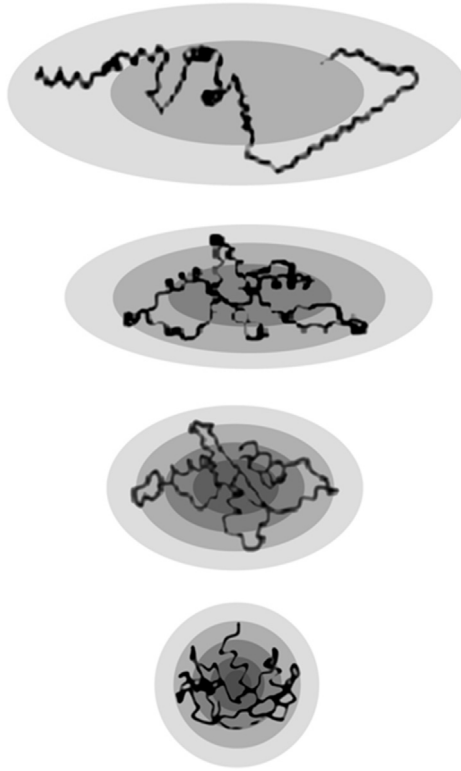
CHAPTER 2

Folding with active participation of water

Dawid Dułak[1], Małgorzata Gadzała[2], Katarzyna Stapor[3], Piotr Fabian[3], Leszek Konieczny[4], Irena Roterman[5]

[1]Department of Biophysics, Faculty of Physics, Astronomy and Applied Computer Science – Jagiellonian University, Kraków, Poland
[2]Cyfronet AGH Academic Computer Centre CYFRONET - University of Science and Technology in Cracow, Kraków, Poland; Currently - Schibsted Tech Polska Sp. z o. o., Kraków, Poland
[3]Department of Theory of Informatics, Institute of Informatics, Silesian University of Technology, Gliwice, Poland
[4]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Krakow, Poland
[5]Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Krakow, Poland

Contents



From Globular Proteins to Amyloids
ISBN: 978-0-08-102981-7
https://doi.org/10.1016/B978-0-08-102981-7.00002-6

The diagram illustrates the basic idea behind protein folding simulations based on the fuzzy oil drop model. The Early Stage (ES) intermediate (top picture) of the polypeptide chain is immersed in an external force field represented by a 3D Gaussian. Optimization (minimization) of nonbonding interactions proceeds in parallel with stepwise alignment between T and O, for each residue separately. Minimizing differences between both distributions directs hydrophobic residues toward the center of the emerging structure, while hydrophilic residues are exposed on its surface. The encapsulating ellipsoid slowly shrinks as the folding process progresses. The resulting increase in packing is reflected by a greater gradient of hydrophobicity between the protein's surface and its center.

Several sample proteins have been subjected to folding simulations based on the proposed model. For the sake of simplicity, we selected proteins whose native conformations are consistent with the 3D Gaussian form — i.e. spherical proteins which well-defined hydrophobic cores [1].

Traditional folding algorithms involve minimization of nonbonding interatomic interactions within the protein along with optimization of the corresponding torsion potentials. Water is typically modeled as a pool of external molecules (mono-, bi- or triatomic models) which interact with amino acids in a pairwise manner. Under these assumptions, producing a conformation which exposes hydrophilic residues is often a time-consuming process, requiring intensive computations.

In the fuzzy oil drop model the external force field is simulated by introducing an additional optimization step which reconciles the placement of each residue with the eponymous "fuzzy oil drop" (mathematically represented by a 3D Gaussian) [2]. This step is interleaved with optimization of the molecule's internal energy and can be implemented using the GROMACS software package [3–7].

The Gaussian is constructed individually for each iteration to match the existing intermediate structure. In successive iterations the capsule shrinks until an acceptable degree of packing has been attained (depending on nonbonding interactions). The final result is then evaluated using CASP similarity metrics, enabling quantitative comparisons between models and their corresponding targets [8].

ES (Early Stage) — the staring structure is generated according to ES model described in details in Refs. [9,10].

Fig. 2.1 visualizes successive steps of internal free energy optimization (labeled E) followed by optimization of hydrophobic forces (labeled H). The corresponding stepwise changes in FOD status are shown in Fig. 2.2,

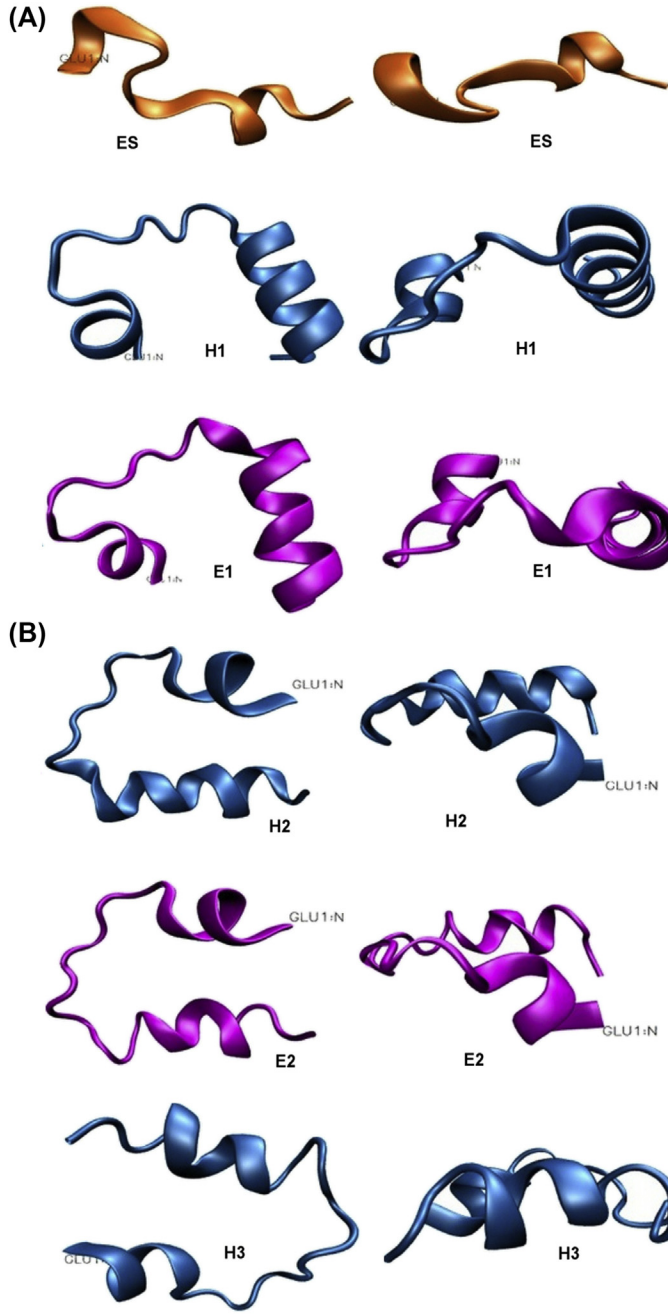


Fig. 2.1 Sample de novo protein (1FME) in successive iterations of the folding process. For each iteration the corresponding internal energy value is listed in Fig. 2.2. (as calculated by GROMACS [3]). Symbols E and H correspond to internal energy and hydrophobicity optimization steps respectively.

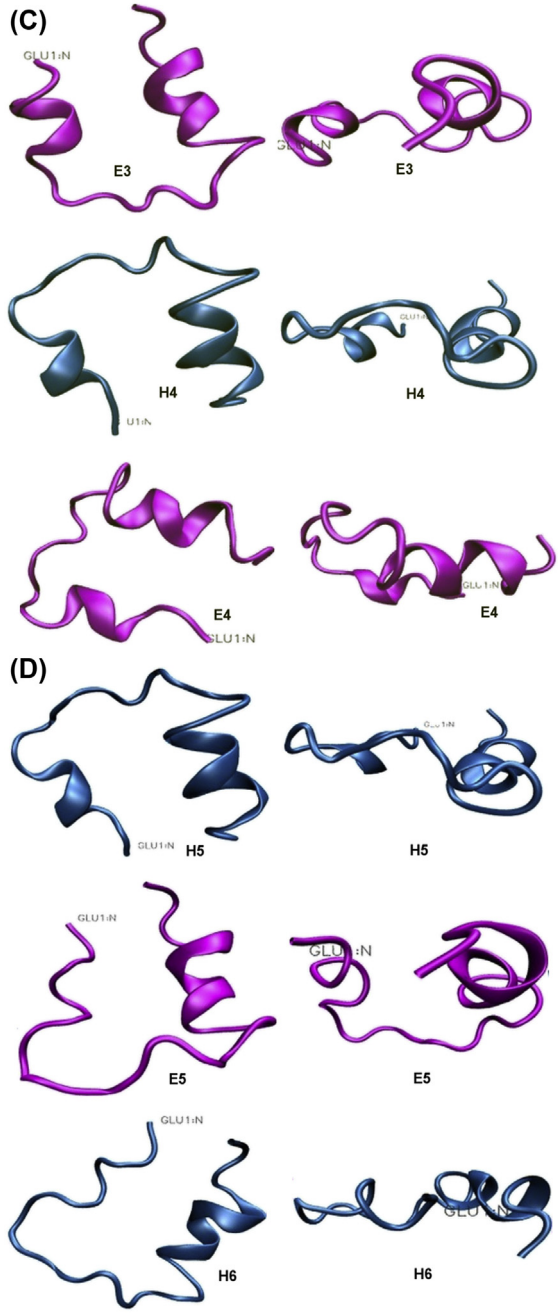


Fig. 2.1 (***continued***).

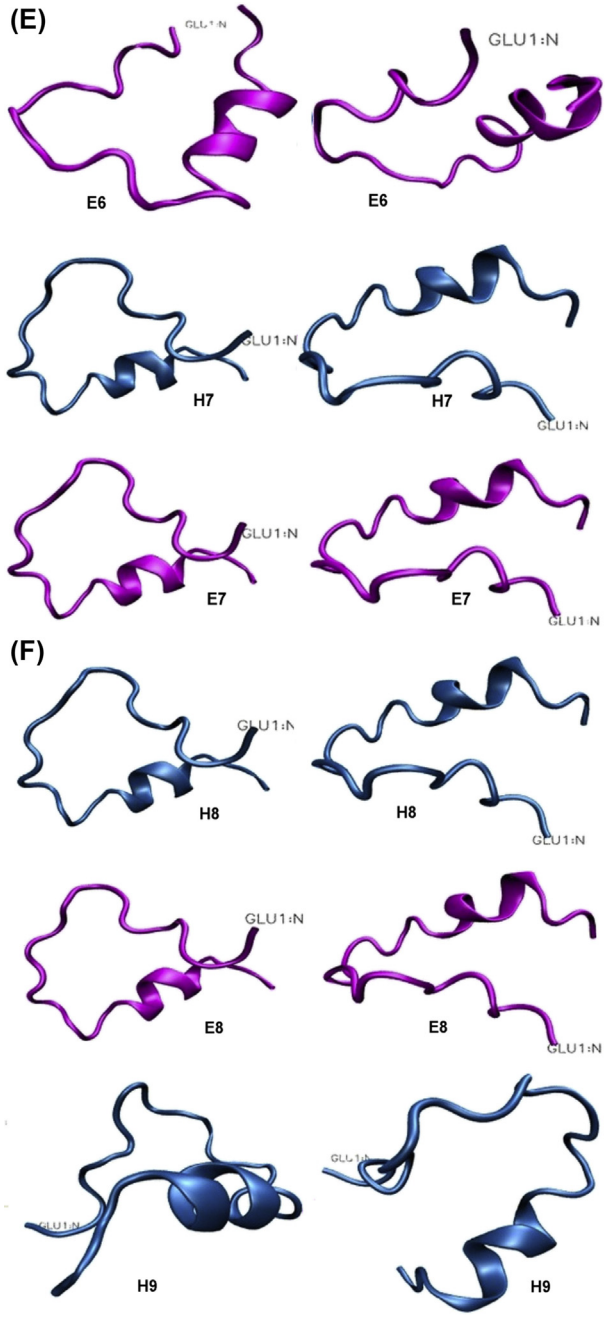


Fig. 2.1 (*continued*).

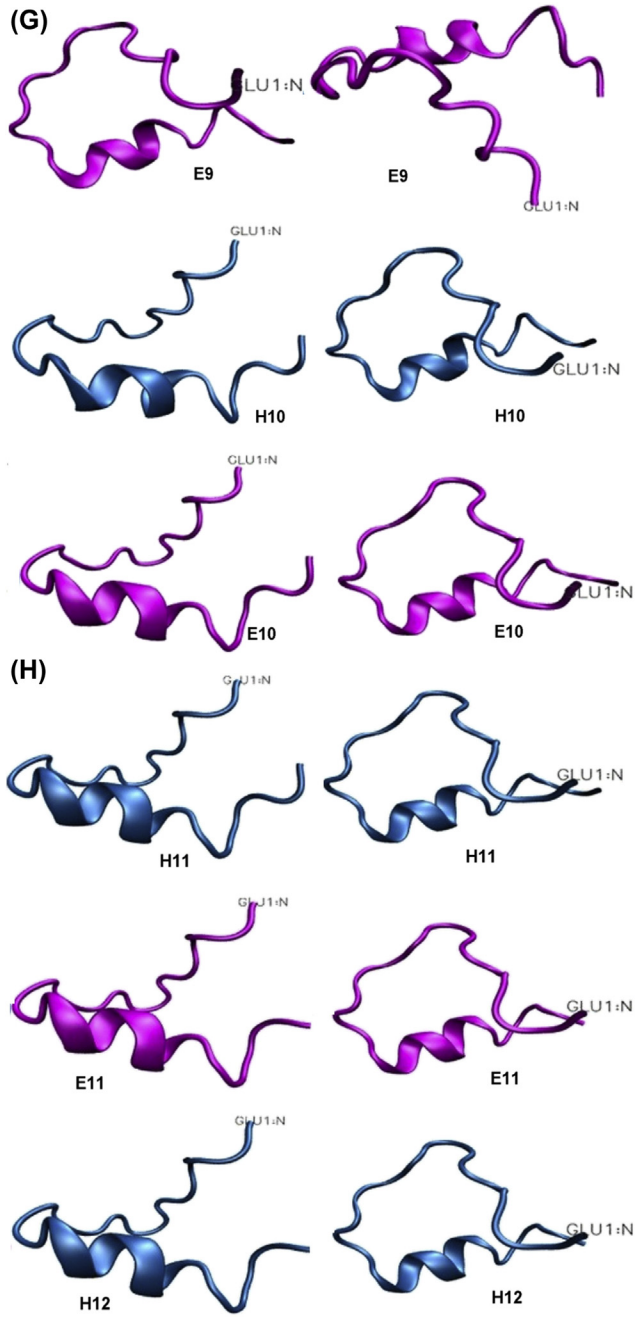


Fig. 2.1 (*continued*).

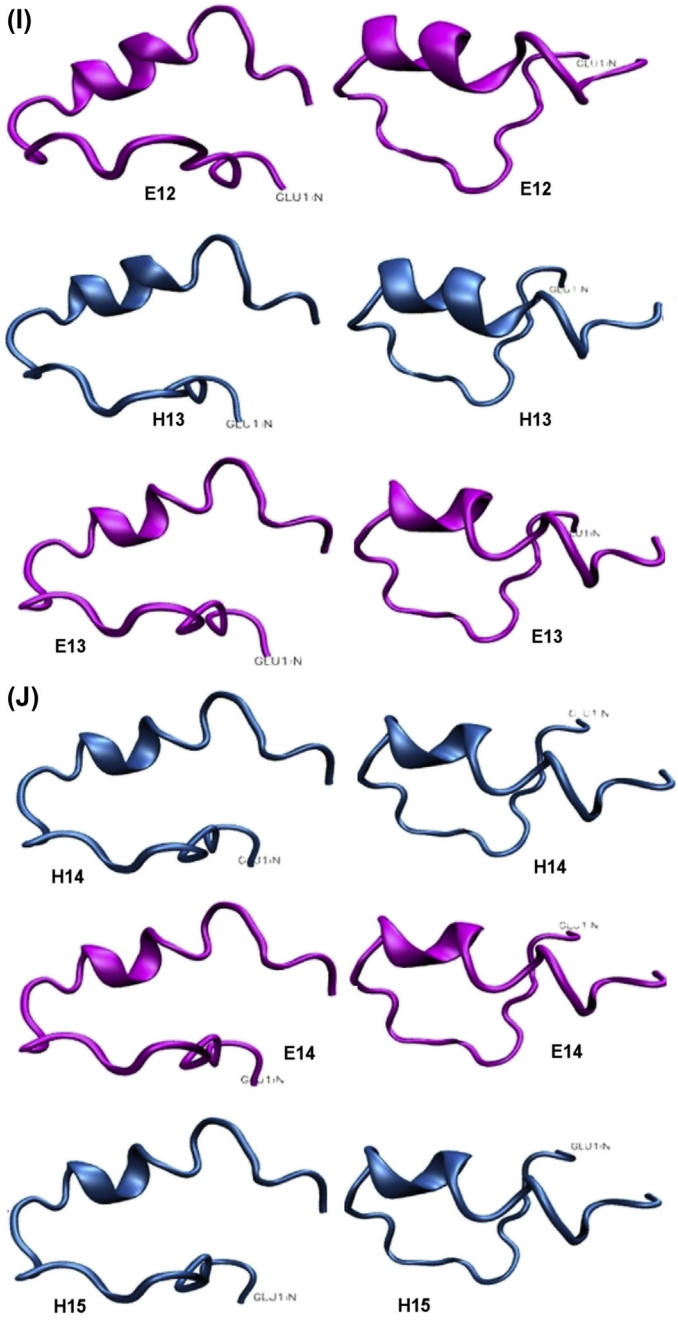


Fig. 2.1 (*continued*).

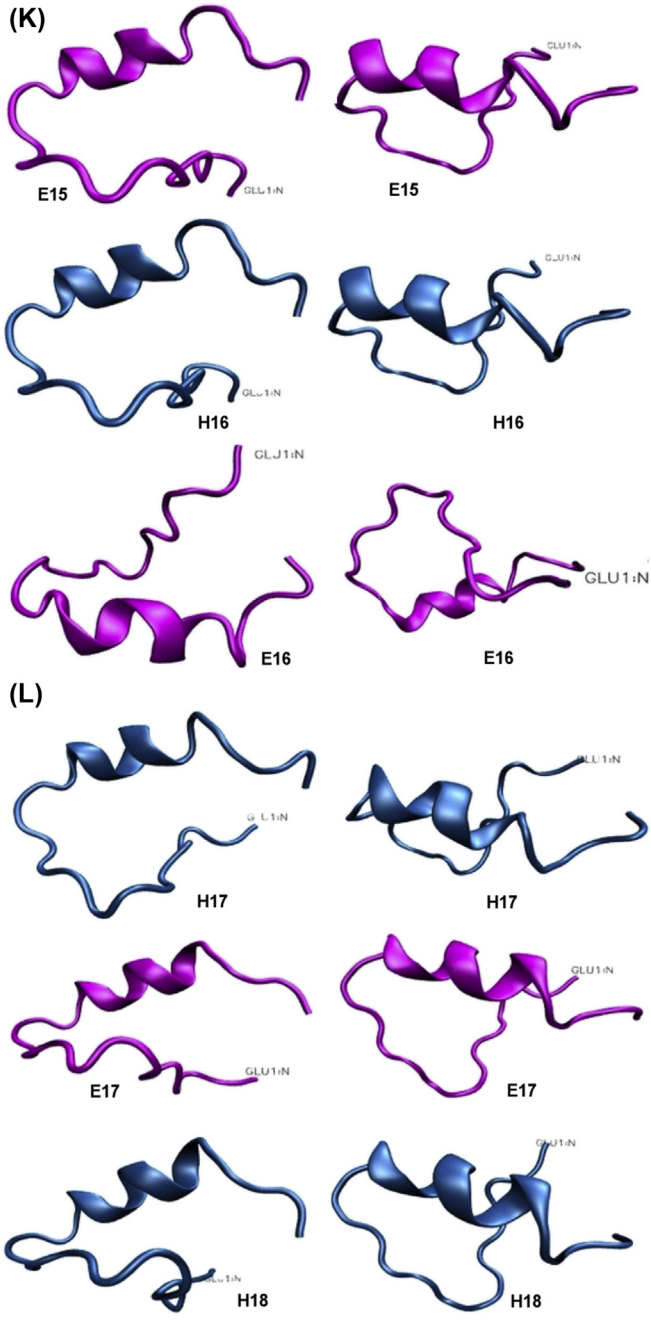


Fig. 2.1 (***continued***).

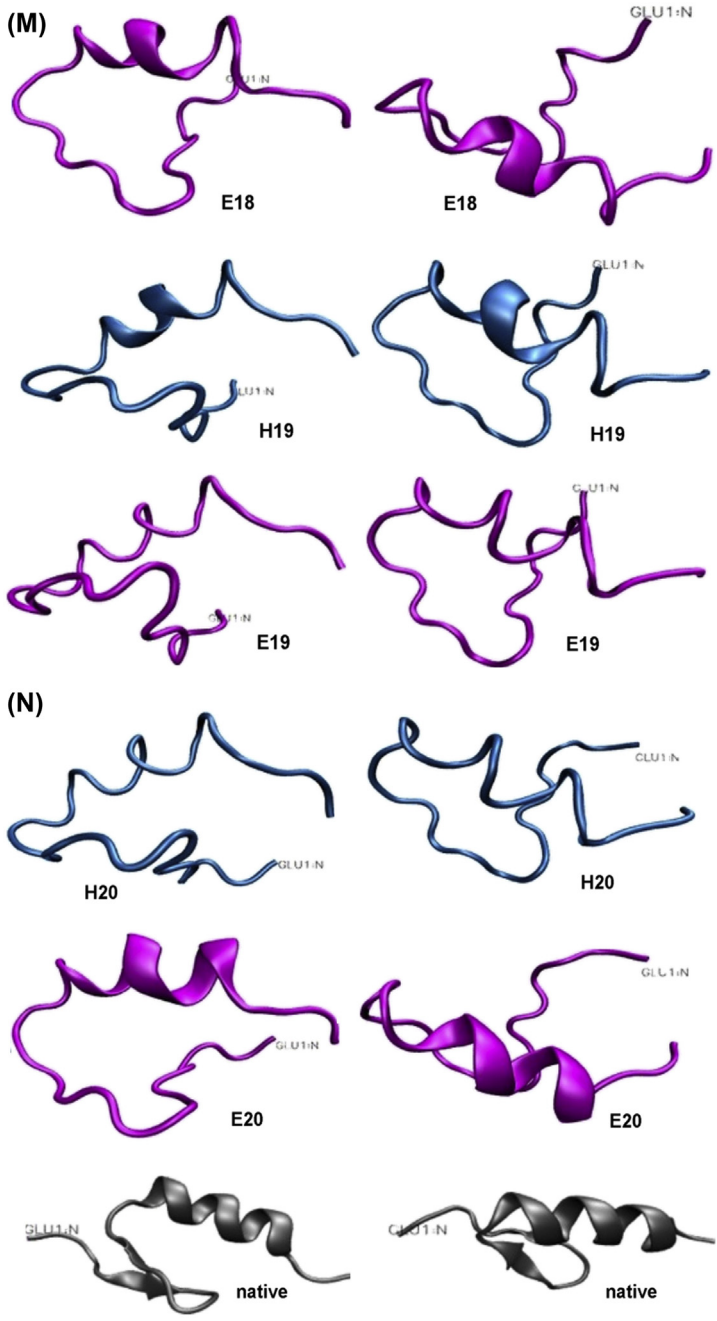


Fig. 2.1 (*continued*).

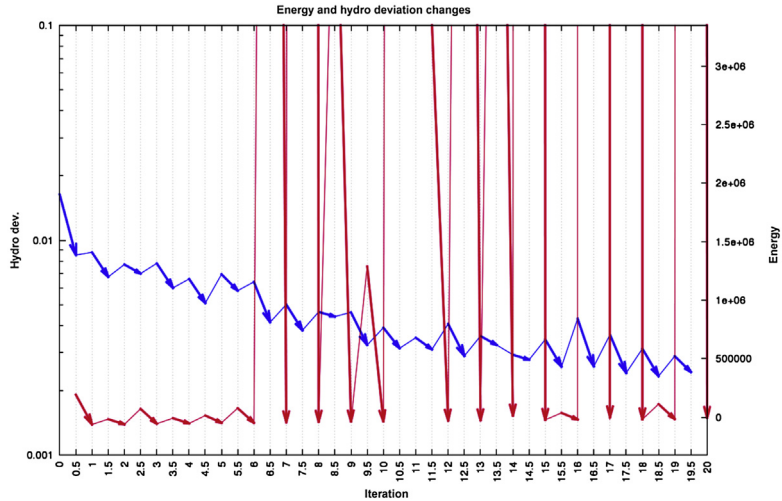


Fig. 2.2 Changes accompanying successive iterations of the folding process. Red line: internal free energy optimization (according to GROMACS); blue line: fuzzy oil drop model fitting (axis called Hydrophobicity). The chart corresponds to structures shown in Fig. 2.1.

while Fig. 2.3 presents the alignment between the resulting model and the target.

The presented example shows that a hydrophobic core emerges as the protein adopts a micelle-like conformation. Other examples of protein folding simulation are also presented in Chapter 10.

The profiles illustrated in Figs. 2.2 and 2.3 reveal a contradiction between optimization of free energy and optimization of hydrophobic interactions. In light of this fact, it seems useful and reasonable to end the process with a final internal energy optimization step. It is due to the fact, that the hydrophobicity optimization is performed on effective atoms. Thus it may introduce the inter-atomic collisions when the structure is transformed to all atom model. Fig. 2.2 reveals progressive formation of a hydrophobic core. It seems that the best alignment is obtained in the 18th iteration, where RMS-D reaches its lowest value, while GDT is almost at its maximum level Fig. 2.3.

The final structures shown in Fig. 2.4 differ somewhat from the target, particularly with regard to their N- and C-terminal fragment. Fig. 2.5 reveals a concentration of hydrophobic residues at the center of the protein, along with exposure of hydrophilic residues on its surface. The target structure, 1FME, is characterized by the following parameters: RD(T-O-R):

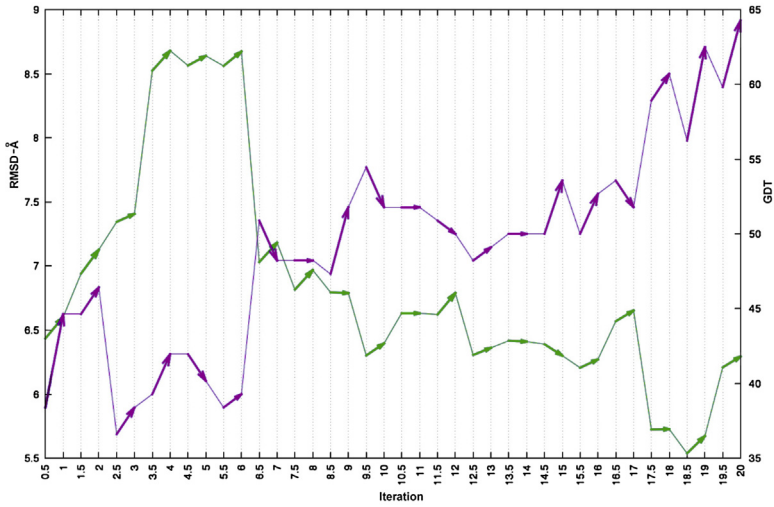


Fig. 2.3 Changes in RMS-D (Å) and GDT (no unit — parameter used in CASP project to assess the similarity between target and model [8]) during folding. The chart corresponds to structures shown in Fig. 2.1.

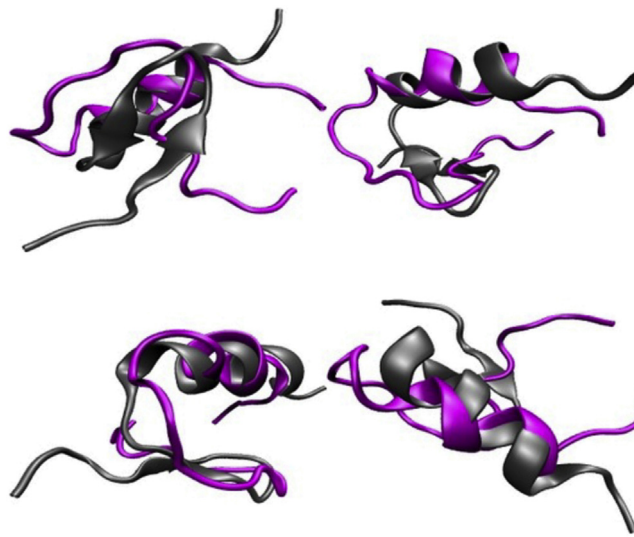


Fig. 2.4 Final models (pink) superimposed onto the native structure of de novo protein (1FME).

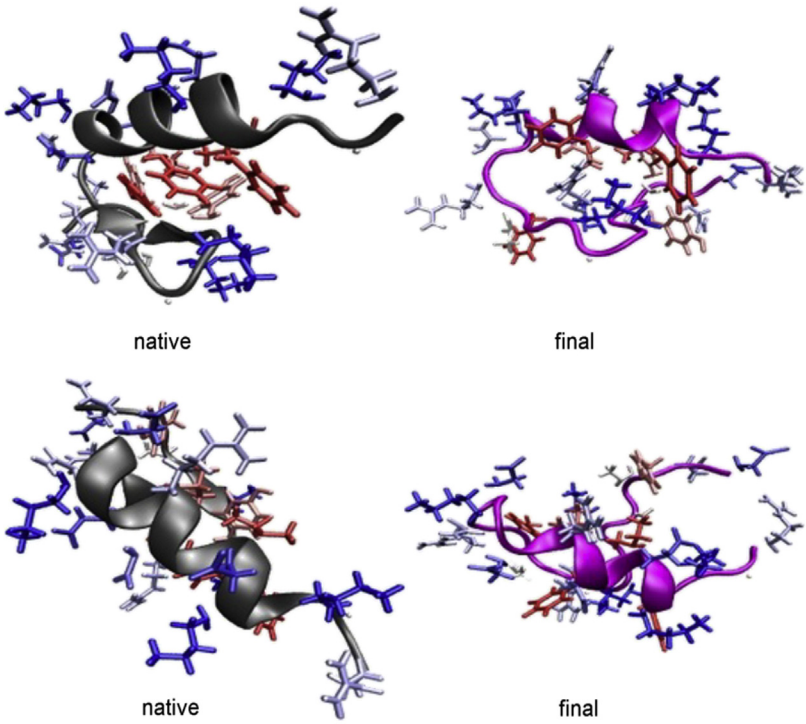


Fig. 2.5 Native and final structures showing hydrophobic (red) and hydrophilic (blue) residues (1FME).

0.298; RD(T-O-H): 0.249; HvT: 0.706; TvO: 0.821; HvO: 0.863. Consequently, this protein is regarded as highly accordant with the theoretical model and represents a useful test case. Our model, generated by the folding simulation, has RD values of 0.294 and 0.089 (T-O-R and T-O-H respectively), along with the following correlation coefficients: HvT: 0.460; TvO: 0.770; HvO: 0.609.

Another sample folding simulation based on the presented model is discussed in Ref. [2].

The authors are currently developing a software package where interleaved optimization of internal (nonbonding interactions) and external (fuzzy oil drop model) force fields is replaced by a multicriteria optimization procedure representing a tradeoff between both factors. Introduction of the Pareto front approach is expected to more accurately reflect the balance between internal and external force fields.

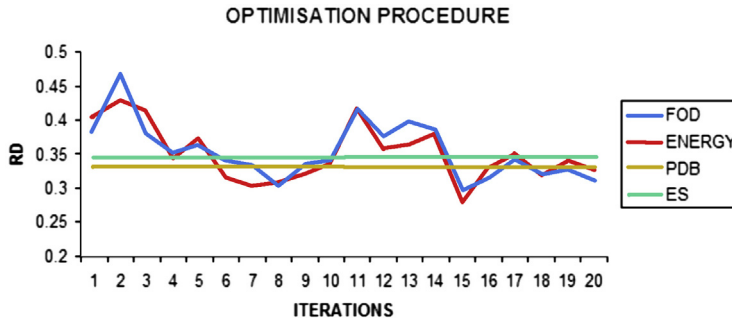


Fig. 2.6 Changes in RD (the measure of idealized − 3D Gauss function distribution) and observed hydrophobicity distribution) during folding. The chart corresponds to structures shown in Fig. 2.1. FOD − status after FOD optimization, Energy − status after energy optimization, PDB − status as it appears in crystal structure, ES − status of early stage model.

To follow the role of FOD influence on folding process the changes of RD are shown in Fig. 2.6.

The energy values represents the internal energy status after the Gromax iteration. The values described as FOD represent the status of the structure after FOD iteration. The convergence can be seen especially in the final steps of procedure. The line described ES represent the status of Early Stage structure generated according to Refs. [9,10]. The line identified as PDB visualize the status of the structure available in PDB. There are 34 models available. The range of RD for all models is as follows: 0.306−0.356. The line represents the averaged RD value for all 34 models treated as reference form.

It would also be highly desirable to validate the model with a broader set of proteins and simulation parameters (e.g. number of iterations).

References

[1] Konieczny L, Brylinski M, Roterman I. Gauss-function-Based model of hydrophobicity density in proteins. In Silico Biology 2006;6(1−2):15−22.

[2] Gadzała M, Dułak D, Kalinowska B, Baster Z, Bryliński M, Konieczny L, et al. The aqueous environment as an active participant in the protein folding process. Journal of Molecular Graphics and Modelling 2019;87:227−39. https://doi.org/10.1016/j.jmgm.2018.12.008.

[3] van Gunsteren WF, Berendsen HJC. Groningen molecular simulation (GROMOS) library manual. 1987. p. 1−221.

[4] van Gunsteren WF, Berendsen HJC. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angewandte Chemie International Edition in English 1990;29(9):992−1023. https://doi.org/10.1002/anie.199009921.

[5] van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kruger P, Mark AE, et al. Biomolecular simulation: the GROMOS96 manual and user guide. Zurich: Verlag der Fachvereine& Berendsen, H. J. C; 1996.
[6] http://www.gromos.net.
[7] van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, et al. Biomolecular modeling: goals, problems, perspectives. Angewandte Chemie International Edition 2006;45(25):4064−92. https://doi.org/10.1002/anie.200502655.
[8] http://predictioncenter.org.
[9] Jurkowski W, Brylinski M, Konieczny L, Wiśniowski Z, Roterman I. Conformational subspace in simulation of early-stage protein folding. Proteins: Structure, Function, and Bioinformatics 2004;55(1):115−27. https://doi.org/10.1002/prot.20002.
[10] Jurkowski W, Baster Z, Dułak D, Roterman-Konieczna I. The early-stage intermediate. Protein Folding in Silico 2012:1−20. https://doi.org/10.1533/9781908818256.1.