

Protein Folding in Silico

With the recent successful completion of the Human Genome Project and related attempts to determine whole genomes it has become obvious that the obtained wealth of data needs to be matched by information on the function and interaction of the huge number of encoded proteins. These polymers are the workhorses in a cell and are responsible for transporting molecules, catalyzing biochemical reactions, or fighting infections.

Proteins are only functional if they assume specific shapes. Despite decades of research it is still an open question how these structures emerge from a protein's chemical composition (the sequence of amino acids as specified in the genome). An answer to this question could lead to a deeper understanding of various diseases that are caused by the miss-folding of proteins, and enable the design of novel drugs with customized properties.

Given a sufficiently accurate description of the forces between the atoms in a protein, and between a protein and the surrounding environment, it is theoretically possible to simulate the folding of a protein. However, the complex type of the interactions containing both repulsive and attractive terms leads to very rough energy landscapes. Hence, sampling of low-energy conformations becomes a hard computational task.

Only with the use of massively parallel computers (such as JUMP and the new BlueGene in Jülich) and the development of advanced simulation techniques are we approaching a point where atomistic simulations of stable domains in proteins (usually of order 50-200 residues)

become feasible. Sampling in generalized ensembles, parallel tempering and energy landscape paving are some of the novel algorithms that allow exploring low-energy configurations without the simulations getting trapped in a local minimum.

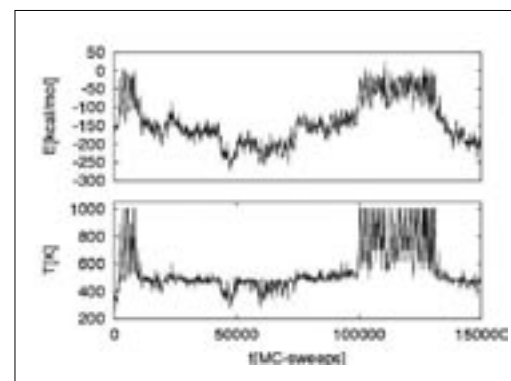


Figure 1: Time series of energy and temperature of one replica in a parallel tempering simulation of the 36-residue protein HP-36

Especially interesting for parallel computing is parallel tempering where one distributes on the nodes of a multiprocessor machine N copies of the molecule, each at a different temperature. In addition to standard Monte Carlo or molecular dynamics moves on each copy, parallel tempering allows with certain probability the exchange of conformations between two copies i and $i+1$. I show as an example in Figure 1 the time series of temperature and energy of one arbitrarily chosen replica as obtained in a parallel tempering simulation of the 36-residue protein HP-36 (the figure is taken from [1]). Note how the resulting random walk in temperature leads to one in energy that enables escapes out of local minima. In this way sampling of low-energy structures will be enhanced. A simple implementation of this and other modern protein simulation techniques can be found in the free program package SMMP (Simple Molecular

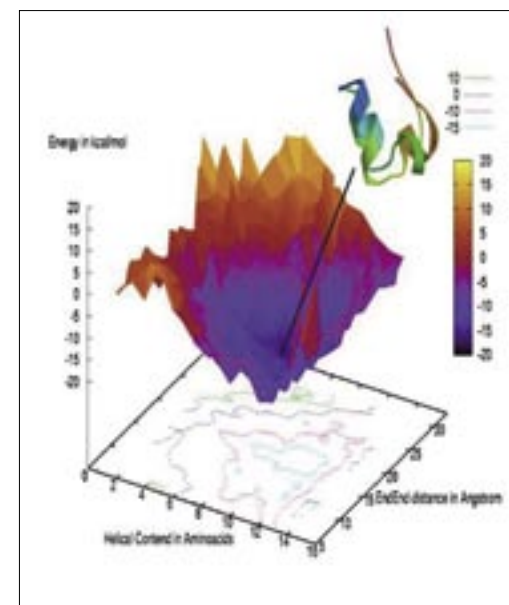


Figure 2: Energy landscape of the trp-cage protein

Mechanics for Proteins) which is available from the web page [2].

Current applications focus on probing the mechanism of folding in small proteins and the conditions under which proteins mis-fold and aggregate. A now widely accepted assumption is that a protein evolves into its biologically active structure through a diffusive process along a funnel-shaped energy landscape. Figure 2 shows as an example a two-dimensional projection of the folding funnel of the 20-residue trp-cage protein as determined in a computer

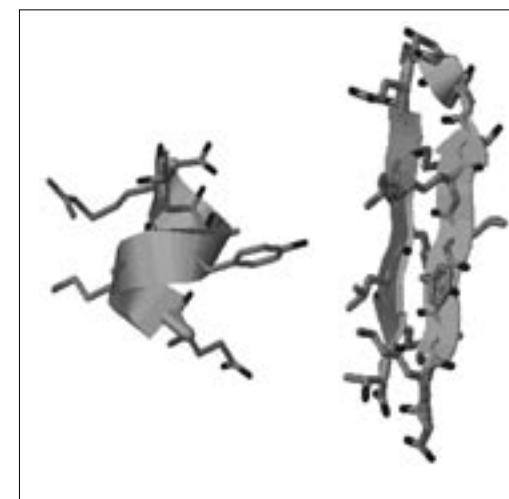


Figure 3a and 3b: Low-energy configurations of the chameleon peptide EKALYRT

simulation. Configurations found at the bottom of the funnel resemble closely the experimentally determined structure (the figure is taken from [3]).

Particularly interesting and important are situations where proteins fold incorrectly. Formation of a β -strand instead of an α -helix in part of a protein can lead to aggregation and formation of fibrils that are often related to the outbreak of neurological diseases. A possible mechanism for the growth of the toxic fibrils may be that the incorrectly folded protein induces mis-folding in close-by molecules. For instance, the peptide EKAYLRT likes to form a β -strand when in the vicinity of another β -strand (Figure 3b), while further away (or isolated) it tends to form an α -helix (Figure 3a). The figure is taken from [4].

One challenge in the next years will be to extend these lines of research to larger and medically relevant proteins. Other research will focus on the interaction of proteins with different biological molecules (flexible docking) in order to understand how biomolecules interact and regulate each other in a cell. Applications of the current research may also include the use of proteins for assembling nanostructures.

References

- [1] C.Y. Lin, C.-K. Hu, U.H.E. Hansmann
Proteins 52 (2003) 436
- [2] www.phy.mtu.edu/biophys/smmp.htm
- [3] A. Schug, W. Wenzel, U.H.E. Hansmann
J. Chem. Phys., 122 (2005) 194711
- [4] Y. Peng, U.H.E. Hansmann
Phys. Rev. E, 68 (2003) 041911

• Ulrich H.E. Hansmann

John von Neumann
Institute for Computing
Research Group
Computational Biology
and Biophysics
Research Centre Jülich