

Modular Multiscale Approach to the Protein-Folding Problem

**Nevena Ilieva¹, Jiaojiao Liu², Peicho Petkov³,
Jianfeng He², Antti Niemi⁴**

¹Institute of Information and Communication Technologies,
Bulgarian Academy of Sciences, Sofia, Bulgaria

²School of Physics, Beijing Institute of Technology, Beijing, PR China

³Faculty of Physics, St. Kliment Ohridski University of Sofia,
5 J. Bourchier Blvd., 1164 Sofia, Bulgaria

⁴Department of Physics and Astronomy, Uppsala University, Uppsala,
Sweden

Abstract. Protein folding is the process of formation of a functional 3D structure from a random coil. The protein folding problem endures as one of the most important unresolved problems in science. Despite Anfinsen's dogma, which states that the native 3D shape (the tertiary structure) of a protein is completely determined by its primary structure (the amino acid sequence), with presently available physics-based methods it is not yet possible to reliably deduce the shape of a biologically active protein from the corresponding aa sequence.

We monitor the evolution of the protein backbone geometry in terms of a complete set of structural order parameters – the virtual backbone bond and torsion angles in the associated with the C_{α} -atoms Frenet frames. We inquire how structure emerges during the process of protein folding. In the vicinity of a local minimum, the expansion of the free energy that models the thermodynamics of the protein relates to the DNLS energy function. The modular building blocks of a folded protein can be described in terms of the kink solution to this equation. We propose that the spin chain interpretation of a protein backbone provides a systematic framework for understanding and description of the protein folding process, since the formation of a super-secondary structure can be understood in terms of a Bloch domain wall that forms along a Heisenberg spin chain, or along a closely related XY chain.

We aim at developing a next generation, computationally effective precision multi-scale approach to predict the structure and to model the dynamics of proteins from their amino acid sequence. The approach we propose builds on complementarity of stochastic and deterministic steps, substantiated in MCMC (Markov chain Monte Carlo) and MD (molecular dynamics) approaches. The protocol allows for accelerated sampling of the conformation space and faster convergence to the biological fold. It also sheds some light on the existence of exceptions from Anfinsen's dogma.

This research was supported in part by Bulgarian Science Fund (Grant DNTS-CN-01/9/2014) and Intergovernmental S&T Cooperation Project at the Ministry of Science and Technology, P.R. China (2014-3