# Constraints in Protein Folding

**Professor Steffen Petersen**

The 3D structure of a protein is intimately linked with its function. If a protein loses its 3D structure, it is likely to abolish its function. If the 3D structure is modified, the protein is expected to modulate its activity and specificity. Protein Engineering is the science and art of doing so. Much can be learned from studying experimentally determined protein 3D structures. The Protein Structural Database PDB contain all publically available knowledge about Protein 3D structures determined either by X-ray diffraction, NMR and in a few cases predicted 3D structures have been deposited as well.

The Holy Grail in structural biology is to be able to predict the 3D structure of a protein based on sequence information. Whereas this may not be achieved fully for many years to come, intelligent use of all available information is the most sensible approach. Presently there are good methods for secondary structure prediction from sequence information, but only rudimentary knowledge about what drives the packing of such secondary structural elements into the folded 3D structure.

In 2001 Per Harald Jonson and Steffen B. Petersen published the paper "A critical view on conservative mutations" (Protein Engineering vol. 14. 397- 403, 2001), which uncovered some hitherto unknown spatial correlations of amino acids in proteins. The protein was perceived as divided into 10 layers based on solvent accessibility of such layers. For each layer spatial neighbor correlations were computed and the results could be viewed as statistical maps in 2D and 3D. We therefore have access to a unique database about packing of amino acids in protein structures. A third semester project succeeded in building a MATLAB based program, ExPro, which allows the user to investigate the amino acid spatial correlations graphically. The program thus reads the database and uses the information for graphical representation.

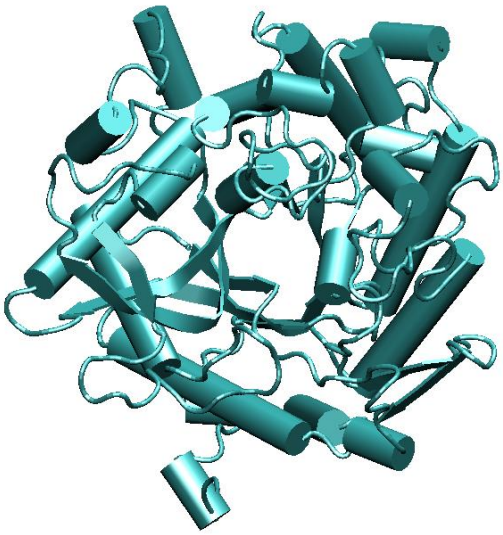
In the present project we will select a few smaller proteins, such as BPTI (Bovine Pancreatic Trypsin Inhibitor), for which the 3D structure has been experimentally determined. We will target the simulated folding of such proteins using a combination of molecular dynamics and the statistical information present in the database.

The present approach to simulating protein dynamics is to assume that the potential energy of a protein can be divided into two terms: bonded and non-bonded. In the bonded class one simulates bond-stretch, angle-bend and rotate-along-bond motions. In the non-bonded class one includes van-der-Waals energy, and electrostatic energy.

$$V(R) = E_{bonded} + E_{non-bonded}$$

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{rotate-along-bond}$$



In the project we will add a term to the non-bonded potential energy function – based on our statistical insight into existing protein structures.

The new statistical term can be viewed as a constraint, or tethering force, i.e. when the MD program attempts to redefine the #D structure of the protein, it will have to include a new potential that constantly drives the protein to minimize this term. There will be 400 such terms, one for each pair of amino acids. The essential potential values, are those we depict using ExPro.

It would make sense if we could interface am existing protein MD program to a MATLAB command module.