Downhill, Ultrafast and Fast Folding Proteins Revised

Mateusz Banach [1], Katarzyna Stapor [2] (ID), Leszek Konieczny [3], Piotr Fabian [2] (ID) and Irena Roterman [1,*] (ID)

[1] Department of Bioinformatics and Telemedicine, Jagiellonian University—Medical College, Lazarza 16, 31-533 Krakow, Poland; mateusz.banach@uj.edu.pl

[2] Institute of Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland; katarzyna.stapor@polsl.pl (K.S.); piotr.fabian@polsl.pl (P.F.)

[3] Chair of Medical Biochemistry, Medical College, Jagiellonian University, Kopernika 7, 31-034 Krakow, Poland; mbkoniec@cyf-kr.edu.pl

* Correspondence: myroterm@cyf-kr.edu.pl; Tel.: +48-12-619-96-93

check for
updates

Abstract: Research on the protein folding problem differentiates the protein folding process with respect to the duration of this process. The current structure encoded in sequence dogma seems to be clearly justified, especially in the case of proteins referred to as fast-folding, ultra-fast-folding or downhill. In the present work, an attempt to determine the characteristics of this group of proteins using fuzzy oil drop model is undertaken. According to the fuzzy oil drop model, a protein is a specific micelle composed of bi-polar molecules such as amino acids. Protein folding is regarded as a spherical micelle formation process. The presence of covalent peptide bonds between amino acids eliminates the possibility of free mutual arrangement of neighbors. An example would be the construction of co-micelles composed of more than one type of bipolar molecules. In the case of fast folding proteins, the amino acid sequence represents the optimal bipolarity system to generate a spherical micelle. In order to achieve the native form, it is enough to have an external force field provided by the water environment which directs the folding process towards the generation of a centric hydrophobic core. The influence of the external field can be expressed using the 3D Gaussian function which is a mathematical model of the folding process orientation towards the concentration of hydrophobic residues in the center with polar residues exposed on the surface. The set of proteins under study reveals a hydrophobicity distribution compatible with a 3D Gaussian distribution, taken as representing an idealized micelle-like distribution. The structure of the present hydrophobic core is also discussed in relation to the distribution of hydrophobic residues in a partially unfolded form.

Keywords: ultrafast folding; downhill folding; hydrophobic core

1. Introduction

Despite many years of research [1], the mechanism of the protein folding process remains unresolved. The protein folding process turns out to be varied, including the time it takes to obtain the native structure. There are proteins with a small number of amino acids in the chain which are characterized by a very fast folding process. This group of proteins is known as fast-folding, ultrafast folding or downhill [2].

Another group of proteins are those that fold on their own, although the participation of the ribosome is required [3]. An example of an extremely demanding process is the folding process which requires the participation of chaperones or chaperonins [4–8].

When discussing the issue of protein folding, the presence of many intrinsically disordered proteins should also be noted [9]. Generally speaking, in statistical terms, the course of the folding process should last a very long time [10]. However, this is not so. Fast-folding proteins reach their

final structural form within a dozen microseconds [11]. The identified multi-stage nature of the folding process introduces the term intermediates, the number of which is characteristic for a given protein. [12–15]. An often identified state of partial unfolding is referred to as the molten globule state [16,17].

New perspectives are emerging with the appearance of techniques based on the "single molecule" experiment, including nuclear magnetic resonance, relaxation dispersion NMR spectroscopy [18] (in particular, photo-protection strategy [19]), as well as single molecule fluorescence spectroscopy sub-millisecond conformational dynamics [20]. Single molecule fluorescence resonance energy transfer (SM-FRET) utilizing fluorescent labeling and immobilization of proteins opens the possibility of quantitative (temporal) analysis of the folding process for fast-folding proteins [21].

Among the possible mechanisms of the protein folding process, hydrophobic collapse is considered [22,23]. Thermodynamic analysis even suggests that the smallest proteins are expected to have very marginal free energy barriers to folding [24–27]. The presence of specific secondary forms, including the Beta-hair pin, is important in obtaining the final structure of the protein [28,29].

A list of widely used experimental techniques is provided in a review [30]. The current analysis takes into account the presence of the water environment in folding the protein structure. The influence of the water environment on the structure of proteins as described by the fuzzy oil drop model has already been discussed many times, including structures with distribution expressed by the 3D Gaussian function and structures showing local incompatibility both in the form of local excess [31] and deficit hydrophobicity [32].

The conclusion drawn from these observations indicates that there are sequences that are capable of generating structures with a highly ordered hydrophobic core. Local non-compatibility with the idealized distribution can even be treated as intentional. This imperfection is a form of encoding the biological function of a given protein. Proteins with micelle-like hydrophobicity distribution are very soluble, but completely lack interaction opportunities with other molecules, except for surface complexing of ions or low-molecular compounds on the basis of charge interaction. Proteins with a high degree of similarity in the hydrophobicity distribution to the idealized one would be devoid of their biological activity. It turns out, however, that the biological function of proteins with micelle-like distribution is of biological importance in the case of antifreeze type III proteins, where high solubility is desirable and where binding or complexation of other molecules is not expected [33].

The aim of the presented analysis is to propose a folding mechanism for proteins characterized by a fast-folding process based on the active participation of the water environment in the generation of a hydrophobic core. This directs the process towards the concentration of hydrophobic residues in the center with simultaneous exposure of polar residues on the surface [34–37].

To demonstrate an additional factor promoting rapid reach of the native structure, partially unfolded structures based on the early-stage model have also been included in the analysis [38–40]. Partially unfolded structures that retain the fragmentary concentration of hydrophobic residues and contain a seed for the construction of a hydrophobic core favor a high prognosis for a reversible process.

2. Results

As previously mentioned, the potential high rate of folding of the polypeptide chain was considered by evaluating the degree of generation of the hydrophobic core, in which the water environment was actively involved as the dominant factor. The hydrophobic core was regarded as an important factor stabilizing the tertiary structure (apart from the disulfide bonds). According to the fuzzy oil drop model, the term "presence of a hydrophobic core" means both the concentration of high hydrophobicity in the center of the molecule together with the presence of a polar surface coating.

The status expressed by the values of RD (Relative Distance) parameters for both the native and partially unfolded forms is given in Table 1. Figure 1 presents the idea of obtaining a partially unfolded form generated based on the Phi_e and Psi_e angles.

Table 1. The values of the RD (Relative Distance) parameter for proteins under study.

PDB ID	RD—LS	RD—LS after Fragment Selection/Elimination		RD—ES	RD—ES after Fragment Selection/Elimination	
2CYU	**0.638**	3-helix system	0.493	**0.540**	18–23	0.431
2L6Q	0.473			**0.595**	7–10,21–26	0.488
2L6R	0.477	34–37 eliminated	0.462	0.439		
1BBL	**0.556**	12–33 selected 38–48 selected	0.349 0.311			
2WXC	0.382			**0.525**	165-C-term	0.441
2A3D	**0.580**	21–24 eliminated	0.474	**0.710**	28–45 selected	0.474
2F21	0.402			**0.761**		
		1–49 selected	0.520	**0.657**		
		50–163 selected	0.310	**0.697**		
2P6J	0.298			**0.504**	helix 5–17	

ES (early, limited conformational sub-space) and LS (late stage, structure available in PDB) forms, respectively. Selected: status of the given fragment showing the local accordance; Eliminated: the RD value after elimination of a given fragment; Helix orientation: only the re-orientation of a given helix expected for lowering the RD value; Values in bold: structure with RD above 0.5.
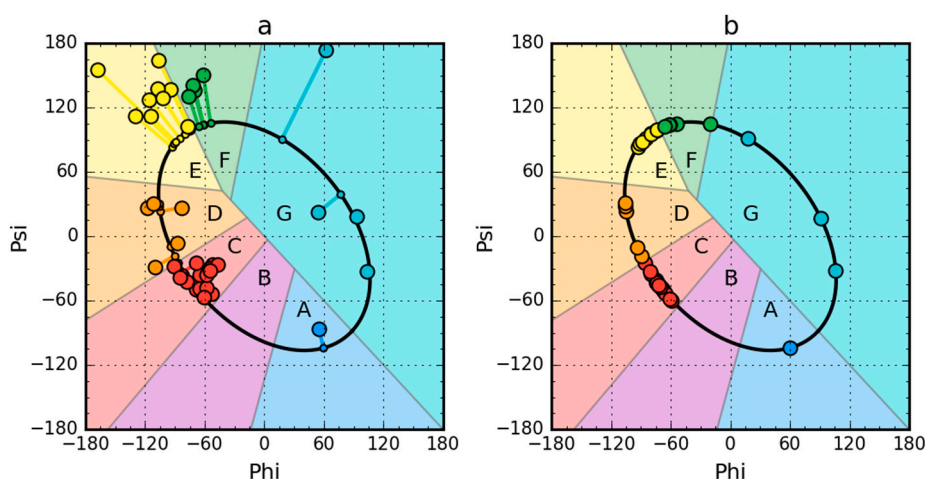

Figure 1. Model of a partially unfolded structure for peripheral-subunit binding domain (PDB ID: 1W4E) protein used as example. (**a**) Phi, Psi angles distribution in the structure available in PDB transformed to the appropriate Phi_e and Psi_e angles (shortest distance criterion). (**b**) Status after transformation and optimization. The colors distinguish the zones. A–G are defined according to seven local maxima on the ellipse path (limited conformational sub-space). The ellipse path defined according to Phi, Psi change causing the change of radius of curvature of the polypeptide.

Partially unfolded structures for all the proteins under study were obtained as shown in Figure 1. The proteins presented in Table 1 show RD values below 0.5 in the vast majority for the native structures available in PDB. The structure of the e3-binding domain (PDB ID: 2CYU) for the complete structure shows an RD greater than 0.5. However, this value is due to loose loops. The system of three helices constituting the main part of the whole molecule shows RD = 0.493, which indicates the formation of a hydrophobic concentration in the center of the molecule.

The molecule 1BBL, also showing an RD above 0.5, reduces the value to 0.462 when the four residues in the loop are eliminated. Moreover, the status of two components of the whole molecule composed of helical fragments 12–33 and 38–48 (neglecting the loose loop) indicates a very high adjustment of the hydrophobicity distribution to the micellar system (both values below 0.4).

The hypothesis that the active participation of the water environment in protein folding, in which the formation of the hydrophobic core is important, has been demonstrated.

2.1. Peripheral-Subunit Binding Domain (1W4E)

The peripheral-subunit binding domain (PDB ID: 1W4E) represents a set of models with an identical sequence. The status of both the native and partially unfolded form shows the presence of a high degree of arrangement compatible with the fuzzy oil drop model. In this case, the reconstruction of a structure with a high degree of packing which satisfies the conditions resulting from this model, is obvious (Figure 2).
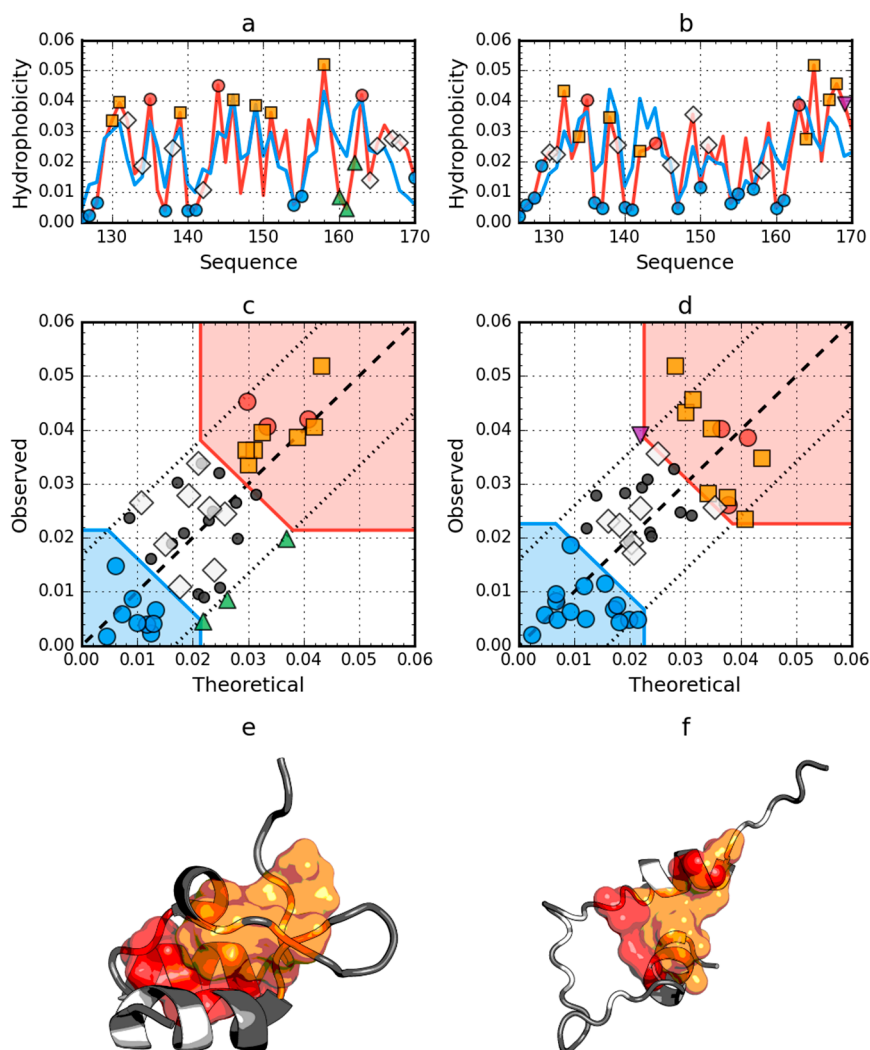

Figure 2. Hydrophobicity density distributions and 3D presentation of native (**a**,**c**,**e**) and early stage (**b**,**d**,**f**) forms of peripheral-subunit binding domain (PDB ID: 1W4E). Blue line on (**a**,**b**): theoretical distribution (*T*); red line on (**a**,**b**): observed distribution (*O*); red zone on (**c**,**d**): hydrophobic core; blue zone on (**c**,**d**): hydrophilic surface; red circle markers: common hydrophobic core members in both models; orange square markers: hydrophobic core members in this model but not in the other model; white rhombus markers: hydrophobic core members in other model but not in this model (possible overlap with other markers); blue circle markers: hydrophilic surface members in this model; green triangle markers: *T* value >> *O* value; purple triangle markers: *T* value << *O* value: black circle markers: hydrophobically insignificant residues. Colored segments on (**e**,**f**) correspond to markers of the same color on (**a**–**d**), respectively (hydrophobic core is shown as surface).

2.2. W Protein of Bacteriophage Lambda (2L6Q)

The structure of W protein of bacteriophage lambda (PDB ID: 2L6Q) shows the presence of an arranged core in its native form (Figure 3 and Table 1). In the 2L6R, the presence of the fuzzy oil drop model arrangement was also demonstrated in a partially unfolded form. In the case of the unfolded 2L6Q, only the segment Beta-strand 22–28 shows a local distribution different from the expected one. The low RD value for the native form, however, suggests that the process of hydrophobic core generation gives this segment a status compatible with the idealized distribution (Figure 3a).
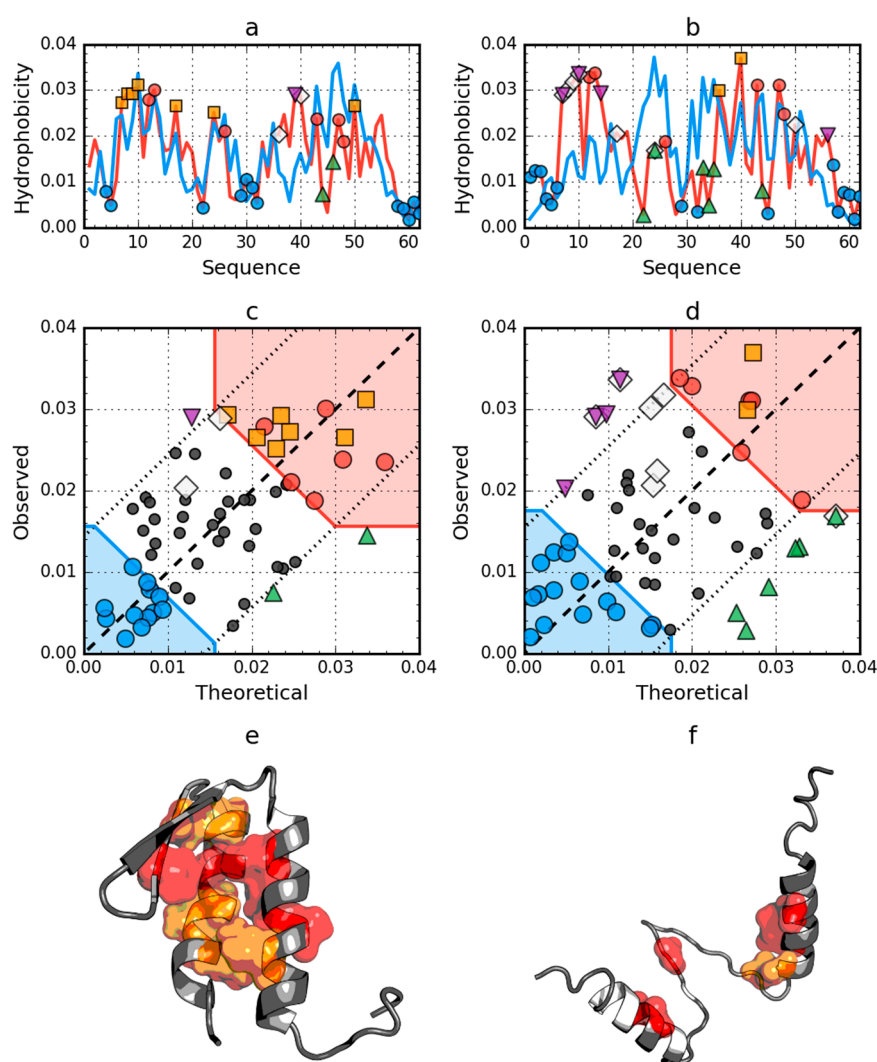

Figure 3. Hydrophobicity density distributions and 3D presentation of native (**a**,**c**,**e**) and early stage (**b**,**d**,**f**) forms of W protein of bacteriophage lambda (PDB ID: 2L6Q). Blue line on (**a**,**b**): theoretical distribution (T); red line on (**a**,**b**): observed distribution (O); red zone on (**c**,**d**): hydrophobic core; blue zone on (**c**,**d**): hydrophilic surface; red circle markers: common hydrophobic core members in both models; orange square markers: hydrophobic core members in this model but not in the other model. White rhombus markers: hydrophobic core members in other model but not in this model (possible overlap with other markers); blue circle markers: hydrophilic surface members in this model; green triangle markers: T value >> O value; purple triangle markers: T value << O value; black circle markers: hydrophobically insignificant residues. Colored segments on (**e**,**f**) correspond to markers of the same color on (**a**–**d**), respectively (hydrophobic core is shown as surface).

2.3. Other Proteins

The analysis of two further sample proteins (thermostable protein variant engrailed homeodomain (PDB ID: 2P6J) and de novo designed triple helix bundle protein (PDB ID: 2A3D)) are included in Supplementary Materials (Figures S1 and S2) to avoid redundancy.

2.4. Comparative Analysis of Partially Unfolded Structures

Using a set of structural codes (A–G) where C stands for the right-handed helix, E stands for the beta structure and G represents the left-handed helix, the presence of these forms of the secondary structure of the proteins in question was determined (Figure 4). The D code, which expresses a structural form intermediate between the helix and the beta form, is also worth mentioning. Also important is the F code, which represents the forms present at the end of the beta strand, introducing a significant reduction in the radius of curvature in relation to the large radius values for the beta-strand form.
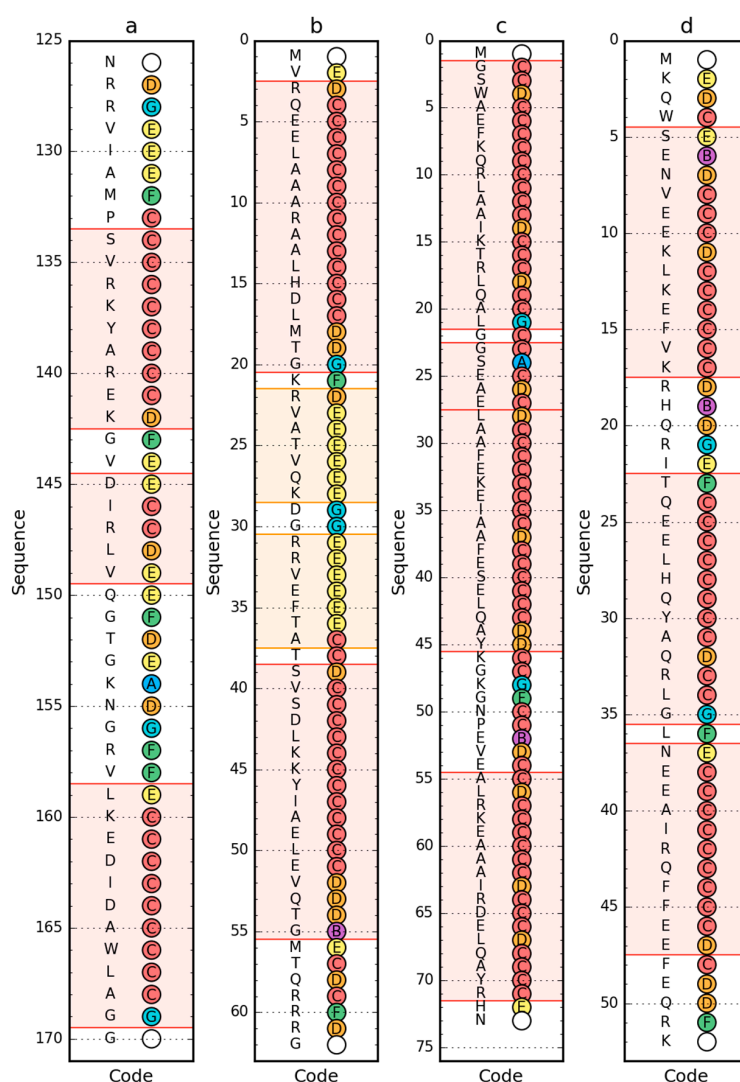

Figure 4. Sequence (left) and structural codes (right) of peripheral-subunit binding domain (PDB ID: 1W4E) (**a**); W protein of bacteriophage lambda (PDB ID: 2L6Q) (**b**); de novo designed triple helix bundle protein (PDB ID: 2A3D) (**c**); and thermostable protein variant, engrailed homeodomain (PDB ID: 2P6J) (**d**). Colors of the circle markers correspond to structural codes (A: blue, B: purple, C: red, D: orange, E: yellow, F: green, G: cyan). Chain termini are marked by blank circles. The frames are as follows: red, helical form (code C expected); yellow/orange, β-structural (code E expected).

The form of secondary structure as classified by PDB [40] with PROMOTIF [41] turns out to be different to those determined by structural codes (Figure 5). This is especially true of engrailed homeodomain (PDB ID: 2P6J), where in helical fragments 5–17 and 23–35 are also defined by structural codes other than C. In other cases, the codes differing from the classified form occur only for the N- and C-terminal residues in individual segments. This is an irrelevant difference from the point of view of determining the status of a given residue in the context of the accompanying segment. However, it strongly influences the formation of ES structural form.
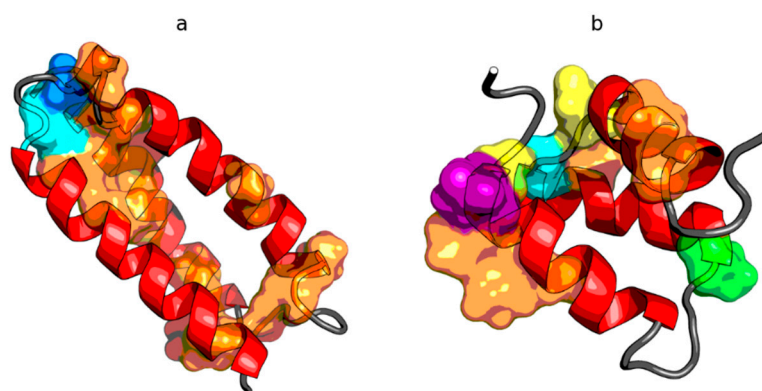

Figure 5. 3D presentation of de novo designed triple helix bundle protein (PDB ID: 2A3D) (**a**) and engrailed homeodomain (PDB ID: 2P6J) (**b**) with localization of residues representing structural codes other than the expected C within the helical fragments. These discrepancies are shown as the surface, with colors matching the color codes in Figure 4c or Figure 4d respectively (the helices themselves are red, just like code C is represented by red on Figure 4).

2.5. Negative Control

Among the proteins acting as an anti-freeze agent is the winter flounder derived protein (PDB ID: 1WFA). This short polypeptide chain (37 aa) does not generate a globular form; instead, it takes the form of a helix that spans the entire chain. This form is unable to generate a hydrophobic core. The use of the fuzzy oil drop model unanimously reveals the status of this protein as not representing a micelle-like form. The RD value for this protein is 0.837. The T and O profiles illustrating the hydrophobicity distribution show the expected hydrophobicity concentration, which was not observed (Figure 6). There was also no significant reduction in the level of observed hydrophobicity as one approached the terminal positions of the amino acids.
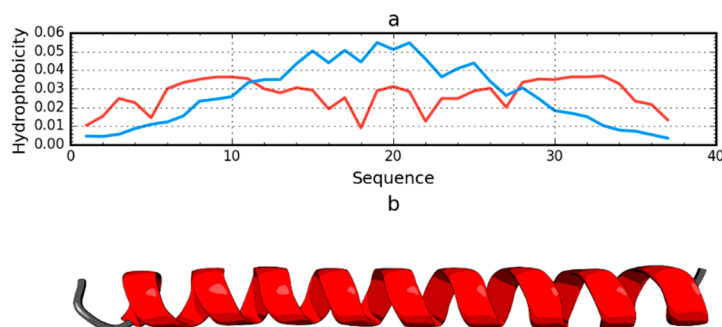

Figure 6. Hydrophobicity density distributions (**a**) and 3D presentation (**b**) of antifreeze protein from winter flounder (PDB ID: 1WFA). Blue line: theoretical distribution (*T*); red line: observed distribution (*O*).

In addition to the drastic example of the incompatibility of the hydrophobicity distribution versus the idealized distribution, human lysozyme is an example of a protein (PDB ID: 1LZ1) with clearly

marked fragments with the distribution *O* and *T* highly compatible with the present chain segments showing local incompatibility (Figure 7).
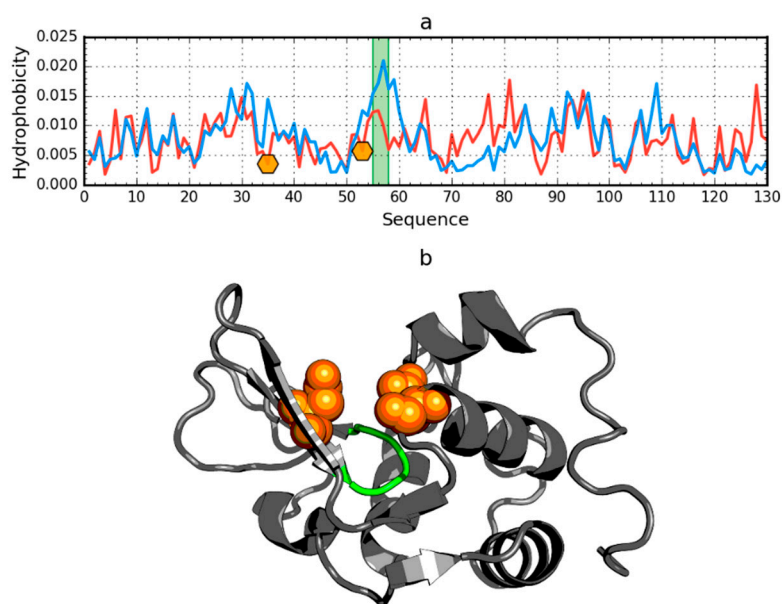

Figure 7. Hydrophobicity density distributions (**a**) and 3D presentation (**b**) of human lysozyme (PDB ID: 1LZ1). Blue line: theoretical distribution (*T*); red line: observed distribution (*O*). Orange markers on (**a**) and orange spheres on (**b**) denote catalytic residues (E35, D53). Green segments on (**a**,**b**) span residue range 55–58.

The region of incompatibility is located in the active center of this enzyme, as shown by the compared *T* and *O* profiles of this protein. The RD value for lysozyme is 0.529. The elimination of the catalytic residues (35 and 53) from the status calculations (according to divergence entropy) results in the reduction of the RD value to 0.515. After further elimination of residues 55–58, the value RD = 0.499 was obtained. It should be noted that residues 55–58 are located in the immediate vicinity of the active site and are probably an integral component of the active site (Figure 7).

The elimination of the indicated residues (catalytic and a segment in the immediate vicinity of the catalytic residues) resulting in an RD value less than 0.5 indicates that the entire remainder of the chain satisfies the micelle-like degradation conditions. It also means that the remainder of the chain guarantees the solubility of the protein in question.

T versus *O* plots of these two proteins (Figure 8) corroborate the above findings. There are no hydrophobic core members in the antifreeze protein (red zone in Figure 8a) and the location of residues in the *T* × *O* space suggests very low correlation between the distributions, which is in fact approximately equal to 0.08. In the lysozyme, catalytic residues are marked by the FOD model as showing hydrophobicity deficiency in the center of the molecule, just like the F57 and Q58, which contribute to the RD value being higher than 0.5. Their elimination (together with the rest of the 55–58 segment) is enough to offset the other increase of RD above 0.5 caused by three exposed cys residues (numbers 6, 77, 128), which can be seen as the top-most triangle markers (purple) on Figure 8b.
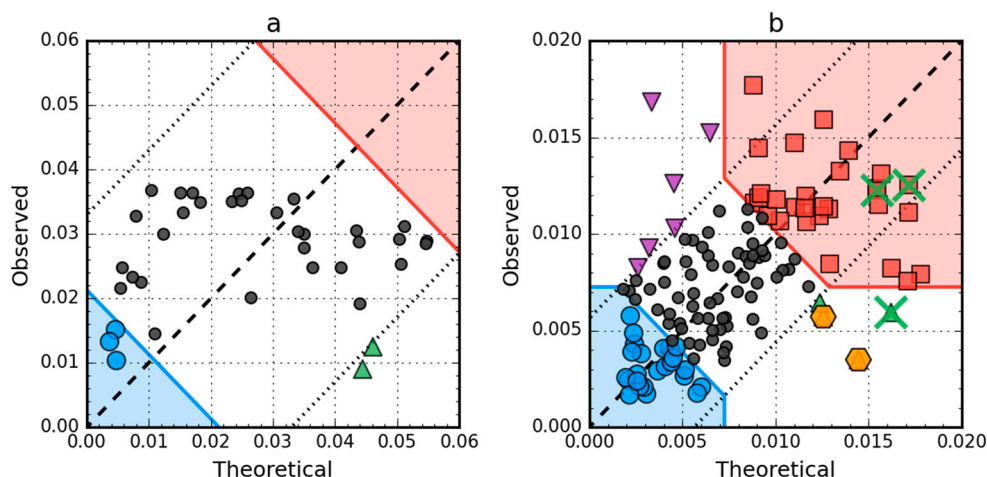
Figure 8. Hydrophobicity density scatter plots (*X*-axis: *T*, *Y*-axis: *O*) of antifreeze protein from winter flounder (PDB ID: 1WFA) (**a**) and human lysozyme (PDB ID: 1LZ1) (**b**). Red square markers: hydrophobic core members; blue circle markers: hydrophilic surface members; green triangle markers: *T* value >> *O* value; purple triangle markers: *T* value << *O* value; black circle markers: hydrophobically insignificant residues. Red and blue backgrounds are areas in which residues are classified as belonging to either core or surface. Orange hexagonal markers on (**b**) denote catalytic residues (E35, D53). Green "×" markers on (**b**) span residue range 55–58 (F57 has been omitted for clarity; it is to the right of the *X*-axis limit, *T* = 0.209).

3. Discussion

Proteins with high RD are mainly proteins representing tertiary structures, where the mutual spatial arrangement of the domains results in the appearance of areas with high RD values. Very often in a protein with the status expressed by the value of RD > 0.5, it is enough to eliminate single residues or small fragments of the chain to obtain the value of RD < 0.5. The residues identified in this way often turn out to be catalytic residues or residues involved in the complexation of another protein. This issue is discussed in the book [42] with a detailed analysis of examples.

For the topic discussed in this article, it is also important to demonstrate the highly ordered hydrophobic core observed in titin [43]. This molecule, used in the SM-FRET method [21] as a frame for the analysis of the protein unfolding process, has a single-step unfolding form. Therefore, easy identification of the unfolding of the frame for other proteins enables a detailed analysis of the unfolding process of other proteins with more complex kinetics. The low RD value for the sandwich domain observed in the immunoglobulin domains just in titin distinguishes it from many other immunoglobulin-like domains [44]. The specificity of the incompatibility of the *T* and *O* distributions in other immunoglobulin-like domains assigns them their specificity.

The early-stage intermediate structure, based on the relaxed arrangement of peptide bond planes, reveals the degree of presence of a seed with a hydrophobic core. The presence of local hydrophobicity maxima compatible with those expected in the structure generated on the basis of Phi_e and Psi_e angles(and thus partially unfolded) means that the further folding process will be an expansion of the hydrophobic core. Protein folding simulations taking into account the influence of the water environment in the form of introducing an external force field directing the folding process consistent with 3D Gauss function were carried out [45]. A large-scale protein folding project taking into account these external conditions is currently being implemented.

The widely discussed phenomenon of cooperativity in protein folding seems to have its mathematical expression [45]. Generation of the hydrophobic core requires cooperativity and is the effect of a specific synergy, which is obtained in the process of folding in a way strongly dependent on the amino acid sequence. For a given polypeptide chain, the degree of ease/difficulty to obtain this synergy is primarily expressed by the folding time. This ease is obviously determined by the

amino-acid sequence, which favors the micelle-like structure more or less. The proteins discussed here are present in hyperthermophilic bacteria. The stable arrangement form of the hydrophobic core is probably the reason for protein stability in these organisms.

The entire non-redundant PDB database showed that a significant number of domains (over 90%) have a hydrophobicity arrangement consistent with the 3D Gaussian distribution [46]. This is in strict accordance with the reports on two-state systems present in protein domains as structural units [47].

The same is true of antifreeze type III proteins, which show hydrophobicity distribution consistent with the 3D Gauss distribution [48]. Interpretation of this phenomenon explains their antifreeze action relying on the arrangement of water molecules in a way resulting from the distribution of charges on the surface of this protein, which is different from those present in the ice structure.

Noteworthy is the de novo protein synthesis experiment, the aim of which was to obtain a significant structural change with a minimum number of introduced mutations. It was obtained for a set of four proteins (56 aa), which differ in turn by one introduced mutation, resulting in alternating structures: 2LHC—3α, 2LHD—4β + α, 2LHG 3α and 2LHE 4β + α [49]. The analysis carried out on the basis of the fuzzy oil drop model showed that the cause of the different secondary structure involved different chain fragments in the construction of the hydrophobic core [37].

What do we learn from this paper that we did not know before?

1. Protein folding is the process of micellization of bipolar molecules (amino acids are molecules with different forms of bipolarity) with limited freedom of movement resulting from the presence of peptide bonds [42].

2. Down-hill proteins (fast-folding, ultrafast-folding) are proteins with an amino acid sequence that allows construction of a micelle-like structure, which in the case of antifreeze proteins guarantees solubility (i.e., the basic condition enabling the performance of a biological function, which is the order of water molecules imposed by the distribution of charges on the protein surface preventing the structuring present in the ice structure). This issue is discussed in this paper.

3. Amino acid sequence determining the protein structure is a sequence with a specific coding system excluding the possibility of generating the perfect micelle, which is the local discordance versus 3D Gauss. The discordance is represented as local hydrophobicity deficiency and usually appears in the ligand or substrate binding cavity [32] or local hydrophobicity excess which is "used" by proteins to construct interfaces in protein–protein complexes [31]. Lysozyme is an example of cavity identification as shown in this paper.

4. The form and degree of the encoded inability to reconstruct the micelle-like pattern determines the specificity of a given protein.

5. In this context, amyloid (formed in physiological conditions) is also a result of micellization. The final product however takes the form of ribbon-like micelle since these polypeptides are not able to construct the globular forms [50].

The above mentioned conclusions are the generalizations resulting from the current work but also from the analysis of numerous proteins described in other publications [33–36,42,43,50]. The present work discusses a set of proteins representing a high agreement of hydrophobicity distribution in relation to an idealized distribution consistent with 3D Gauss function. In summary, proteins can be defined as "intelligent micelles" with encoded local discordance which expresses the specificity of particular protein molecules. "Amino acid sequence determines the structure" can be substituted by, "amino acid sequence determines the form and degree of specific disability in forming the ideal micelle." This disability determines the specificity of the protein.

4. Materials and Methods

4.1. Data

The object of analysis are the proteins listed in Table 2. They represent examples of proteins classified as fast folding. Proteins except de novo, viral and ones derived from Homo sapiens are bacterial proteins.

Table 2. List of proteins under consideration in this paper.

PDB-ID	Protein	Chain Length	Characteristics	Ref
2CYU	e3-binding domain	39 aa	Downhill	[51]
2L6Q	Viral protein	62 aa		[52]
2L6R	Viral protein	62 aa		[53]
1W4E	Peripheral-subunit binding domain	45 aa	Ultrafast	[54]
1W4F		45 aa		[54]
1W4G		45 aa		[54]
1W4H		45 aa		[54]
2BTH		45 aa		[54]
1W4J		51 aa		[54]
1W4K		51 aa		[54]
1BBL		37 aa		[55]
2WXC		47 aa		[56]
Negative Control				
1WFA	Antifreeze	37 aa	Antifreeze	[57]
1LZ1	Lysozyme	130 aa	Enzyme	[58]

The list of proteins that are the subject of analysis also includes a negative example in the sense of being assessed by the fuzzy oil drop model. It is a protein of small size (37 aa) taking a helical form along the entire chain length. This form, not representing the globular form, does not necessarily generate the presence of a hydrophobic core.

4.2. Late Stage Model—Fuzzy Oil Drop

The fuzzy oil drop model (FOD) has already been described in detail [59–61]. The oil drop model introduced by Kauzmann [62] was modified to form the fuzzy oil drop model. There are two levels in this model, which are high in the center of the molecule and low in the surface layer. These have been modified to a continuous form by introducing a 3D Gaussian function spread over the protein molecule. In this situation, each amino acid can be assigned the so-called T_i, which is the level of hydrophobicity resulting from the location of a given residue represented by the effective atom (averaged position of atoms making up the amino acid). This is the so-called idealized level (i.e., the level that a given residue would represent if the distribution was completely compatible with the theoretical distribution (T)).

In order to determine the status consistent with the 3D Gaussian distribution, the following function was used:

$$\widetilde{H}_i^T = \frac{1}{H_{\text{sum}}^T} \exp\left(\frac{-(x_i - \bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_i - \bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_i - \bar{z})^2}{2\sigma_z^2}\right). \tag{1}$$

The point $(\overline{x}, \overline{y}, \overline{z})$ is the position of the geometric center of the protein in the 3D coordinate system, placed in its origin, at (0,0,0). The protein should be oriented, making the line linking the longest distance between two effective atoms in the molecule coaxial with the *X*-axis. It is then rotated around the *X*-axis to make the line linking the two most distant positions of the projections of effective atoms on the YZ plane coaxial with the *Y*-axis. The three parameters σ_x, σ_y, σ_z represent standard deviations of the size of the protein, equal to 1/3 of the highest absolute values of the x-coordinate, y-coordinate and z-coordinate, respectively (according to the 3-sigma rule). The normalizing coefficient H_{sum}^T represents the sum of all H_i^T values of amino acids of the protein, making the H_i^T value unit-less. The only input information for the theoretical distribution is a geometrical term concerning the full protein (i.e., the size of the ellipsoid "drop" containing the protein, characterized by σ_x, σ_y, σ_z). Traditionally, the value of the Gauss function is interpreted as a theoretical idealized hydrophobicity density at any given point.

However, the status of a given amino acid may not be as expected. The actual status is the effect of the magnitude of the interactions of a given residue with neighboring residues. Thus, it depends on the distance between the interacting residues and on their own hydrophobicity. The observed level of hydrophobicity, called O, characterizes the status of a given residue in a given protein.

The equation proposed in [63] was applied to calculate the observed hydrophobicity:

$$H_i^0 = \frac{1}{H_{sum}^0} \sum_j \left\{ \begin{array}{l} \left(H_i^r + H_j^r\right)\left(1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right), \; for \; r_{ij} \leq c, \\ 0, for r_{ij} > c \end{array} \right. \tag{2}$$

where H_i^O denotes the experimentally observed hydrophobic density at a particular point which collects the hydrophobic interaction in distance-dependent form, as given in the formula with the cutoff distance (*c*) assumed according to original work 9Å. The denominator H_{sum}^O (sum of all H_i^O) makes the value in normalized form. H_i^r and H_j^r express the intrinsic hydrophobicity of *i*-th and *j*-th residues, which can be taken according to the arbitrarily selected scale.

After normalizing distributions *T* and *O*, it becomes possible to compare them quantitatively using divergence entropy D_{KL} introduced by Kullback-Leibler [64].

$$D_{KL}(P\|Q) = \sum_i P(i)\log_2 \frac{P(i)}{Q(i)}, \tag{3}$$

where $P(i)$ denotes the observed probability (hydrophobicity density) localized on the *i*-th residue, which in this paper is called O_i and $Q(i)$ denotes the expected (target distribution) hydrophobicity localized on the same residue, which in this paper is called T_i (theoretical one) corresponds to the distance between *O* and *T*, the latter of which is regarded as the reference.

Equation (3) takes the following form in our paper:

$$D_{KL}(O|T) = \sum_{i=1}^N O_i \log_2(O_i/T_i), \tag{4}$$

with T_i treated as reference distribution.

D_{KL} expresses the formal distance between both distributions (*T* and *O*). However, since it constitutes a measure of entropy, it cannot be interpreted on its own and a second reference model must be provided. Since *T* simulates a "perfect" hydrophobic core, we may add a reference distribution which lacks any concentration of hydrophobicity at any point in the protein body. This type of distribution—called the unified distribution (*R*)—assigns hydrophobicity of 1/N to each residue (N being the number of residues in the chain). It represents the status deprived of any form of hydrophobicity differentiation in the protein body.

$$D_{KL}(O|R) = \sum_{i=1}^N O_i \log_2(O_i/R_i), \tag{5}$$

with R_i as the reference distribution.

When considering O and R, the value of D_{KL} tells us to what degree the observed distribution approximates the unified distribution. Comparing both values (for O/T and O/R) provides a description of the protein's status. When $O/T < O/R$, the observed distribution is aligned with the theoretical distribution, and therefore the protein may be assumed to contain a hydrophobic core. In the opposite case ($O/T > O/R$) the protein lacks a prominent core.

By applying the 3D Gaussian model and calculating divergence entropy, we obtain a fine-grained description of the protein's status, a procedure which would not be possible under the original oil drop model. In order to avoid having to deal with two distinct values of D_{KL}, we compute another parameter referred to as relative distance (RD):

$$\mathrm{RD} = \frac{D_{KL}(O|T)}{D_{KL}(O|T) + D_{KL}(O|R)}. \tag{6}$$

The parameter RD is used to determine the status of a given molecule or domain. When assuming values less than 0.5, this indicates an arrangement consistent with the idealized system. Otherwise, the situation differs from the centric system of the hydrophobic core. The status determined may refer to the whole protein or to a domain and an appropriate 3D Gaussian function is generated for such a defined structural unit. It is also possible to determine the status of a selected chain fragment within a given unit (protein, domain), or to identify those residues which together represent a status deviating from the centric arrangement. Such a calculation is used, for example, to determine the status of the segment forming a disulfide bond, the surroundings of the catalytic residue, or the status of the residues involved in ligand binding. This procedure has been used in the analysis discussed here.

The RD parameter can be calculated for any structural unit under consideration. The status of the selected chain fragment can also be estimated. The status of the chain fragment can be treated only as part of a formerly defined structural unit. The status of a chain fragment requires normalization (sum of T_i and O_i belonging to selected chain fragments shall be equal to 1.0). Then the RD can be calculated after defining the structural unit it has been selected from. It may happen that the status of selected a chain fragment can be expressed by RD > 0.5 in respect to the chain or the domain of the status expressed by RD < 0.5.

4.3. Early-Stage Intermediate Model—Partially Unfolded Protein

When looking for sources of generating clusters of hydrophobic residues in the center of the molecule, an early-stage intermediate model was used for the analyzed proteins. It should be noted that it was constructed based solely on the preferences of the backbone itself. The system of peptide bonds—including the system of peptide bond planes in particular—showed that the relaxed forms of these planes determine the relationship of Phi and Psi rotation values and the radius of curvature for the pentapeptide is treated as a unit. The value of the angle between the planes of peptide bonds (with the virtual Cα-Cα bond as the axis of rotation) determines the value of the radius of curvature. Following angle changes from 0° to 180°, the radius of curvature changes from the smallest present in the helix to the largest present in the extender or beta form. These changes, determined with the use of the approximation function, indicate on the Ramachandran map an elliptical path connecting all areas with a specific secondary structure. It is assumed that as long as the backbone alone determines the structural form, we are dealing with an early state of the folding process. Replacing the Phi and Psi angles with the corresponding Phi$_e$ and Psi$_e$ angles (angles belonging to the elliptical path determined by the shortest distance between Phi and Psi and the ellipse) generates the form of an early intermediate when inter amino acid interactions do not take place yet.

Early intermediate structures are obtained using the Phi$_e$ and Psi$_e$ values. Such an analysis leads to the recognition of the 3D structure, in which—if the non-binding interactions present in proteins are included in the calculations—the final form is obtained.

The simulation of the folding process, where the starting structure is determined precisely on the basis of the described model and the folding process relies on minimizing non-binding interactions and taking into account the active participation of the water environment by introducing an external force field in the form of a 3D Gaussian function, has been presented [65].

Here, the early intermediate structures determined for the proteins in question will be analyzed comparatively in order to identify a common factor leading to downhill or fast-folding structures characterized by a high degree of agreement between the *T* and *O* distribution. This observation is interpreted as an example of the folding process dominated by a tendency to the generation of a hydrophobic core. The unequivocal orientation of the folding process as dominated by the environment explains the speed of the folding process and its reversibility. The distribution of the Phi_e and Psi_e angles obtained for the non-redundant protein database shows the presence of seven local maxima who have been assigned codes A through G. The C code represents the helical form, the G code is the left-handed helix, the E code is the beta structure, and the F code represents the structure of the end of beta structure segment. The interesting state is represented by the D code that connects the helical region to the beta structure region.

The early intermediate structures of the proteins in question will be expressed in the codes given above. With the help of the determined Phi_e and Psi_e values, it is possible to construct a partially unfolded structure determined due to the optimal arrangement of the planes of peptide bonds.

4.4. Tools and Software

3D images of the protein structures were rendered with PyMOL [66] while data charts were plotted using Matplotlib library [67]. Online calculations of fuzzy oil drop hydrophobicity profiles and structural codes are available at http://fod.cm-uj.krakow.pl web server.

5. Conclusions

In the present study, significant participation of the arrangement of hydrophobicity distribution compatible with that present in the spherical micelle and expressed by means of 3D Gauss function in fast and ultrafast folding as well as downhill proteins were demonstrated [68]. The reversibility of this process results from the presence of the seed of the hydrophobic core in the early intermediate structure, which is the partially folded structure. In the present work, the participation of hydrophobic interactions in the process of shaping the structure of proteins has been expressed by means of a mathematical model [42,69]. The fuzzy oil drop model is nothing more than the mathematical expression of the commonly observed hydrophobic core generation. The model allows quantitative description and comparison of the status of different protein molecules.

The practical use of the fuzzy oil drop model allows for the quantitative assessment of the status of a given protein, enabling comparative analysis. The higher the RD value, the lower the stability guaranteed by the presence of the hydrophobic core. For proteins with low RD values, a one-step unfolding process is predicted. The low value of RD observed for most domains treated as individual structural units (3D Gaussian function determined for the domain) suggests their spontaneous folding as an expression of the influence of the external field, which is the water environment, directing this process towards the generation of a hydrophobic core.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

FOD	Fuzzy oil drop
PDB	Protein data bank
RD	Relative distance

References

1. Dill, K.A.; Maccallum, J.L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042–1046. [CrossRef] [PubMed]
2. Muñoz, V. Conformational Dynamics and Ensembles in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395–412. [CrossRef]
3. Schmeing, T.M.; Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nat. Cell Biol.* **2009**, *461*, 1234–1242. [CrossRef]
4. Ellis, R.J.; Van Der Vies, S.M. Molecular Chaperones. *Annu. Rev. Biochem.* **1991**, *60*, 321–347. [CrossRef] [PubMed]
5. Ellis, J. Proteins as molecular chaperones. *Nat. Cell Biol.* **1987**, *328*, 378–379. [CrossRef] [PubMed]
6. Fernández-Fernández, M.R.; Valpuesta, J.M. Hsp70 chaperone: A master player in protein homeostasis. *F1000Research* **2018**, *7*, 1497. [CrossRef] [PubMed]
7. Mayer, M.P.; Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **2005**, *62*, 670–684. [CrossRef]
8. Gruber, R.; Horovitz, A. Allosteric Mechanisms in Chaperonin Machines. *Chem. Rev.* **2016**, *116*, 6588–6606. [CrossRef] [PubMed]
9. Oldfield, C.J.; Dunker, A.K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **2014**, *83*, 553–584. [CrossRef]
10. Zwanzig, R.; Szabó, A.; Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 20–22. [CrossRef]
11. Cavalli, A.; Haberthür, U.; Paci, E.; Caflisch, A. Fast protein folding on downhill energy landscape. *Protein Sci.* **2003**, *12*, 1801–1803. [CrossRef] [PubMed]
12. Creighton, T.E.; Darby, N.J.; Kemmink, J. The roles of partly folded intermediates in protein folding. *FASEB J.* **1996**, *10*, 110–118. [CrossRef]
13. Gelman, H.; Gruebele, M. Fast protein folding kinetics. *Q. Rev. Biophys.* **2014**, *47*, 95–142. [CrossRef] [PubMed]
14. Eaton, W.A.; Muñoz, V.; Hagen, S.J.; Jas, G.S.; Lapidus, L.J.; Henry, E.R.; Hofrichter, J. Fast Kinetics and Mechanisms in Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 327–359. [CrossRef] [PubMed]
15. Baldwin, R.L. The Search for Folding Intermediates and the Mechanism of Protein Folding. *Annu. Rev. Biophys.* **2008**, *37*, 1–21. [CrossRef] [PubMed]
16. Creighton, T.E. How important is the molten globule for correct protein folding? *Trends Biochem. Sci.* **1997**, *22*, 6–10. [CrossRef]
17. Kuwajima, K. The Molten Globule, and Two-State vs. Non-Two-State Folding of Globular Proteins. *Biomolecules* **2020**, *10*, 407. [CrossRef]
18. Sanchez-Medina, C.; Sekhar, A.; Vallurupalli, P.; Cerminara, M.; Muñoz, V.; Kay, L.E. Probing the Free Energy Landscape of the Fast-Folding gpW Protein by Relaxation Dispersion NMR. *J. Am. Chem. Soc.* **2014**, *136*, 7444–7451. [CrossRef]
19. Campos, L.A.; Liu, J.; Wang, X.; Ramanathan, R.; English, D.S.; Muñoz, V. A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy. *Nat. Methods* **2011**, *8*, 143–146. [CrossRef]
20. Ramanathan, R.; Muñoz, V. A Method for Extracting the Free Energy Surface and Conformational Dynamics of Fast-Folding Proteins from Single Molecule Photon Trajectories. *J. Phys. Chem. B* **2015**, *119*, 7944–7956. [CrossRef]

21. Wang, Z.; Campos, L.A.; Muñoz, V. Single-Molecule Fluorescence Studies of Fast Protein Folding. In *Methods in Enzymology*; Elsevier BV: Amsterdam, The Netherlands, 2016; Volume 581, pp. 417–459. [CrossRef]

22. Sadqi, M.; Lapidus, L.J.; Muñoz, V. How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12117–12122. [CrossRef] [PubMed]

23. Oliva, F.Y.; Muñoz, V. A Simple Thermodynamic Test to Discriminate between Two-State and Downhill Folding. *J. Am. Chem. Soc.* **2004**, *126*, 8596–8597. [CrossRef] [PubMed]

24. Doshi, U.R.; Muñoz, V. The Principles of α-Helix Formation: Explaining Complex Kinetics with Nucleation−Elongation Theory. *J. Phys. Chem. B* **2004**, *108*, 8497–8506. [CrossRef]

25. Naganathan, A.N.; Muñoz, V. Scaling of Folding Times with Protein Size. *J. Am. Chem. Soc.* **2005**, *127*, 480–481. [CrossRef] [PubMed]

26. Akmal, A.; Muñoz, V. The nature of the free energy barriers to two-state folding. *Proteins Struct. Funct. Bioinform.* **2004**, *57*, 142–152. [CrossRef] [PubMed]

27. Naganathan, A.N.; Li, P.; Perez-Jimenez, R.; Sanchez-Ruiz, J.M.; Muñoz, V. Navigating the Downhill Protein Folding Regime via Structural Homologues. *J. Am. Chem. Soc.* **2010**, *132*, 11183–11190. [CrossRef] [PubMed]

28. Muñoz, V.; Thompson, P.A.; Hofrichter, J.; Eaton, W.A. Folding dynamics and mechanism of β-hairpin formation. *Nat. Cell Biol.* **1997**, *390*, 196–199. [CrossRef] [PubMed]

29. Munoz, V.; Henry, E.; Hofrichter, J.; Eaton, W.A. A statistical mechanical model for -hairpin kinetics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5872–5879. [CrossRef] [PubMed]

30. Muñoz, V.; Cerminara, M. When fast is better: Protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem. J.* **2016**, *473*, 2545–2559. [CrossRef] [PubMed]

31. Banach, M.; Konieczny, L.; Roterman, I. Protein-protein interaction encoded as an exposure of hydrophobic residues on the surface. In *From Globular Proteins to Amyloids*; Elsevier BV: Amsterdam, The Netherlands, 2020; pp. 79–89. [CrossRef]

32. Banach, M.; Konieczny, L.; Roterman, I. Ligand binding cavity encoded as a local hydrophobicity deficiency. In *From Globular Proteins to Amyloids*; Elsevier BV: Amsterdam, The Netherlands, 2020; pp. 91–93. [CrossRef]

33. Banach, M.; Konieczny, L.; Roterman, I. Why do antifreeze proteins require a solenoid? *Biochimie* **2018**, *144*, 74–84. [CrossRef] [PubMed]

34. Kalinowska, B.; Banach, M.; Król, M.; Irena, R. Application of Divergence Entropy to Characterize the Structure of the Hydrophobic Core in DNA Interacting Proteins. *Entropy* **2015**, *17*, 1477–1507. [CrossRef]

35. Kalinowska, B.; Banach, M.; Wiśniowski, Z.; Konieczny, L.; Irena, R. Is the hydrophobic core a universal structural element in proteins? *J. Mol. Model.* **2017**, *23*, 205. [CrossRef] [PubMed]

36. Banach, M.; Fabian, P.; Stapor, K.; Król, M.; Irena, R. Structure of the Hydrophobic Core Determines the 3D Protein Structure—Verification by Single Mutation Proteins. *Biomolecules* **2020**, *10*, 767. [CrossRef] [PubMed]

37. Roterman, I. The geometrical analysis of peptide backbone structure and its local deformations. *Biochimie* **1995**, *77*, 204–216. [CrossRef]

38. Roterman, I. Modelling the Optimal Simulation Path in the Peptide Chain Folding–Studies Based on Geometry of Alanine Heptapeptide. *J. Theor. Biol.* **1995**, *177*, 283–288. [CrossRef] [PubMed]

39. Fabian, P.; Stapor, K.; Irena, R. Model of Early Stage Intermediate in Respect to Its Final Structure. *Biomolecules* **2019**, *9*, 866. [CrossRef] [PubMed]

40. Berman, H.M.; Westbrook, J.D.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]

41. Hutchinson, E.G.; Thornton, J.M. PROMOTIF-A program to identify and analyze structural motifs in proteins. *Protein Sci.* **1996**, *5*, 212–220. [CrossRef]

42. Konieczny, L.; Roterman, I. Introduction. In *From Globular Proteins to Amyloids*; Elsevier BV: Amsterdam, The Netherlands, 2020; pp. xiii–xx. [CrossRef]

43. 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. *J. Theor. Biol.* **2014**, *359*, 6–17. [CrossRef]

44. Dułak, D.; Gadzała, M.; Stapor, K.; Fabian, P.; Konieczny, L.; Roterman, I. Folding with active participation of water. In *From Globular Proteins to Amyloids*; Elsevier BV: Amsterdam, The Netherlands, 2020; pp. 13–26. [CrossRef]

45. Muñoz, V.; A Campos, L.; Sadqi, M. Limited cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **2016**, *36*, 58–66. [CrossRef]

46. Salapa, K.; Kalinowska, B.; Jadczyk, T.; Roterman, I. Measurement of Hydrophobicity Distribution in Proteins—Non-redundant Protein Data Bank. *Bio-Algorithms Med.-Syst.* **2012**, *8*, 327–338. [CrossRef]

47. Dantas, G.; Kuhlman, B.; Callender, D.; Wong, M.; Baker, D. A Large Scale Test of Computational Protein Design: Folding and Stability of Nine Completely Redesigned Globular Proteins. *J. Mol. Biol.* **2003**, *332*, 449–460. [CrossRef]

48. Banach, M.; Roterman, I. Solenoid—An amyloid under control. In *From Globular Proteins to Amyloids*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 95–115. [CrossRef]

49. He, Y.; Chen, Y.; Alexander, P.A.; Bryan, P.N.; Orban, J. Mutational Tipping Points for Switching Protein Folds and Functions. *Structure* **2012**, *20*, 283–291. [CrossRef] [PubMed]

50. Fabian, P.; Banach, M.; Stapor, K.; Król, M.; Ptak-Kaczor, M.; Irena, R. The Structure of Amyloid Versus the Structure of Globular Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 4683. [CrossRef]

51. Sadqi, M.; Fushman, D.; Muñoz, V. Atom-by-atom analysis of global downhill protein folding. *Nat. Cell Biol.* **2006**, *442*, 317–321. [CrossRef]

52. Sborgi, L.; Verma, A.; Muñoz, V.; De Alba, E. Revisiting the NMR Structure of the Ultrafast Downhill Folding Protein gpW from Bacteriophage λ. *PLoS ONE* **2011**, *6*, e26409. [CrossRef]

53. Sborgi, L.; Verma, A.; Muñoz, V.; De Alba, E. High Resolution NMR Structure of gpW (W Protein of Bacteriophage Lambda) at Acidic pH. 2011. Available online: https://www.rcsb.org/structure/2l6r (accessed on 7 October 2020). [CrossRef]

54. Ferguson, N.; Sharpe, T.D.; Schartau, P.J.; Sato, S.; Allen, M.D.; Johnson, C.M.; Rutherford, T.J.; Fersht, A.R. Ultra-fast Barrier-limited Folding in the Peripheral Subunit-binding Domain Family. *J. Mol. Biol.* **2005**, *353*, 427–446. [CrossRef]

55. Robien, M.A.; Clore, G.M.; Omichinski, J.G.; Perham, R.N.; Appella, E.; Sakaguchi, K.; Gronenborn, A.M. Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli. Biochemistry* **1992**, *31*, 3463–3471. [CrossRef]

56. Neuweiler, H.; Sharpe, T.D.; Rutherford, T.J.; Johnson, C.M.; Allen, M.D.; Ferguson, N.; Fersht, A.R. The Folding Mechanism of BBL: Plasticity of Transition-State Structure Observed within an Ultrafast Folding Protein Family. *J. Mol. Biol.* **2009**, *390*, 1060–1073. [CrossRef]

57. Sicheri, F.; Yang, D.S.C. Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nat. Cell Biol.* **1995**, *375*, 427–431. [CrossRef]

58. Artymiuk, P.; Blake, C. Refinement of human lysozyme at 1.5 Å resolution analysis of non-bonded and hydrogen-bond interactions. *J. Mol. Biol.* **1981**, *152*, 737–762. [CrossRef]

59. Banach, M.; Konieczny, L.; Roterman, I. Fuzzy Oil Drop Model Application—From Globular Proteins to Amyloids. In *The Functional Role of Critical Dynamics in Neural Systems*; Springer Science and Business Media LLC.: Berlin, Germany, 2018; pp. 639–658.

60. Fabian, P.; Stapor, K.; Banach, M.; Ptak-Kaczor, M.; Król, M.; Irena, R. Alternative Hydrophobic Core in Proteins—The Effect of Specific Synergy. *Symmetry* **2020**, *12*, 273. [CrossRef]

61. Jurkowski, W.; Baster, Z.; Dułak, D.; Roterman-Konieczna, I. The early-stage intermediate. In *Protein Folding in Silico*; Elsevier BV: Amsterdam, The Netherlands, 2012; pp. 1–20, ISBN-13: 9781907568176.

62. Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1–63. [CrossRef] [PubMed]

63. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59–107. [CrossRef]

64. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

65. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss-function-based model of hydrophobicity density in proteins. *In Silico Biol.* **2006**, *6*, 15–22. [PubMed]

66. *The PyMOL Molecular Graphics System*; Version 2.0; Schrödinger, LLC.: New York, NY, USA, 2015.

67. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

68. Roterman, I.; Konieczny, L.; Jurkowski, W.; Prymula, K.; Banach, M. Two-intermediate model to characterize the structure of fast-folding proteins. *J. Theor. Biol.* **2011**, *283*, 60–70. [CrossRef]

69. Mok, K.H.; Kuhn, L.T.; Goez, M.; Day, I.J.; Lin, J.C.; Andersen, N.H.; Hore, P.J. A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nat. Cell Biol.* **2007**, *447*, 106–109. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.