CHAPTER 5

Proteins structured as spherical micelles

Mateusz Banach[1], **Leszek Konieczny**[2], **Irena Roterman**[1]
[1]Department of Bioinformatics and Telemedicine, Jagiellonian University—Medical College, Krakow, Poland
[2]Chair of Medical Biochemistry, Jagiellonian University—Medical College, Krakow, Poland

Contents

The image depicts a highly regular, spherically symmetrical structure. Shades of gray correspond to increasing concentrations of hydrophobicity, which is low on the

From Globular Proteins to Amyloids
ISBN: 978-0-08-102981-7
https://doi.org/10.1016/B978-0-08-102981-7.00005-1

surface but high at the center. The only difference between the presented image and the theoretical Gaussian lies in the fact that the Gaussian distribution is continuous rather than discrete.

The model presented in this work, referred to as the fuzzy oil drop model (FOD), is supported by the existence of proteins which conform to it with near-perfect accuracy. Such proteins exhibit excellent agreement between the observed distribution of hydrophobicity (abbreviated as O) and the corresponding theoretical distribution (abbreviated as T) which is expressed as a 3D Gaussian. Accordingly, these proteins may be thought of as spherical micelles.

The image attached to the title of this chapter visualizes a highly ordered structure which exhibits spherical symmetry. Shades of gray correspond to increasing concentrations of hydrophobicity, which is low on the surface but peaks at the center. The only difference between the presented image and the theoretical Gaussian lies in the fact that the Gaussian distribution is continuous rather than discrete.

Proteins rated as highly consistent with the theoretical model include type III antifreeze proteins, muscle tissue proteins such as titin, as well as certain fast-folding proteins. In all these cases tertiary conformations are stabilized by prominent hydrophobic cores rather than by other factors (i.e. disulfide bonds).

With regard to type III antifreeze proteins, their conformance to the fuzzy oil drop model is associated with exposure of polar residues on the surface. This has an effect on the structural ordering of water molecules, however the resulting structuralization of the solvent differs from conditions encountered in ice crystals. Consequently, type III antifreeze proteins lower the freezing point of water simply by virtue of being present.

Muscle tissue is subject to frequent stretching and deformations caused by external forces. In such cases it is desirable for its constituent proteins to revert to their original form when no external forces are present. A well-ordered hydrophobic core provides this useful reversibility, which may be observed i.e. in titin.

Finally, the group of strongly accordant structures includes proteins referred to as "fast-folding" (or even "ultrafast-folding"). In their case, the missing information required at the folding stage is clearly provided by the aqueous solvent.

Type III antifreeze proteins

This group is represented by an antifreeze protein isolated from ocean pout, *Macrozoarces americanus*, listed in PDB under ID 1AME [1]. It consists of 66 residues arranged into four helices and four β-strands forming two anti-parallel strands. In addition, it also contains short disordered fragments.

Fig. 5.1 illustrates the theoretical and observed distribution of hydrophobicity in antifreeze protein type III (1AME).

For this protein, RD values for T-O-R and T-O-H are equal to 0.300 and 0.294 respectively. Correlation coefficients are as follows: HvT = 0.370; TvO = 0.796; HvO = 0.707. These figures indicate the presence of a prominent hydrophobic core (RD < 0.5), and the protein is therefore classified as accordant with the model. In addition, high values of both TvO and HvO suggest that the protein's tendency to form a near-perfect spherical micelle is consistent with the intrinsic properties of its individual residues (notwithstanding limitations described in the introduction, where micellar protein structures are compared to surfactant

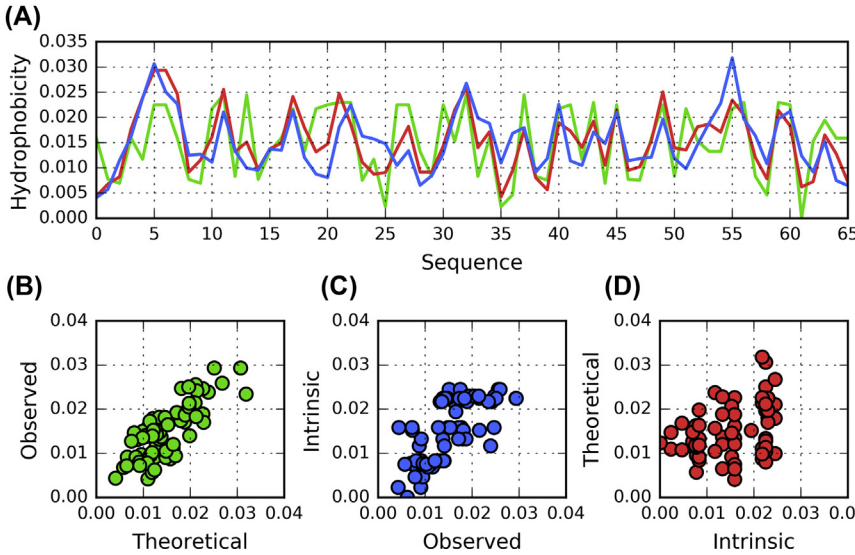

Fig. 5.1 FOD characteristic of antifreeze protein type III (1AME): (A) hydrophobicity distribution profiles: T (blue (dark gray in print version)), O (red (gray in print version)) and H (green (light gray in print version)), (B), (C), (D) correlation scatter plots (TvO, HvO, and HvT respectively). Each circle in (B), (C) and (D) represents one residue in the sequence.

micelles). Note that a low value of HvT would suggest that the theoretical and intrinsic distributions contradict each other. The presented protein serves as proof that, in spite of problems caused by covalent bonds between residues, as well as variable hydrophobicity of the residues themselves, it is nevertheless possible for the protein to achieve a fold which closely corresponds to the spherical model. Good agreement between T and O indicates that the protein's surface consists of polar residues while hydrophobic residues congregate in its central part.

Fig. 5.2 provides view of structure of good accordance with the fuzzy oil drop model. The biological role of this protein is to induce a specific type of order in the surrounding water (though exposure of polarity), hindering the formation of ice crystals. The antifreeze properties of 1AME are thus explained.

Later on we will discuss other possible ways of counteracting crystallization of water (see chapter 7).

Titin

Titin is primarily found in muscle tissue. It is a giant protein, consisting of a sequence of β-sandwich domains. Its overall length may reach 1 μm [2]. Titin works by mimicking a coil spring: it deforms under the influence of external forces, and reverts to its original shape in the absence of such forces.

The immunoglobulin-like domain found in titin differs from immunoglobulins in that it lacks disulfide bonds. Under such conditions the only factor capable of stabilizing the protein's tertiary conformation is a well-ordered hydrophobic core. The core enables the domain to revert to its original structure when no deforming forces are present.

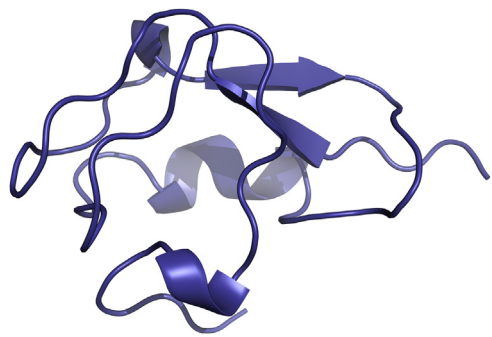

Fig. 5.2 3D presentation of antifreeze protein type III (1AME).

Since the structure of titin (PDB ID: 1TIT) was thoroughly described in another publication [3], here we will focus on the so-called human titin-obscurin-1 complex (PDB ID: 2WWM) [4], which consists of two titin domains and two obscurin-like domains (referred here as obscurin for brevity). Published research underscores the medical importance of this complex since point mutations affecting its sequence may results in hereditary muscle diseases [5].

From the point of view of our current subject it is important that each domain participating in the complex retains a highly ordered hydrophobic core.

As noted, the 2WWM complex consists of four chains labeled O, C (obscurin domains) and T, D (titin domains) respectively. Both proteins are described as "mainly β-sandwich" (2.60.40.10) in the CATH classification. Their length is also comparable: 100 aa for titin and 98 aa for the obscurin domain.

Due to the chain-like arrangement of T/O and D/C dimers which are separated by sparsely packed areas, there is no reason to expect that the four-chain complex as a whole will contain a well-ordered hydrophobic core. Our analysis therefore focuses on individual domains. The status of each chain from titin and obscurin is listed in Table 5.1 and Fig. 5.3. Low RD values for both T-O-R and T-O-H) indicate that the observed distribution of hydrophobicity is, in each case, closely aligned with the theoretical distribution and therefore the domain contains a prominent hydrophobic core. Correlation coefficients appear balanced, with no distinctive outliers. This suggests synergistic self-ordering of residues in the process of shaping the core of each domain.

Table 5.1 Comparison of fuzzy oil drop parameters, showing that each domain contains a prominent hydrophobic core but the complex as a whole does not contain a shared core as observed in titin-obscurin (2WWM).

	RD		CC		
Chain	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
Complex	0.768				
Chain C	0.442	0.458	0.729	0.572	0.437
Chain D	0.391	0.345	0.440	0.671	0.755
Chain O	0.464	0.492	0.413	0.530	0.758
Chain T	0.395	0.359	0.479	0.644	0.793

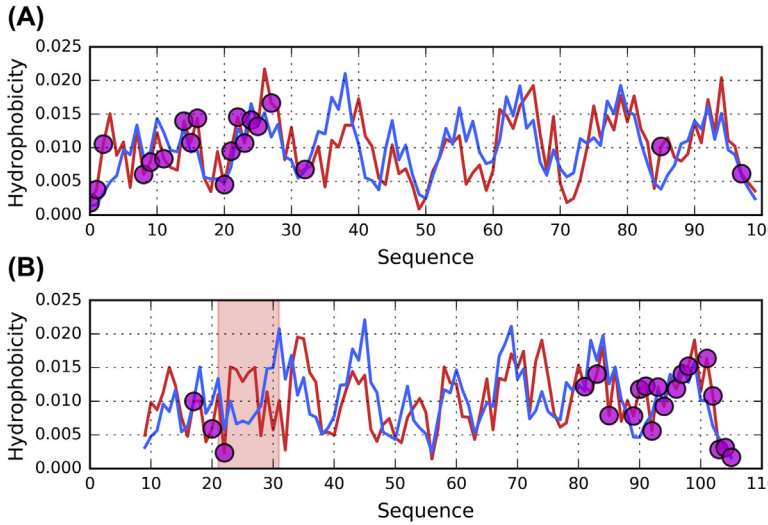

Fig. 5.3 Theoretical (T, blue (dark gray in print version)) and observed (O, red (gray in print version)) hydrophobicity distribution profiles for titin-obscurin complex (2WWM): (A) titin (chain D). (B) obscurin (chain O).

Magenta circles denote residues engaged in protein-protein interactions between the chains. The sole local discordance between the hydrophobicity distributions in obscurin (highlighted in red) does not cause the entire domain to be regarded as discordant with the FOD model.

Hydrophobicity distribution profiles (Fig. 5.3) also reveal good agreement between T and O, suggesting the presence of a strong, central hydrophobic core. In effect, both domains should be regarded as highly accordant with the model and structurally stabilized by their respective hydrophobic cores (Fig. 5.4.).

Such strong accordance with the monocentric structural pattern is also related to the protein's biological purpose. As a component of muscle tissue, it is frequently subjected to external forces and must revert to its original conformation in their absence. Given that no disulfide bonds are present, tertiary conformational stability is mediated solely by the hydrophobic core, and such core must be prominent enough to exert a powerful effect upon the protein's structure. An additional argument which supports our hypothesis regarding the purpose of titin's well-structured hydrophobic core relates to the use of this molecule in protein unfolding analyzes [6]. Studies which focus on (possibly multistage) unfolding often measure the elongation of a protein as a function of the applied longitudinal force. In

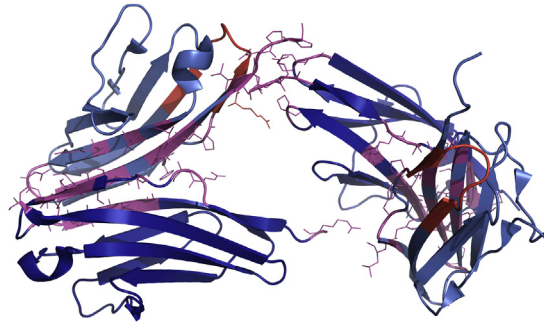

Fig. 5.4 3D presentation of titin (blue (dark gray in print version)) and obscurin (marine (gray in print version)) domains in titin-obscurin complex (2WWM). The fragments marked in each chain of obscurin in red (21–31) are locally discordant. Residues engaged in protein-protein interactions are magenta-colored and have side chains displayed.

order to achieve a macroscopic effect, titin is often used as a convenient scaffold for the target protein. Notably, titin itself undergoes single-step unfolding, thus any elongation becomes easy to detect and identify. This phenomenon appears to relate to the properties of titin's hydrophobic core which undergoes rapid destruction when the applied force exceeds a certain threshold.

An in-depth comparative analysis of β-sandwich domains, particularly in the context of immunoglobulins and their constituent parts, can be found [3]. In this chapter we focus on identification of prominent hydrophobic cores in muscle tissue proteins.

Ultrafast-folding proteins

This group of proteins is represented by mutant chicken villin subdomain hp-35 (PDB ID: 2F4K) with two point mutations: K65L and K70L. Swapping lysine for leucine at both positions produces a protein referred to as fast-folding, or even ultrafast-folding [7]. Given the strong polarity of K and the equally strong hydrophobicity of L, this change (as seen in Fig. 5.5) greatly strengthens the protein's hydrophobic core. The corresponding RD values are: 0.319 (T-O-R) and 0.204 (T-O-H), with correlation coefficient 0.468, 0.764 and 0.735 (HvT, TvO and HvO respectively). Notably, these are among the lowest RD values we have ever encountered throughout our years of work with the fuzzy oil drop model.

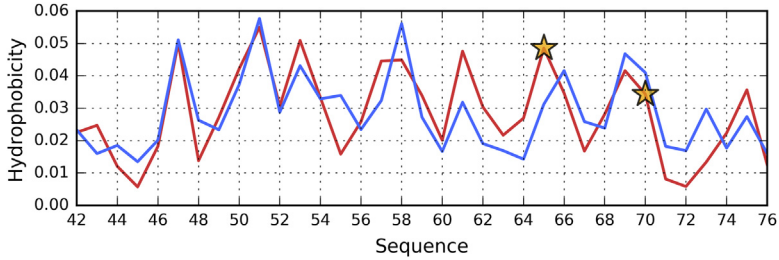

Fig. 5.5 Theoretical (T, blue (dark gray in print version)) and observed (O, red (gray in print version)) hydrophobicity distribution profiles for Chicken villin subdomain (2F4K). Orange stars mark the mutation loci (K65L and K70L).

Taken together, the above values indicate the presence of a particularly prominent monocentric hydrophobic core. Excellent alignment between O and T (regardless of which other reference distribution—R or H—is selected) and balanced correlation coefficients with notable domination of TvO confirm strong micellar properties of this molecule. This observation is further supported by plotting theoretical and observed hydrophobicity distributions, as shown in Fig. 5.5.

Visual inspection of Fig. 5.5 leads to unequivocal conclusion—both distributions are very similar. The role of both previously mentioned mutations is also visualized: introduction of two hydrophobic residues (L) in place of hydrophilic residues (K) at positions where the theoretical model expects hydrophobicity to remain high (in the center of the molecule, as shown in Fig. 5.6) promotes the formation of a hydrophobic core and may explain the extremely rapid ("ultrafast") nature of this process.

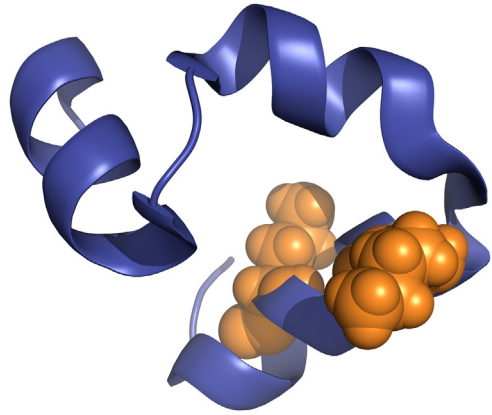

Fig. 5.6 3D presentation of chicken villin subdomain (2F4K). Orange spheres (gray in print version) mark the mutation loci (K65L and K70L).

It might be interesting to consider the presented case in a broader context. Such considerations would touch upon the status of the aqueous solvent which appears to play an active role in the folding process, guiding its course and determining the formation of a hydrophobic core—all in line with the basic assumptions of the fuzzy oil drop model. In Ref. [7] the authors state the following: "The ultrafast folding rate, very accurate X-ray structure, and small size make this engineered villin subdomain an ideal system for simulation by atomistic molecular dynamics with explicit solvent" (excerpt from the abstract). Clearly, the presented protein represents a convenient study subject, enabling validation of the model itself.

Screening analysis of a comprehensive non-redundant set of proteins [8] confirms that the vast majority of domains folds in accordance with the fuzzy oil drop model. This serves as further proof of the model's correctness and applicability in simulating various secondary and supersecondary folds [9].

Regarding immunoglobulin-like domains, despite strong topological similarities their constituent β-strands tend to exhibit variable accordance with the model [3].

The presented protein chicken villin subdomain (2F4K) contains no disulfides. This suggests that its structural stability is solely due to hydrophobic interactions. Calculation of fuzzy oil drop parameters appears to confirm this view. Another protein stabilized by its hydrophobic core will be presented in the chapter which discusses the relation between structural stability and disulphide bonds (Chapter 7) as well as in Refs. [10,11]. As it turns out, both factors, i.e. hydrophobic interactions and disulfides, may either reinforce or counteract each other and therefore either promote or frustrate tertiary stability.

Stabilization in the presence of disulfide bonds

Modern biochemistry textbooks single out hydrophobic cores and disulfide bonds as factors which stabilize the tertiary conformations of proteins. In the example presented in this chapter both factors act cooperatively and are mutually reinforcing. Our analysis concerns a bioactive peptide named τ-AnmTx Ueq 12-1 (short name: Ueq 12-1), isolated from the sea anemone *Urticina eque* listed in PDB under ID 5LAH—a short protein (45 aa) which nevertheless contains five disulfide bonds [10].

Bioactive peptide—Ueq 12-1 (5LAH) is a highly peculiar protein owing to the arrangement of its cysteines. Three adjacent Cys residues form three separate disulfide bonds—a rarely observed pattern. The multitude of

disulfides renders the protein very stable; what is more, the hydrophobic core is also extraordinarily well ordered and consistent with the theoretical distribution of hydrophobicity (Table 5.2)—this goes for individual secondary folds as well as for fragments connecting Cys residues which form disulfide bonds. To enable comparative analysis we also list the corresponding correlation coefficients.

Taken together, the presented parameters reveal high consistency between T and O, which, in turn, shows that the protein adopts a micelle-like conformation. The network of disulfide bonds appears to further reinforce this structure. Both factors (hydrophobicity and disulfide bonds) work toward the same goal. Detailed analysis of individual folds shows that each of them is also consistent with the theoretical model (Fig. 5.7, Fig. 5.8).

Table 5.2 Values of fuzzy oil drop parameters calculated for the structure of 5LAH.

	RD		Correlation coefficients		
Bioactive peptide (5LAH)	**T-O-R**	**T-O-H**	**HvT**	**TvO**	**HvO**
PROTEIN	0.294	0.321			
SS					
1-8	0.228	0.309	0.798	0.787	0.668
11-42	0.306	0.348	0.432	0.764	0.744
17-35	0.353	0.321	0.309	0.702	0.688
22-43	0.318	0.450	0.447	0.754	0.793
29-44	0.296	0.549	0.550	0.755	0.869
β-structure					
9-11	0.141	0.119	0.549	0.996	0.478
32-34	0.184	0.104	0.671	0.962	0.847
β-sheet	0.178	0.095	0.350	0.926	0.600
β-structure					
16-18	0.446	0.063	0.889	0.988	0.810
41-44	0.490	0.973	0.310	0.337	0.987
27-29	0.373	0.734	0.493	0.703	0.965
β-sheet	0.422	0.577	0.424	0.562	0.844

"SS" corresponds to fragments bracketed by disulfide bonds. Individual β-strands and the resulting β-sheets are also listed.

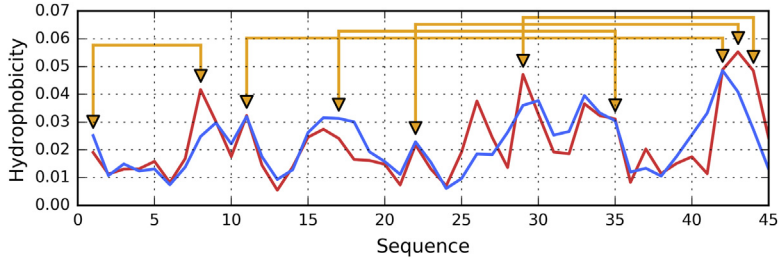

Fig. 5.7 Theoretical (T, blue (dark gray in print version)) and observed (O, red (gray in print version)) hydrophobicity distribution profiles for bioactive peptide Ueq 12-1 (5LAH). Orange lines correspond to SS bonds formed by Cys residues.

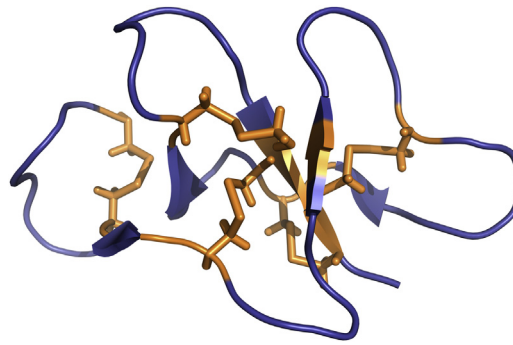

Fig. 5.8 3D presentation of bioactive peptide Ueq 12-1 (5LAH) showing the location of disulfides.

Proteins consistent with the theoretical model—spherical micelles

Referring back to the hypothesis stated in the introduction, that the presence of a hydrophobic core is a result of interactions between the peptide and its aqueous environment, we propose that the protein resembles a micelle whose structure is determined by hydrophobic interactions (notwithstanding differences between the building blocks of proteins and surfactant micelles). The final product—when characterized by a monocentric distribution of hydrophobicity—may be likened to a spherical micelle. In this section we discuss proteins which, regardless of structural factors not found in surfactant micelles (variable intrinsic hydrophobicity of amino acids; covalent bonding between components), adopt a spherical micelle-like conformation. Ultrafast-folding proteins provide a particularly fitting example of this group.

With regard to our previous discussion concerning sources of information required in the folding process, the presented proteins show that—in addition to information carried by the peptide sequence itself—the process also relies on information obtained from the aqueous solvent, which directs polar residues to the surface while shielding hydrophobic residues from contact with water. As can be seen on the example of 5LAH, the system of disulfide bonds may remain in perfect harmony with the hydrophobic core structure—although there are also proteins where such bonds prevent an organized hydrophobic core from forming (an example is discussed in Ref. [11]).

For the sake of contrast we will now discuss a protein which strongly diverges from the expected distribution of hydrophobicity.

Counter-example—prefoldin

The prefoldin β-subunit from *Thermococcus strain KS-1* (PDB ID: 2ZQM) exhibits major deviations from the expected distribution of hydrophobicity. It consists of two anti-parallel helices linked by a protruding β-hairpin. While it essentially lacks a tertiary conformation, CATH categorizes is as 1.10.287.370—mainly alpha orthogonal bundle [12].

Fuzzy oil drop model parameters calculated for this protein are as follows: RD(T-O-R) = 0.621; RD(T-O-H) = 0.619; HvT = 0.035; TvO = 0.249; HvO = 0.715. Taken together, these values indicate that the protein's structure is dominated by the conformational preferences of individual residues and that it lacks a monocentric hydrophobic core (note the low value of TvO and the correspondingly high value of HvO).

Fig. 5.9 illustrates the scope of discordance between T and O. Together with Fig. 5.10 it proves that micellar characteristics are by no means universal among proteins.

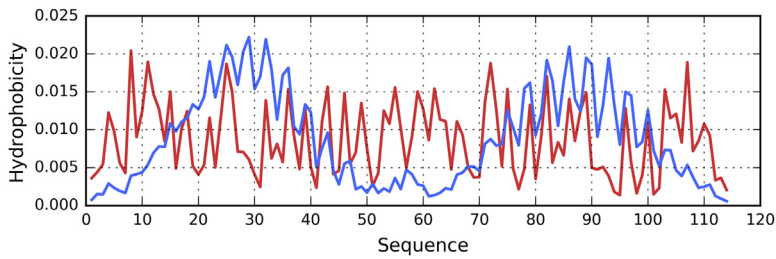

Fig. 5.9 Theoretical (T, blue (dark gray in print version)) and observed (O, red (gray in print version)) hydrophobicity distribution profiles for prefoldin β-subunit (2ZQM) revealing protein's major deviations from the expected distribution of hydrophobicity.

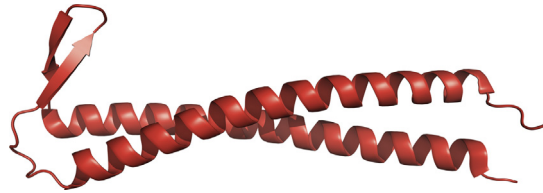

Fig. 5.10 3D presentation of prefoldin β-subunit (2ZQM).

Small proteins in particular (<60 aa) exhibit a whole range of discordances between O and T. Nevertheless, the presented work aims on validating the core assumptions of the fuzzy oil drop model, which is why we focus on proteins (largely) accordant with the model.

References

[1] Ye Q, Leinala E, Jia Z. Structure of type III antifreeze protein at 277 K. Acta Crystallographica Section D Biological Crystallography 1998;54(4):700–2. https://doi.org/10.1107/s0907444997020040.

[2] Labeit S, Barlow DP, Gautel M, Gibson T, Holt J, Hsieh C-L, Trinick J. A regular pattern of two types of 100-residue motif in the sequence of titin. Nature 1990; 345(6272):273–6. https://doi.org/10.1038/345273a0.

[3] Banach M, Konieczny L, Roterman I. The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. Journal of Theoretical Biology 2014;359:6–17. https://doi.org/10.1016/j.jtbi.2014.05.007.

[4] Pernigo S, Fukuzawa A, Bertz M, Holt M, Rief M, Steiner RA, Gautel M. Structural insight into M-band assembly and mechanics from the titin-obscurin-like-1 complex. Proceedings of the National Academy of Sciences 2010;107(7):2908–13. https://doi.org/10.1073/pnas.0913736107.

[5] Göktepe S, Abilez OJ, Parker KK, Kuhl E. A multiscale model for eccentric and concentric cardiac growth through sarcomerogenesis. Journal of Theoretical Biology 2010;265(3):433–42. https://doi.org/10.1016/j.jtbi.2010.04.023.

[6] Munoz V, Cerminara M. When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. Biochemical Journal 2016;473(17):2545–59. https://doi.org/10.1042/bcj20160107.

[7] Kubelka J, Chiu TK, Davies DR, Eaton WA, Hofrichter J. Sub-microsecond protein folding. Journal of Molecular Biology 2006;359(3):546–53. https://doi.org/10.1016/j.jmb.2006.03.034.

[8] Sałapa K, Kalinowska B, Jadczyk T, Roterman I. Measurement of hydrophobicity distribution in proteins - non-redundant protein data bank. Bio-Algorithms and Med-Systems 2012;8. https://doi.org/10.2478/bams-2012-0023.

[9] Kalinowska B, Banach M, Wiśniowski Z, Konieczny L, Roterman I. Is the hydrophobic core a universal structural element in proteins? Journal of Molecular Modeling 2017;23(7). https://doi.org/10.1007/s00894-017-3367-z.

[10] Logashina YA, Solstad RG, Mineev KS, Korolkova YV, Mosharova IV, Dyachenko IA, Andreev YA. New disulfide-stabilized fold provides sea anemone peptide to exhibit both antimicrobial and TRPA1 potentiating properties. Toxins 2017; 9(5):154. https://doi.org/10.3390/toxins9050154.

[11] Banach M, Kalinowska B, Konieczny L, Roterman I. Role of disulfide bonds in stabilizing the conformation of selected enzymes—an approach based on divergence entropy applied to the structure of hydrophobic core in proteins. Entropy 2016;18(3):67. https://doi.org/10.3390/e18030067.

[12] Kida H, Sugano Y, Iizuka R, Fujihashi M, Yohda M, Miki K. Structural and molecular characterization of the prefoldin β subunit from Thermococcus strain KS-1. Journal of Molecular Biology 2008;383(3):465–74. https://doi.org/10.1016/j.jmb.2008.08.041.