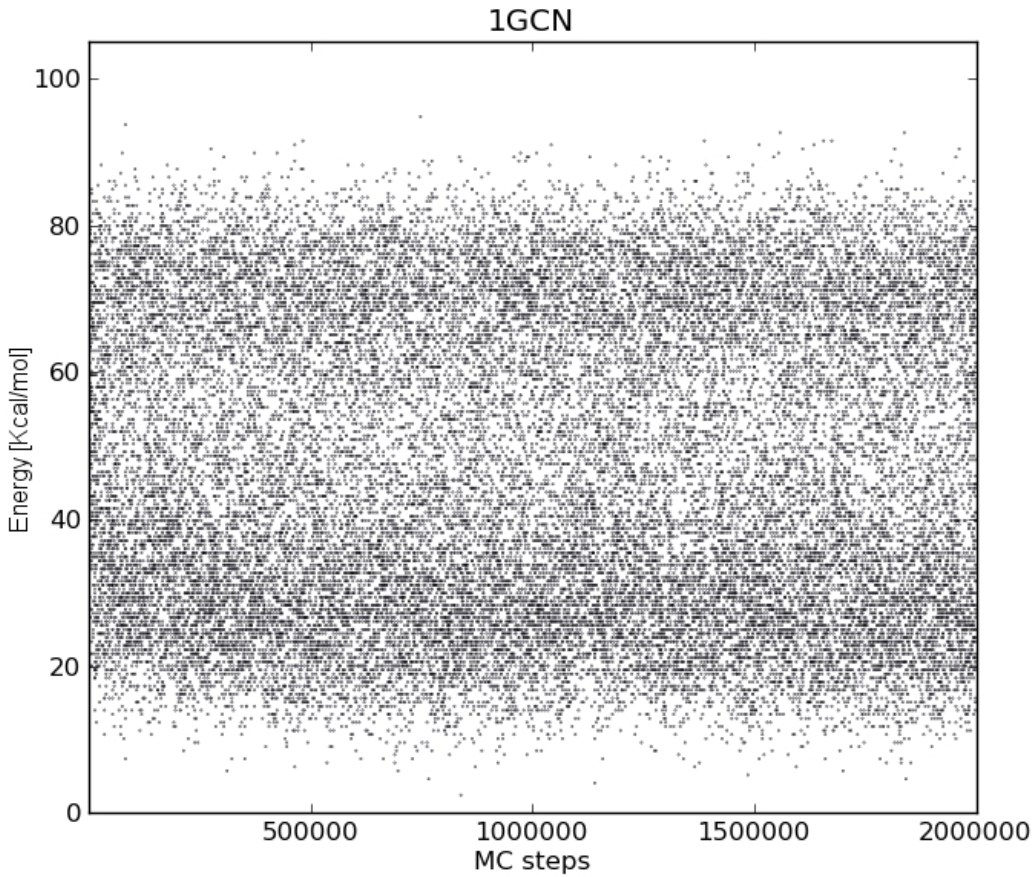


1GCN



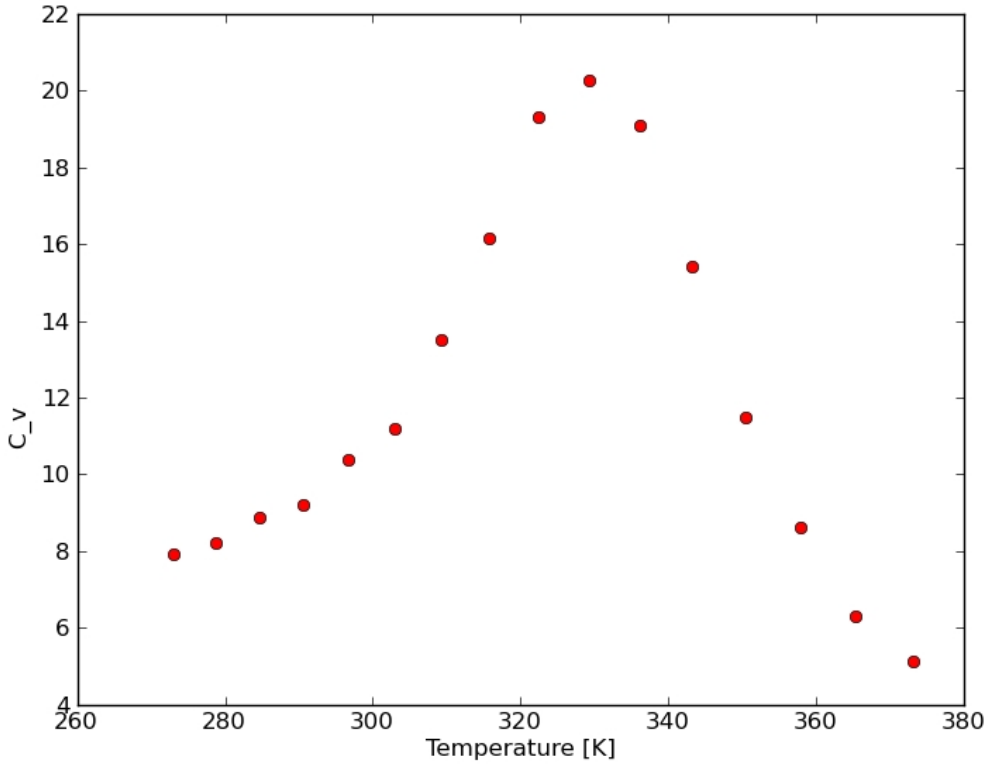
- It has two states $\rightarrow$ protein has folded

1GCN



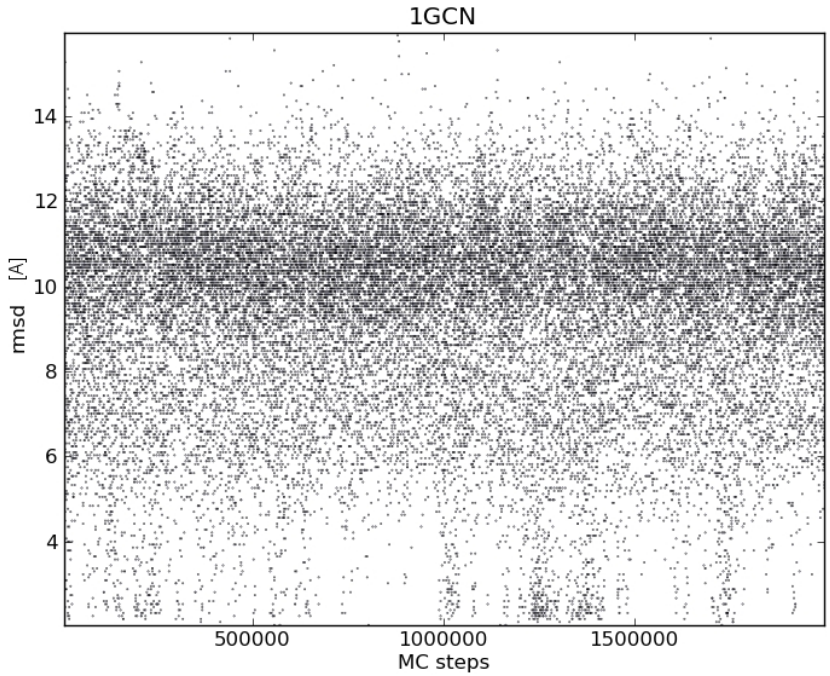
- It did not get stuck

1GCN



- There is a transition from the unfolded into the folded state

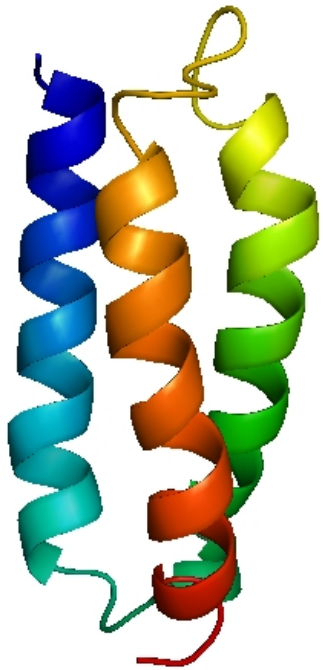
How is the resemblance with the experimental structure ?



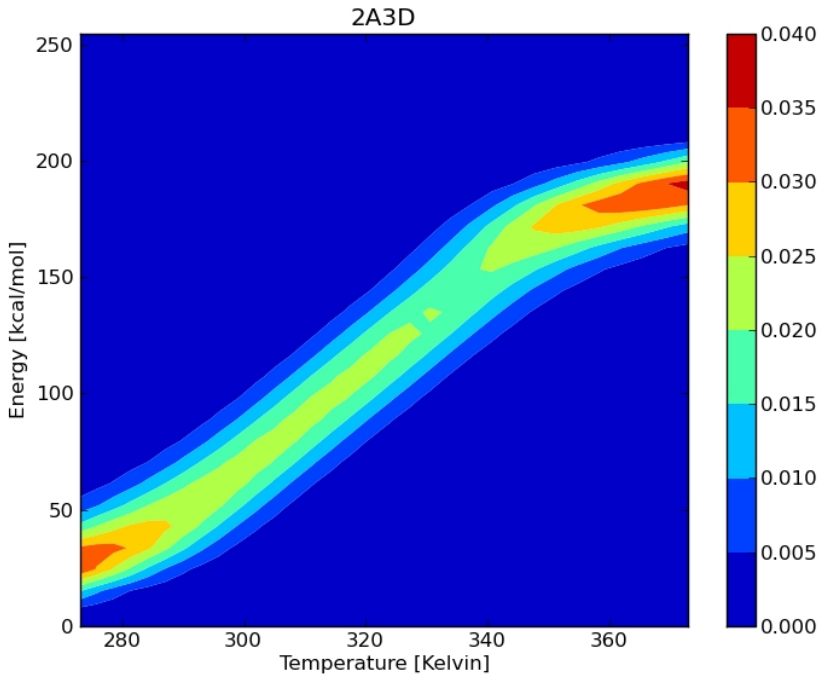
- The lowest rmsd sampled is 1.6 Å
- Most of the time the rmsd was in the range of 10-12 Å

2A3D

- Designed protein, 73 residues, 1140 atoms, 333 degrees of freedom
- 32 temperatures : 270 - 370 Kelvin
- 1024 processors
- 1 day, 1.300.000 MC sweeps

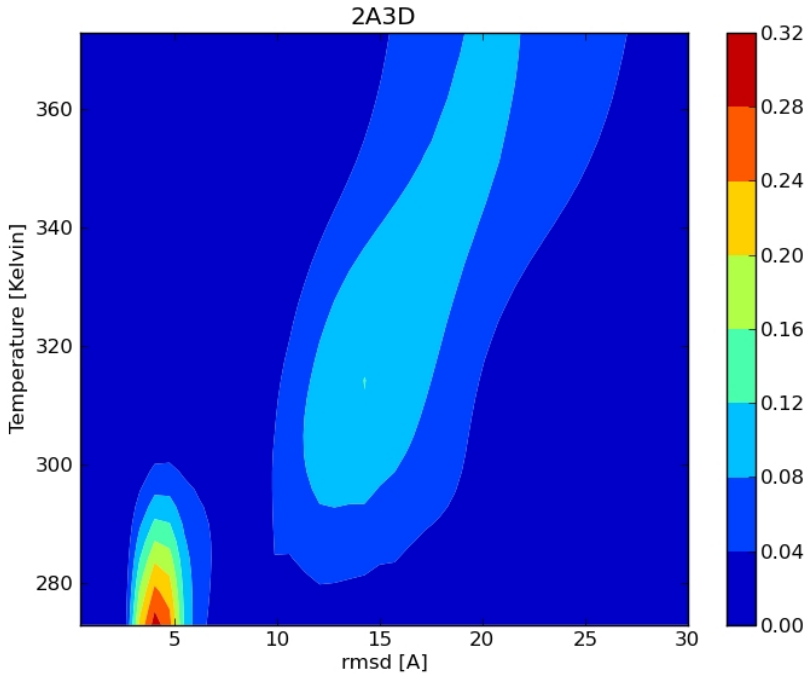


2A3D



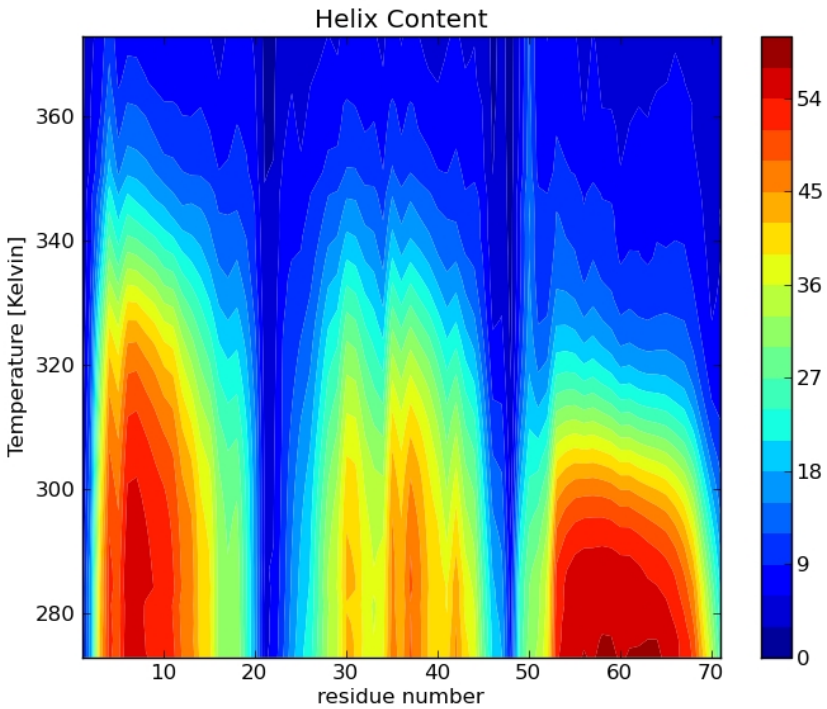
- Two distinct energy states $\rightarrow$ protein has folded

2A3D



- Maximum at low rmsd $\rightarrow$ it has folded into the correct structure

2A3D

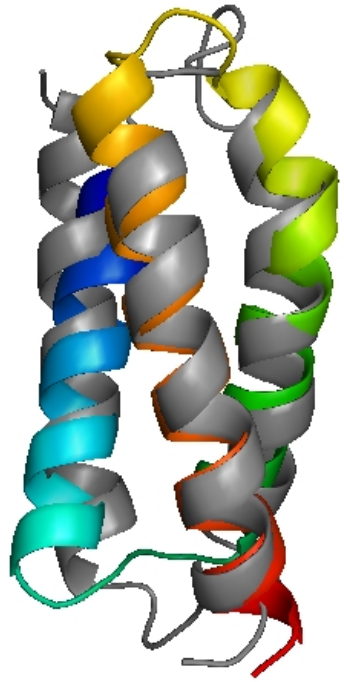


- Central helix is less stable than the two other helices
- First helix folds at higher temperatures

2A3D

- movie17 movie 118
- Movie shows the trajectory of accepted conformations
- It is observable that the different helices do not fold simultaneously
- White hydrophobic parts arrange between the helices, in order to hide from water

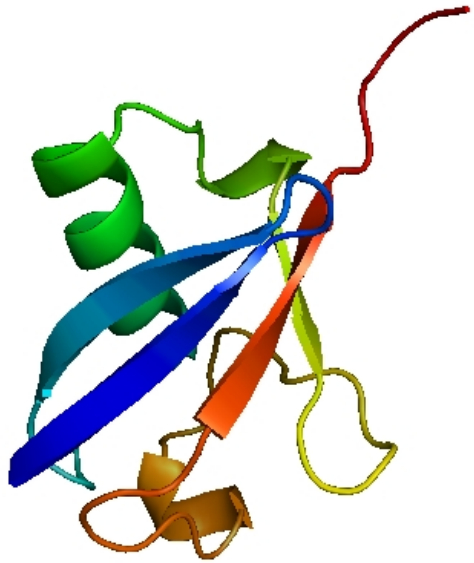
2A3D



- Lowest energy structure has rmsd of 3.7 Å
- Good result for protein of this size

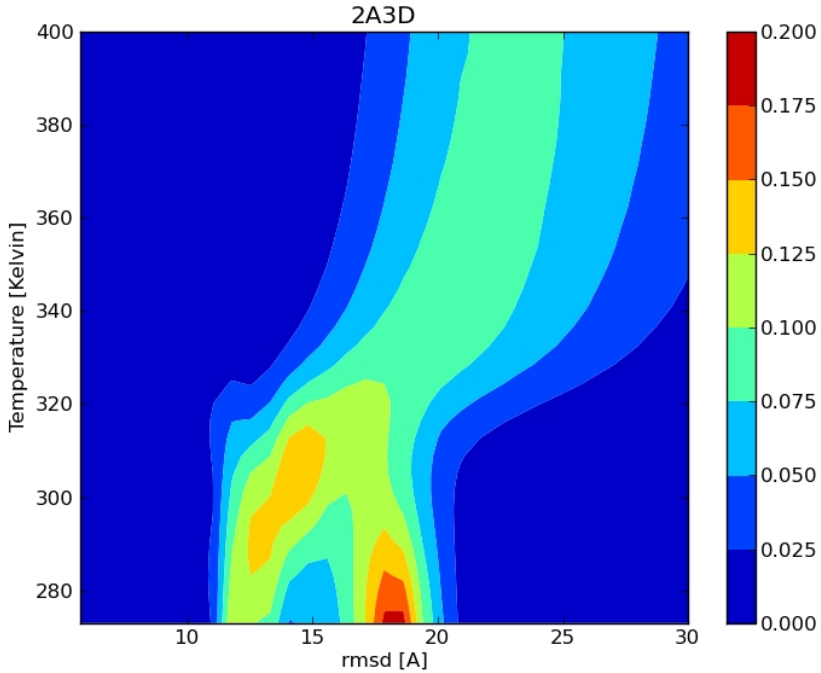
1UBQ

- 76 residues, 1231 atoms, 368 degrees of freedom
- 1 β-sheet, 1 hairpin
- 3 helices
- 32 temperatures in range of 273 - 400 Kelvin
- 64 processors
- 1 day, 1.679.000 MC sweeps



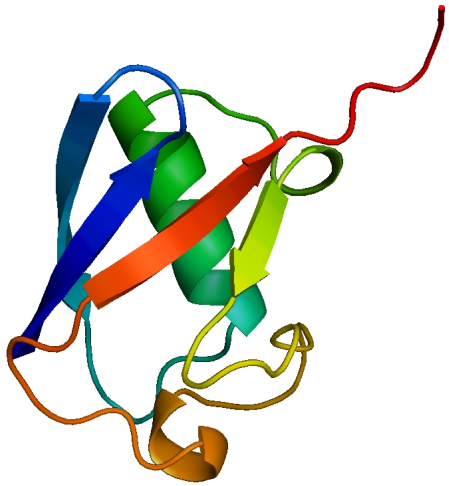
1UBQ

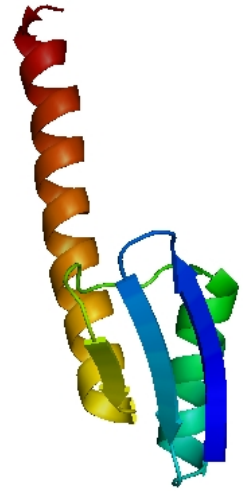
- So far, none of the replica folded into the correct structure



- What would it take for the molecule to fold ?

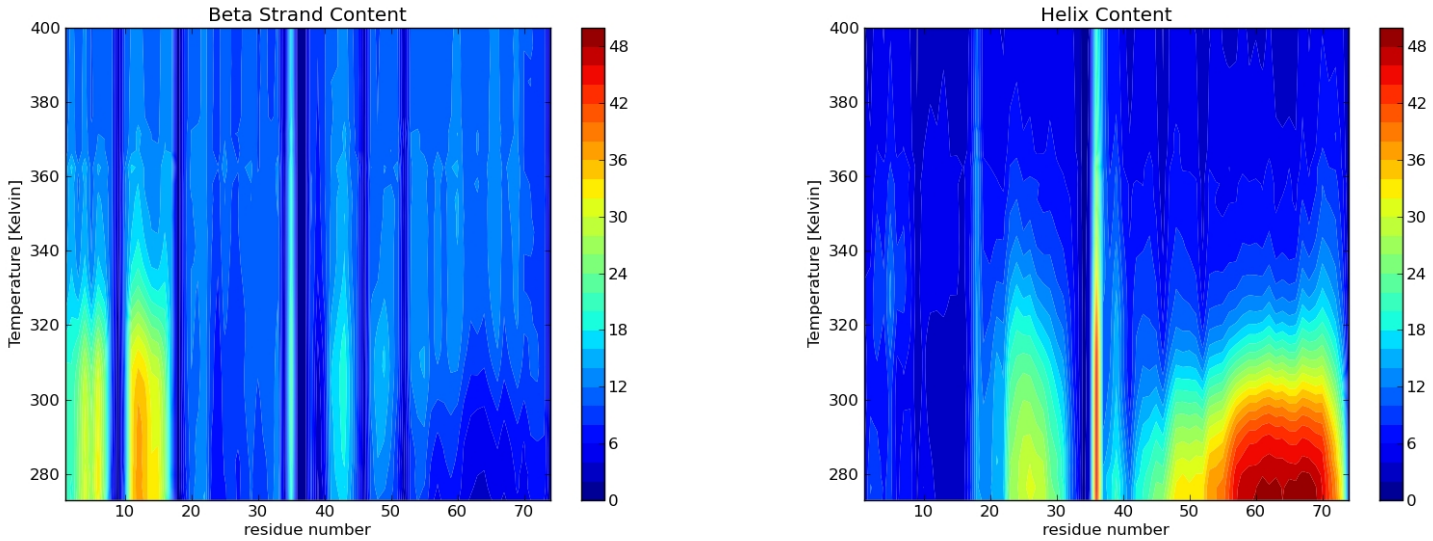
What Went Wrong with 1UBQ ?





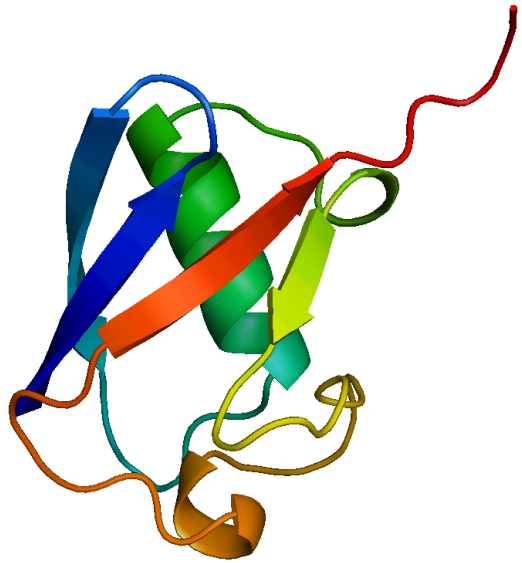
Instead of folding an β-sheet it folded an α-helix at the end of protein

What Went Wrong with 1UBQ ?



- At lower temperatures the red α-helix is more stable than the hairpin

What Went Wrong with 1UBQ ?



- Once the hairpin and the yellow β-strand have folded, they get attached $\rightarrow$ the end of the protein does not have a place to arrange between them
- In this case it is energetically favourable to form an α-helix

Simulation with Constraints

Dihedral Constraints

- Instead of changing degrees of freedom/dihedral angles randomly, choosing new angles of a defined distribution
- Angles which are more likely to be in the native conformation are tried more often

Simulation with Constraints

Why Using Constraints ?

- If some features of the molecule are already known, they can be used as constraints $\rightarrow$ in order to make simulation faster and more efficient
- Using constraints on some angles can help to find out what went wrong in the simulation $\rightarrow$ can lead to an improved force field

How Guide 1UBQ into Its Native Structure?

- Try to find minimum set of constraints which are sufficient
  - Setting temperature dependent constraints $\rightarrow$ decreasing constraints with increasing temperature
  - Setting stronger constraints at the transition temperature
- Constraints make it harder to escape from a local minimum, that can be compensated by increasing the maximum temperature
- Constraining the part that works well

What to Do Next?

- Analysis of the constraint runs needs to be done
- Try to reduce constraints and find smallest necessary set of constraints $\rightarrow$ clue to the force field

Conclusion

- We got excellent results for the folding of the 73 residue designed 3-helix bundle protein, 2A3D, making it the largest protein folded with this model
- Natural proteins, like ubiquitin, constitute a greater challenge. The existing force field was unable to fold the natural proteins tried during the project
- We tested a newly developed method to do simulations with constraints, which will serve, among other things, as an additional aid in force field research

Julie Krainau
Humboldt Universität Berlin