Computational Biology and Chemistry 35 (2011) 169–173

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Brief communication

Multi-nucleation and vectorial folding pathways of large helix protein

Rui Zhou[a], Yi He[b], Yi Xiao[a,*]

[a] Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA

ARTICLE INFO

Article history:
Received 19 January 2011
Received in revised form 31 March 2011
Accepted 17 April 2011

Keywords:
Helix protein
Internal repeats
United-residue model
Folding pathway

ABSTRACT

At present we have already had the detailed knowledge of the folding of small model proteins, but a unified picture of how large proteins fold is still absent. We simulated the folding of a large eight-helix-bundle protein with a length of 145 amino acids by using a united-residue protein model. We observed a multiple nucleation folding pathway: the formation of secondary structures was followed by the nucleation of helices at the two terminal parts and also at the middle of the chain, and then the nuclei grew and combined with each other to form the tertiary structure. Surprisingly, we also found a vectorial folding pathway that was shown recently for co-translational folding in the ribosome exit tunnel. Furthermore, we found that all three-helix subunits in the chain can fold into native-like conformations independently, especially those at the two terminal parts and the middle of the chain, which may be responsible for the nucleation's. These results may be helpful to understand the folding mechanism of large repeat helical proteins.

1. Introduction

The protein folding mechanism is regarded as a grand challenge of molecular biology (Dill et al., 2008). The folding pathways of large proteins are complex and in the past decades different models have been proposed to elucidate them (Baldwin and Rose, 1999a,b; Daggett and Fersht, 2003; Fersht, 2008; Karplus and Weaver, 1994; Lesk and Rose, 1981; Ptitsyn, 1996, 1995; Wetlaufer, 1973). Recent experiments showed that the complex folding behaviors of proteins are related to the number of nucleation motifs containing in the native topology (Lindberg and Oliveberg, 2007). Meanwhile, due to biological importance and particular architecture, repeat proteins (e.g., ankyrin repeat, leucine-rich repeat) aroused more and more interest (Forrer et al., 2004; Lee and Blaber, 2010; Mosavi et al., 2002). It was shown that the sequence similarity of repetitive units in repeat proteins can result in multiple folding cores and lead to diverse folding pathways (Kloss et al., 2008). Therefore, repeat protein can also be considered as an ideal model to investigate the folding pathways of large proteins.

Molecular dynamics (MD) simulations have been made huge progress in studying protein folding mechanism. One of the most significant works was done by Duan and Kollman in 1998 and they observed the intermediate state of villin headpiece subdomain with 35 residues in 1-microsecond MD simulation in explicit water (Duan and Kollman, 1998). Recently Lei and Duan further revealed the folding landscape by high-accuracy *ab initio* folding(Lei et al., 2007). Yet, it is still difficult to use all-atom MD simulation to study the folding processes of large protein (Chen and Xiao, 2006, 2008; Fersht and Daggett, 2002). However, the united-residue (UNRES) force field provides alternative approach to attack this problem, especially that of helical proteins (Liwo et al., 2007, 2005). Compared to other coarse-grained model which are largely knowledge-base, UNRES is purely physics-based force field, determined by a cluster-cumulant expansion of the effective free energy of a protein plus the surrounding solvent (Liwo et al., 2001). Because of simplified representation (Liwo et al., 2005) UNRES is capable to carry out large-scale simulations in real time (Khalili et al., 2006; Liwo et al., 2010). The UNRES have been successfully applied to study the structures and dynamics of many proteins. For examples, the UNRES 4P force field was used to simulate the folding processes of seven proteins including 4 α proteins, 1 β proteins and 2 α+β proteins with lengths from 28 to 75. All α-helical proteins and a α+β folded to the native-like structures (Liwo et al., 2007, 2005). Carr and Wales (2005) investigated the folding of the villin headpiece subdomain using the UNRES force field. The UNRES F2 force field was used to study the folding pathways of the B-domain of staphylococcal protein A (46 residues helical protein) (Khalili et al., 2006; Maisuradze et al., 2010). Among 400 35 ns trajectories of simulations started from the extended state, 380 of them are folded to the native structure. The results agreed with the composition of the intermediate deduced by experimental data and its short lifetime. Recently, the UNRES force field was also applied to investigate β-amyloid peptide fibrils (Rojas et al., 2010) and PDZ binding to the BAR domain of PICK1 (He et al., 2011).

* Corresponding author.
E-mail address: yxiao@mail.hust.edu.cn (Y. Xiao).

1476-9271/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2011.04.004

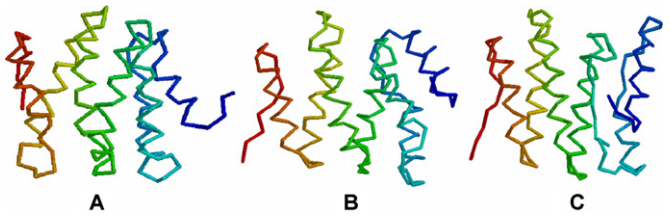

Fig. 1. Experimetal and simulated structures of 1DVP. (A) The experimental structure. (B) The simulated structure with minimal Cα-rmsd 3.35 Å. (C) The most populated conformation with Cα-rmsd 4.17 Å.

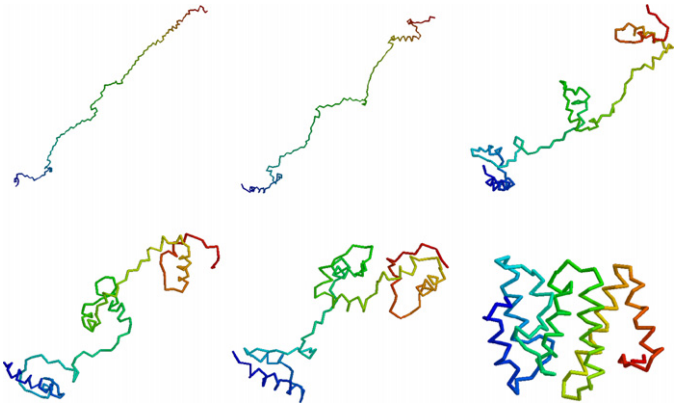

Fig. 2. A typical simulated multi-nucleation folding pathway of 1DVP.

In a previous paper (He et al., 2009c), we reported an UNRES molecular dynamics simulation of the folding processes of a six-helix protein (Harris et al., 2004) and found that the folding started simultaneously from the two ends and formed two native-like compact units and then these two units assembled into the final tertiary structure. We also found the first-half and last-half parts of this protein could fold into their native conformations independently (He et al., 2009b). This implies that this protein could be divided into two foldable halves, which might be responsible for the multiple folding nuclei behaviors. To see the generality of this kind of folding pathway, in this paper we study the folding processes of another large helix-bundle protein of 145 amino acids (PDB:1DVP; Mao et al., 2000) by UNRES molecular dynamics simulation.

2. Methods

We used the 1GAB force field here, which had been optimized by 1GAB as a training protein and proved to be highly predictive for α-helical proteins (Liwo et al., 2007). The details of the UNRES force field and implementation of the Lagrangian and Langevin formalism with UNRES can be found elsewhere (He et al., 2009a; Khalili et al., 2005a,b; Liwo et al., 2005; Oldziej et al., 2005). The canonical MD was employed to simulate the folding processes of 1DVP from extended conformations with the Berendsen thermostat at a constant temperature of 300 K. We did not use any restraints derived from the known structure in our simulations. The time increment for integrating the equation of motion was 4.89fs. 48 independent trajectories were simulated for 293 ns.

3. Results and discussions

1DVP is an eight-helix bundle protein that is believed to play a key role in tyrosine kinase receptor signaling (Lohi and Lehto, 1998) and was proposed as a multipurpose docking adapter that localize proteins to the membrane through interactions with the membrane and the endocytic machinery (Mao et al., 2000).

From the simulations, the protein successfully folded into native conformation with minimal Cα-rmsd (root mean square deviation) of 3.35 Å from experimental structure. The minimal Cα-rmsd structure, the experimental structure, and the most populated structure (with a Cα-rmsd of 4.17 Å to the experimental structure) are shown in Fig. 1. Since 1DVP is a large protein for *ab initio* molecular dynamics simulation, we define the structure whose Cα-rmsd value is lower than 5.50 Å as a native-like structure. If a trajectory reaches the native-like state and continuously stays in the state which has Cα-rmsd value lower than 7 Å for longer than 35 ns, we define the trajectory as a successfully folded one. According to the definition, there are 10 trajectories that folded into native state successfully, and the average mean first passage time, which is defined as the average time of arriving the native-like structure at first time in the 10 folded trajectories, is 53 ns in UNRES time.

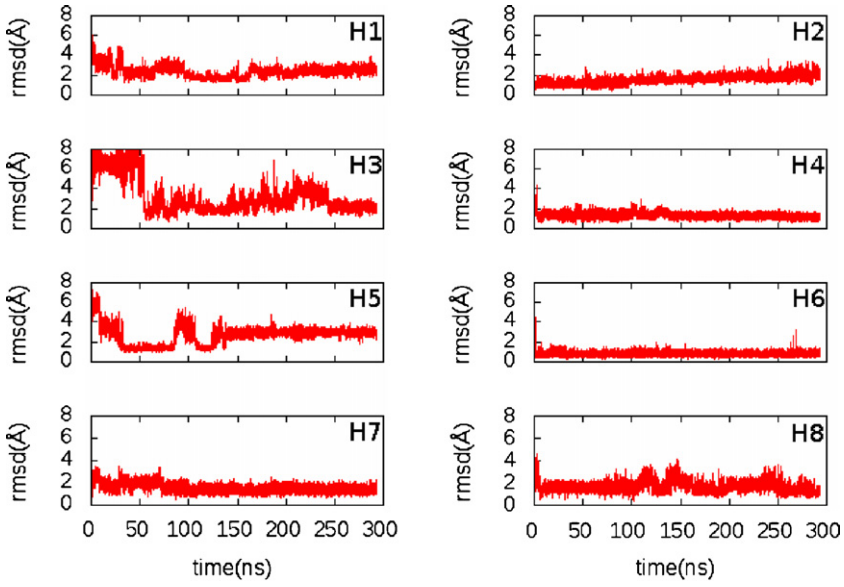

Fig. 3. The Cα-rmsd values versus time of each helix in a typical folding trajectory with minimal Cα-rmsd structure. Helices are numbered from N-terminus and designated as H1 to H8.

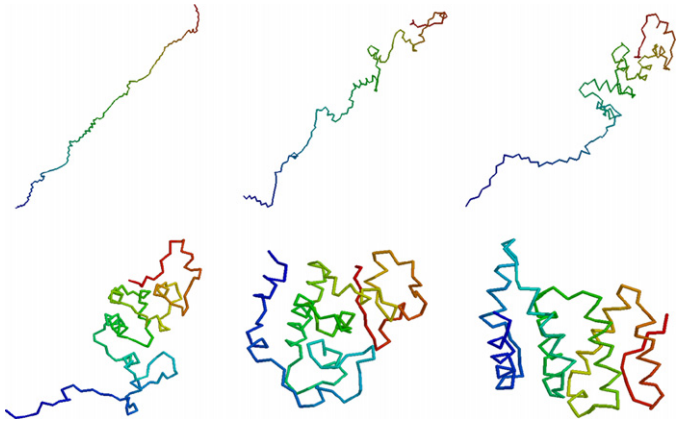

Fig. 4. A vectorial folding pathway of 1DVP.

We observed that nine of ten folded trajectories show very similar folding pathway: the folding starts by the formation of local secondary structures and follows by the simultaneous nucleation of helices at the two ends and the middle part which grow and form three compact units. Then these units are assembled together into a tertiary structure. Finally the whole protein adjusts itself into the native state. Fig. 2 is a typical multi-nucleation folding pathway of 1DVP. To understand the details of folding process, we further analyzed the Cα-rmsd values versus time of each helix of a typical trajectory with minimal Cα-rmsd structure, shown in Fig. 3. It shows that the 2ed, 4th, 6th and 7th helices (H2, H4, H6, and H7) form faster than the others and the Cα-rmsd values of them are below 2 Å at most times. The Cα-rmsd values of the first, third and fifth helices (H1, H3, H5) fluctuate around 3 Å. The results demonstrate that the H2, H4, H6 and H7 are more stable than other parts during folding process. At the early stage of folding, the folding nuclei appear and grow at these places. Our results suggest that multi-nucleation folding pathway may be common for large helix-bundle proteins.

On the other hand, we also find that 1DVP can fold into its native state through a vectorial pathway, i.e., the fully extended chain starts folding by the formation of local helical structures followed by the nucleation of the helices in the C-terminal part and then by vectorial and sequential folding of the remaining part (Fig. 4). This kind of folding pathway was recently found in co-translational folding of an ankyrin repeat protein (a kind of helix-bundle proteins) in the ribosome exit tunnel (Lee et al., 2010). Therefore, it is surprising that this kind of folding pathway can also be realized in *in vitro* folding.

To understand the multi-nucleation at the early stage of the folding of 1DVP, we cut the protein chain into consecutive triple-helix bundles: H123 (H1 to H3), H234 (H2 to H4), H345 (H3 to H5), H456 (H4 to H6), H567 (H5 to H7) and H678 (H6 to H8). We carried out 48 293 ns independent molecular dynamics simulation for each triple-helix bundle from extended states by using UNRES Langevin equations of motion with a time step of 0.1 UNRES time (4.89fs). We found that all the triple-helix bundles can fold into their native-like states but the Cα-rmsd values of H234, H345 and H567 have higher fluctuations than those of H123, H456 and H678, i.e., the later are more stable than the others (Fig. 5). This may explain why 1DVP prefers to form nuclei at the ends and middle parts of the chain.

We also analyzed the internal sequence repeats of 1DVP. In PDBsum, the structure of this protein is also described as a super-helix horseshoe consisted of eight alpha helices. This suggests that the sequence of this protein may have internal repeats. To see whether this is true, we investigated the internal structure-related sequence repetition of 1DVP. There are many methods to detect the internal repetitive units of proteins in sequence and structure levels (Chen et al., 2009; Fischer et al., 1992; Giuliani et al., 2002; He et al., 2009b; Heger and Holm, 2000; Ji et al., 2007; Konopka, 1994, 2003; Konopka and Chatterjee, 1988; Konopka and Smythers, 1987; Rackovsky, 1998; Szklarczyk and Heringa, 2004; Taylor et al., 2002; Turutina et al., 2006; Vriend and Sander, 1991; Xu and Xiao, 2005). By using our previous method (Huang and Xiao, 2007; Xu and Xiao, 2005), we found that all the segments with helical conformations have strong similarity in sequence with each other and can be regarded as repeats (Fig. 6). This may be one of reasons why all helical triplets can fold into the native-like structures independently and why 1DVP folds from multi-nuclei.

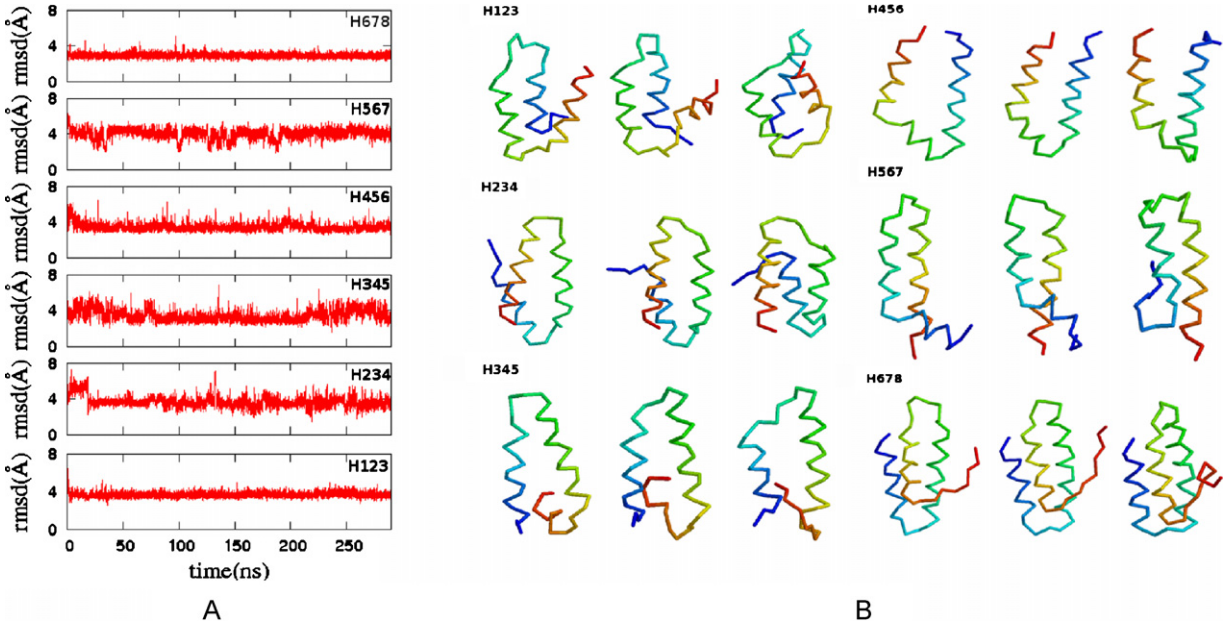

Fig. 5. (A) The rmsd values versus time of typical trajectories of each fragments. (B) Experimetal and simulated structures of 3-helix fragments. They are in the order form left to right as the experimental structures, the simulated structures with minimal Cα-rmsd and the most populated structures.

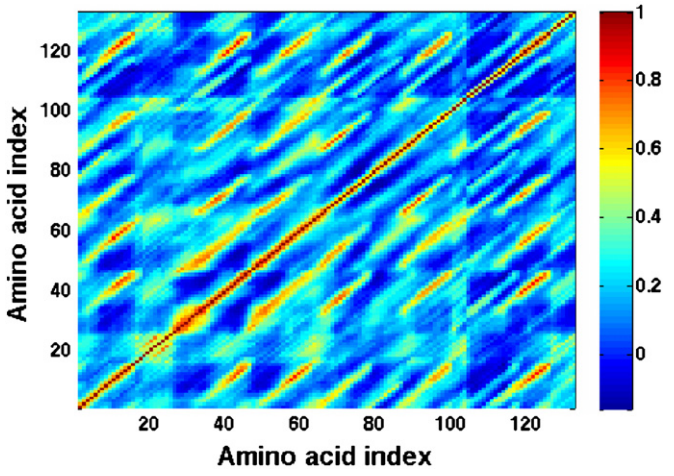

Fig. 6. The internal sequence repeats of 1DVP. The color denotes the values of the Pearson's correlation coefficient (or similarity) between two subsequences with length of 14 amino acids beginning from the amino acids indicated by the amino acid indices of *x*- and *y*-axis.

Although there is no direct experimental evidence that confirms our results at present, the folding pathways found here may not be from the simulation methodology. Firstly, the experimentally determined energy landscapes for helical repeat proteins show that the folding of the repeat protein formed completely from sequentially local helical contacts may simply zip-up from a single core, e.g., the Notch ankyrin domain, or have parallel folding pathways, like the consensus Tetratrico peptide repeat (TPR) protein (Kloss et al., 2008). The repeat proteins may have pathways folding from single core or multi-cores, depending on the sequence identity of the repeats or distribution of the regions of greatest local stability. Our analysis of 1DVP sequence above showed that all the segments with helical conformations have strong similarity in sequence with each other and we also showed H123, H456 and H678 are more stable than H234, H345 and H567. This may be the reason why 1DVP folds from multi-cores. Secondly, our simulated folding processes of three-helix bundle proteins by UNRES force filed are in agreement with experiments and other simulations by using other force fields (He et al., 2009b; Huang et al., 2008; Hubner et al., 2006; Yang et al., 2008). Thirdly, one of the simulated folding pathways, i.e., vectorial folding pathway, was also found in another repeat protein by experiment and simulations, as mentioned above. Finally, as mentioned in Introduction, previous investigations showed that the UNRES force field could give reasonable results for helical proteins.

In summary, we simulated the folding of an eight-helix bundle protein with a length of 145 amino acids. It is found that the folding pathway of 1DVP is very similar to that of 1Q2Z studied in our previous paper and it prefers a multi-nucleation folding pathway. 1DVP may also follow a vectorial folding pathway with nucleation only at one of the ends of the chain. We found that all three-helix subunits in the chain can fold into native-like conformations independently, especially those at the two terminal parts and the middle of the chain, which may be responsible for the nucleation. These results may be helpful to understand the folding mechanism of large repeat helical proteins.

Acknowledgements

We are grateful to Professor Adam Liwo of University of Gdansk in Poland for valuable suggestions and discussions. RZ also thanks Prof. Liwo very much for his hospitality. This work is supported by NSFC under Grant No. 11074084.

References

Baldwin, R.L., Rose, G.D., 1999a. Is protein folding hierarchic? I Local structure and peptide folding. Trends Biochem. Sci. 24, 26–33.
Baldwin, R.L., Rose, G.D., 1999b. Is protein folding hierarchic? II Folding intermediates and transition states. Trends Biochem. Sci. 24, 77–83.
Carr, J.M., Wales, D.J., 2005. Global optimization and folding pathways of selected alpha-helical proteins. J. Chem. Phys. 123, 234901.
Chen, C., Xiao, Y., 2006. Molecular dynamics simulations of folding processes of a beta-hairpin in an implicit solvent. Phys. Biol. 3, 161–171.
Chen, C., Xiao, Y., 2008. Observation of multiple folding pathways of beta-hairpin trpzip2 from independent continuous folding trajectories. Bioinformatics 24, 659–665.
Chen, H., Huang, Y., Xiao, Y., 2009. A simple method of identifying symmetric substructures of proteins. Comput. Biol. Chem. 33, 100–107.
Daggett, V., Fersht, A.R., 2003. Is there a unifying mechanism for protein folding? Trends Biochem. Sci. 28, 18–25.
Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R., 2008. The protein folding problem. Annu. Rev. Biophys. 37, 289–316.
Duan, Y., Kollman, P.A., 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282, 740–744.
Fersht, A.R., 2008. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. Nat. Rev. Mol. Cell Biol. 9, 650–654.
Fersht, A.R., Daggett, V., 2002. Protein folding and unfolding at atomic resolution. Cell 108, 573–582.
Fischer, D., Bachar, O., Nussinov, R., Wolfson, H., 1992. An efficient automated computer vision based technique for detection of three dimensional structural motifs in proteins. J. Biomol. Struct. Dyn. 9, 769–789.
Forrer, P., Binz, H.K., Stumpp, M.T., Pluckthun, A., 2004. Consensus design of repeat proteins. Chem. Biol. Chem. 5, 183–189.
Giuliani, A., Benigni, R., Zbilut, J.P., Webber Jr., C.L., Sirabella, P., Colosimo, A., 2002. Nonlinear signal analysis methods in the elucidation of protein sequence–structure relationships. Chem. Rev. 102, 1471–1492.
Harris, R., Esposito, D., Sankar, A., Maman, J.D., Hinks, J.A., Pearl, L.H., Driscoll, P.C., 2004. The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). J. Mol. Biol. 335, 573–582.
He, Y., Liwo, A., Weinstein, H., Scheraga, H.A., 2011. PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics. J. Mol. Biol. 405, 298–314.
He, Y., Xiao, Y., Liwo, A., Scheraga, H.A., 2009a. Exploring the parameter space of the coarse-grained UNRES force field by random search: selecting a transferable medium-resolution force field. J. Comput. Chem. 30, 2127–2135.
He, Y., Zhou, R., Huang, Y., Xiao, Y., 2009b. Foldable subunits of helix protein. Comput. Biol. Chem. 33, 325–328.
He, Y., Zhou, R., Xiao, Y., 2009c. Study of folding behaviors of a six-helix protein by ab initio molecular dynamics folding simulations of UNRES. Int. J. Mod. Phys. C 20, 373–382.
Heger, A., Holm, L., 2000. Rapid automatic detection and alignment of repeats in protein sequences. Proteins 41, 224–237.
Huang, F., Settanni, G., Fersht, A.R., 2008. Fluorescence resonance energy transfer analysis of the folding pathway of Engrailed Homeodomain. Protein Eng. Des. Sel. 21, 131–146.
Huang, Y., Xiao, Y., 2007. Detection of gene duplication signals of Ig folds from their amino acid sequences. Proteins 68, 267–272.
Hubner, I.A., Deeds, E.J., Shakhnovich, E.I., 2006. Understanding ensemble protein folding at atomic detail. Proc. Natl. Acad. Sci. U.S.A. 103, 17747–17752.
Ji, X., Chen, H., Xiao, Y., 2007. Hidden symmetries in the primary sequences of beta-barrel family. Comput. Biol. Chem. 31, 61–63.
Karplus, M., Weaver, D.L., 1994. Protein folding dynamics: the diffusion–collision model and experimental data. Protein Sci. 3, 650–668.
Khalili, M., Liwo, A., Jagielska, A., Scheraga, H.A., 2005a. Molecular dynamics with the united-residue model of polypeptide chains II. Langevin and Berendsen-Bath dynamics and tests on model alpha-helical systems. J. Phys. Chem. B 109, 13798–13810.
Khalili, M., Liwo, A., Rakowski, F., Grochowski, P., Scheraga, H.A., 2005b. Molecular dynamics with the united-residue model of polypeptide chains I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. J. Phys. Chem. B 109, 13785–13797.
Khalili, M., Liwo, A., Scheraga, H.A., 2006. Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. J. Mol. Biol. 355, 536–547.
Kloss, E., Courtemanche, N., Barrick, D., 2008. Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys. 469, 83–99.
Konopka, A.K., 1994. Sequences and codes: fundamentals of biomolecular cryptology. In: Smith, D. (Ed.), Biocomputing: Informatics and Genome Projects. Academic Press, San Diego, pp. 119–174.
Konopka, A.K., 2003. Sequence complexity and composition. In: Cooper, D.N. (Ed.), Nature Encyclopedia of the Human Genome, vol. 5, pp. 217–224.
Konopka, A.K., Chatterjee, D., 1988. Distance analysis and sequence properties of functional domains in nucleic acids and proteins. Gene Anal. Technol. 5, 87–93.
Konopka, A.K., Smythers, G.W., 1987. DISTAN—a program which detects significant distances between short oligonucleotides. Comput. Appl. Biosci. 3, 193–201.
Lee, J., Blaber, M., 2010. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proc. Natl. Acad. Sci. U.S.A. 108, 126–130.

Lee, W., Zeng, X., Zhou, H.X., Bennett, V., Yang, W., Marszalek, P.E., 2010. Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations. J. Biol. Chem. 285, 38167–38172.
Lei, H., Wu, C., Liu, H., Duan, Y., 2007. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. 104, 4925–4930.
Lesk, A.M., Rose, G.D., 1981. Folding units in globular proteins. Proc. Natl. Acad. Sci. U.S.A. 78, 4304–4308.
Lindberg, M.O., Oliveberg, M., 2007. Malleability of protein folding pathways: a simple reason for complex behaviour. Curr. Opin. Struct. Biol. 17, 21–29.
Liwo, A., Czaplewski, C., Pillardy, J., Scheraga, H.A., 2001. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. J. Chem. Phys. 115, 2323–2347.
Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Oldziej, S., Wachucik, K., Scheraga, H.A., 2007. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. J. Phys. Chem. B 111, 260–285.
Liwo, A., Khalili, M., Scheraga, H.A., 2005. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. Proc. Natl. Acad. Sci. U.S.A. 102, 2362–2367.
Liwo, A., Oldziej, S., Czaplewski, C., Kleinerman, D.S., Blood, P., Scheraga, H.A., 2010. Implementation of molecular dynamics and its extensions with the coarse-grained UNRES force field on massively parallel systems; towards millisecond-scale simulations of protein structure, dynamics, and thermodynamics. J. Chem. Theory Comput. 6, 890–909.
Lohi, O., Lehto, V.P., 1998. VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. FEBS Lett. 440, 255–257.
Maisuradze, G.G., Senet, P., Czaplewski, C., Liwo, A., Scheraga, H.A., 2010. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. J. Phys. Chem. A 114, 4471–4485.
Mao, Y., Nickitenko, A., Duan, X., Lloyd, T.E., Wu, M.N., Bellen, H., Quiocho, F.A., 2000. Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. Cell 100, 447–456.
Mosavi, L.K., Minor Jr., D.L., Peng, Z.Y., 2002. Consensus-derived structural determinants of the ankyrin repeat motif. Proc. Natl. Acad. Sci. U.S.A. 99, 16029–16034.
Oldziej, S., Czaplewski, C., Liwo, A., Chinchio, M., Nanias, M., Vila, J.A., Khalili, M., Arnautova, Y.A., Jagielska, A., Makowski, M., et al., 2005. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. Proc. Natl. Acad. Sci. U.S.A. 102, 7547–7552.
Ptitsyn, O., 1996. How molten is the molten globule? Nat. Struct. Biol. 3, 488–490.
Ptitsyn, O.B., 1995. How the molten globule became. Trends Biochem. Sci. 20, 376–379.
Rackovsky, S., 1998. "Hidden" sequence periodicities and protein architecture. Proc. Natl. Acad. Sci. U.S.A. 95, 8580–8584.
Rojas, A., Liwo, A., Browne, D., Scheraga, H.A., 2010. Mechanism of fiber assembly: treatment of Abeta peptide aggregation with a coarse-grained united-residue force field. J. Mol. Biol. 404, 537–552.
Szklarczyk, R., Heringa, J., 2004. Tracking repeats using significance and transitivity. Bioinformatics 20 (suppl. 1), i311–i317.
Taylor, W.R., Heringa, J., Baud, F., Flores, T.P., 2002. A Fourier analysis of symmetry in protein structure. Protein Eng. 15, 79–89.
Turutina, V.P., Laskin, A.A., Kudryashov, N.A., Skryabin, K.G., Korotkov, E.V., 2006. Identification of amino acid latent periodicity within 94 protein families. J. Comput. Biol. 13, 946–964.
Vriend, G., Sander, C., 1991. Detection of common three-dimensional substructures in proteins. Proteins 11, 52–58.
Wetlaufer, D.B., 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. U.S.A. 70, 697–701.
Xu, R., Xiao, Y., 2005. A common sequence-associated physicochemical feature for proteins of beta-trefoil family. Comput. Biol. Chem. 29, 79–82.
Yang, J.S., Wallin, S., Shakhnovich, E.I., 2008. Universality and diversity of folding mechanics for three-helix bundle proteins. Proc. Natl. Acad. Sci. U.S.A. 105, 895–900.