Journal of Theoretical Biology 359 (2014) 6–17

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function

Mateusz Banach [a,b], Leszek Konieczny [c], Irena Roterman [a,*]

[a] Department of Bioinformatics and Telemedicine – Jagiellonian University – Medical College, Krakow, Poland
[b] Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland
[c] Chair of Medical Chemistry – Jagiellonian University – Medical College, Krakow, Poland

HIGHLIGHTS

- Differentiation of hydrophobic core structure of IgG-like domains is presented.
- Relation between hydrophobic core structure and biological function is shown.
- Potential amyloidogenic character of immunoglobulin-like domains is presented.
- The fuzzy-oil-drop model is shown to characterize the structure of domains.
- Fuzzy-oil-drop model is potentially applicable for protein structure prediction.

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form
25 April 2014
Accepted 5 May 2014
Available online 21 May 2014

Keywords:
Water environment
Hydrophobic core
Amyloidosis
Thansthytetin
Immunoglobulin-like fold

ABSTRACT

In this paper we show that the fuzzy oil drop model represents a general framework for describing the generation of hydrophobic cores in proteins and thus provides insight into the influence of the water environment upon protein structure and stability. The model has been successfully applied in the study of a wide range of proteins, however this paper focuses specifically on domains representing immunoglobulin-like folds. Here we provide evidence that immunoglobulin-like domains, despite being structurally similar, differ with respect to their participation in the generation of hydrophobic core. It is shown that β-structural fragments in β-barrels participate in hydrophobic core formation in a highly differentiated manner. Quantitatively measured participation in core formation helps explain the variable stability of proteins and is shown to be related to their biological properties. This also includes the known tendency of immunoglobulin domains to form amyloids, as shown using transthyretin to reveal the clear relation between amyloidogenic properties and structural characteristics based on the fuzzy oil drop model.

1. Introduction

The fuzzy oil drop model suggests that the formation of a hydrophobic core in proteins is more closely related to their water environment than to pair-wise interactions, which only acknowledge local residue neighborhoods. The hydrophobic core is the result of a unified cooperative effect of ordering on the molecular level and seems less related to individual pair-wise interactions between residues. Local deformations in the hydrophobic core are shown to be related to the biological function of proteins.

Despite half a century of intense research in the field of protein folding, many unresolved questions remain (Dill and MacCallum, 2012). One of those questions pertains to the influence of the water environment upon the folding process. The presence of the hydrophobic core and its role in ensuring tertiary structural stabilization is well established (Hamill et al., 2000; Nelson and Grishin, 2006; Lazar and Handel, 1998; Yang et al., 2012; Koide et al., 2000; Ryu et al., 1996; Chen and Stites, 1959). The reason we refer to a fairly old work (Kauzmann 1959) is to highlight the commonalities between the model proposed in this publication and our own fuzzy oil drop model. Despite the large number of available models (Arakawa et al., 2013; Li et al., 2013; Sirk et al., 2013; Huggins, 2012a, 2012b; Sala et al., 2010, Liu et al., 2012), a general model explaining the influence of water on proteins – a critical factor for cellular mechanisms – and the corresponding

* Corresponding author.
E-mail address: myroterm@cyf-kr.edu.pl (I. Roterman).

http://dx.doi.org/10.1016/j.jtbi.2014.05.007

changes in the structure of these proteins is still missing. The majority of models, implemented as IT solutions, take hydrophobic interactions into account when predicting protein-structure relationships; however the presence of water is modeled as pair-wise interactions between atoms, where each atom belonging to the protein molecule may interact with individual water molecules (in a two- or three-atom model). These interactions are expressed and measured using electrostatic and Leonard-Jones potentials which promote the exposure of hydrophilic residues on the surfaces of the proteins, where contact with water is energetically advantageous. Consequently, it is expected that hydrophobic residues will be internalized, forming the so-called hydrophobic core. Our fuzzy oil drop model views hydrophobic core formation as a holistic process – unlike pair-wise optimization, which focuses on local neighborhoods of individual residues.

In this work the fuzzy oil drop model was applied to protein domains representing immunoglobulin-like folds. Interpretation (including quantitative interpretation) of the obtained results suggests significant differentiation in the status of individual β-structural fragments despite their substantial structural similarity. As it turns out, such differentiation is associated with biological function and with changes which those domains undergo e.g. amyloidogenesis. Transthyretin is used here as an example of an amyloidogenic protein.

2. Materials and methods

2.1. The fuzzy oil drop model

The model explored in this work adopts a global approach to describing the influence of water on a protein, in that both high hydrophobicity in the central part of the protein body and exposure of the hydrophilic residues on its surface is ensured. According to the presented model the hydrophobic core is defined as a concentration of hydrophobicity in the central part of the protein body, encapsulated by a hydrophobicity gradient which reaches near-zero values on the surface. The presence of a hydrophobic core is identified if the observed hydrophobicity density distribution (resulting from inter-residual interactions) remains in accordance with the assumed "idealized" distribution, which predicts a high concentration of hydrophobicity at the center of the protein. Thus, the following two distributions are distinguished: observed (O) and theoretical (T).

2.2. Observed hydrophobicity density

If we assume that the distribution of hydrophobicity follows from the interactions between side chains (the hydrophobicity value, whether experimentally determined or numerically calculated, is ascribed to the effective atom placed at the geometric center of each side chain), it can be expressed using Levitt's formula (Levitt, 1976)

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^{N} (H_i^r + H_j^r) \begin{cases} \left[1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right] & \text{for } r_{ij} \le c \\ 0 \text{ for } r_{ij} > c \end{cases} \tag{1}$$

In Eq. (1), the j index denotes the effective atom (the averaged position of the side chain of the j-th residue). Ho_j is the aggregate value describing the interactions with the neighboring residues (labeled with the i index) at distances not greater than 9 Å (this distance – c – is treated as the cutoff value for the hydrophobic interactions, following the original model Levitt, 1976). Applying a cutoff value implies that hydrophobic interactions are considered local and depend on the location of each residue. This function is empirically determined and, according to Levitt (1976), expresses the strength of hydrophobic interactions. H_i^r and H_j^r represent the hydrophobicity parameter (constant for each residue) ascribed to each amino acid using a predetermined scale, which can be arbitrary (in our study, the relevant scale is derived from Brylinski et al. (2007)). r_{ij} is the distance between the i-th and the j-th residue while N is the total number of residues in the chain.

2.3. Theoretical (idealized) hydrophobicity density distribution

The resulting empirical distribution can be compared to a corresponding theoretical distribution which reflects the presence of a hydrophobic core. Distribution is understood as follows: imagine a very small detector moving from a selected site within the protein molecule to an antipodal site, passing through the center of the molecule and measuring hydrophobicity density at each point along its path. This detector would be expected to find the greatest hydrophobicity density at the center of the molecule, with density values decreasing along with distance from the center and approaching 0 on the surface. Such conditions are most closely approximated by a 3D Gaussian.

$$\tilde{H}t_j = \frac{1}{\tilde{H}t_{sum}} \exp\left(-\frac{(x_j-\bar{x})^2}{2\sigma_x^2}\right) \exp\left(-\frac{(y_j-\bar{y})^2}{2\sigma_y^2}\right) \exp\left(-\frac{(z_j-\bar{z})^2}{2\sigma_z^2}\right) \tag{2}$$

The Gaussian expresses the distribution of probability in an ellipsoid capsule the size of which is determined by the values of σ along each cardinal direction. If we fine-tune these σ values in such a way that the Gaussian envelope completely encapsulates the molecule, then the value of the function will express the expected (theoretical) distribution of hydrophobicity at each point within the protein body. The parameters $\bar{x}, \bar{y}, \bar{z}$ reflect the placement of the center of the ellipsoid – at the origin of the coordinate system all three values become equal to 0. σ values are calculated as 1/3 of the greatest distance between an effective atom and the origin of the system once the molecule has been oriented in such a way that its greatest spatial extension coincides with a system axis (for each axis separately).

The $1/Ht_{sum}$ and $1/Ho_{sum}$ coefficients ensure normalization of both distributions (empirical and theoretical) and therefore enable comparative analysis. This function was applied in the structural analysis and folding simulations described in Konieczny et al. (2006).

2.4. Quantitative analysis of the hydrophobic core status

Each protein molecule can be characterized by two distributions. Quantitative similarities (or differences) between the theoretical and observed profiles can be measured using Kullback–Leibler's (divergence or distance) entropy, as given in Kullback and Leibler (1951)

$$D_{KL}(p|p^0) = \sum_{i=1}^{N} p_i \log_2(p_i/p_i^0) \tag{3}$$

The value of D_{KL} is proportional to the distance between the empirical (p) and target (p^0) distributions. In our case the target distribution is given by the 3D Gaussian function. According to its definition, D_{KL} is a measure of entropy and its value cannot be interpreted in absolute terms. Therefore an independent target distribution is required in addition to the 3D Gaussian distribution: one in which there is no concentration of hydrophobicity near the center of the molecule. In this so-called unified distribution (R), each residue is assigned an equal hydrophobicity density value of $1/N$, where N is the number of residues in the polypeptide chain.

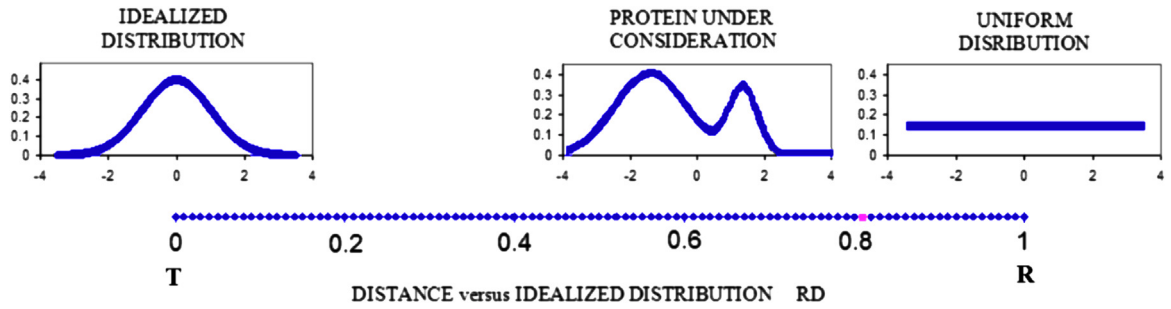

Fig. 1. Graphical representation of the comparison between the theoretical and observed distributions in a virtual molecule to visualize all introduced parameters. Shown along the horizontal axis is the relative distance (RD; Eq. (6)) between the 3D Gaussian distribution (idealized hydrophobicity core, represented by the leftmost panel) and the actual hydrophobicity profile, which differ in comparison to the idealized one. The right-hand panel presents a uniform distribution, i.e. one that does not include a hydrophobic core. The pink dot represents the RD value of the virtual molecule, shown in the central panel.

For simplicity's sake we introduce the following notation to express quantities calculated according to Eq. (3):

$$O/T = \sum_{i=1}^{N} O_i \log_2 O_i/T_i \tag{4}$$

$$O/R = \sum_{i=1}^{N} O_i \log_2 O_i/R_i \tag{5}$$

where O/T expresses the distance between the observed and theoretical (idealized) distributions, and O/R is the corresponding distance between the observed distribution and a distribution in which each residue carries the same hydrophobicity density value (no core at all). Furthermore, we introduce a binary model where $O/T < O/R$ indicates the presence of a hydrophobic core (i.e. close correspondence between the theoretical and observed distributions). To avoid using two parameters we introduce one factor, called relative distance (RD), which is defined as follows:

$$\mathrm{RD} = \frac{O/T}{O/T + O/R} \tag{6}$$

Clearly, the lower the value of RD the more closely a given protein approximates the theoretical model (Fig. 1).

Two boundary conditions are of key importance to our model. One boundary condition is the idealized hydrophobic core structure in the shape of a Gaussian function (distance versus T is equal to 0) while the opposite condition reflect complete lack of hydrophobicity concentration (right panel). The distance between both boundaries is defined as 1. Real proteins exhibit partially distorted hydrophobicity distributions, as shown in the central panel (hypothetical distribution) – in this case the distance versus the idealized distribution is equal to 0.8. The hypothetical distribution, if found in any actual protein, would suggest that this protein does not contain a well defined hydrophobic core. The intuitive accordance criteria (idealized and observed hydrophobicity density distribution) can be denoted as $O/T < O/R$, meaning that $\mathrm{RD} < 0.5$ (the observed distribution is "closer" to the idealized one).

2.5. Definition of structural units in the fuzzy oil drop model

The structural unit in the fuzzy oil drop model can be represented by a multi-chain protein complex (where applicable), an individual chain or a specific domain. This definition is important for theoretical distributions where the ellipsoid is expected to encapsulate the complete unit.

O/T and O/R values can also be calculated for fragments of polypeptide chains. This form of calculation requires prior normalization of T and O values for selected fragments. Calculating O/T and O/R values for fragments reveals their local contribution (regarding the fuzzy oil drop model) to the status of a given structural unit. This fragment-oriented calculation of O/T and O/R was performed for all β-structural fragments and loops present in selected proteins.

Transthyretin (1DVQ – Klabunde et al., 2000) – known for its highly amyloidogenic character – was additionally studied in the context of its other-than-secondary structural fragments in order to analyze their contribution to hydrophobic core formation. This additional calculation of O/T and O/R was performed using a moving 20-aa frame, in a progressive manner (overlapping system). The Overlapped Open Reading Frame (OORF) calculation for transthyretin was performed as follows: the initial window (20 aa), starting with residues numbered 1–20, was used for O/T and O/R calculations. Subsequently, the second frame was defined as comprising residues numbered 2–21 and so forth reaching the last residue in a chain. Such calculation can be used to characterize the entire protein instead of focusing on particular secondary fragments.

2.6. Presentations of structures

Identification of secondary fragments was performed using the data given in PDBSum, as well as the schemes listed in Laskowski (2009). The VMD program (Humphrey et al., 1996; http://www.ks.uiuc.edu/Research/vmd/) was used for 3D structural presentation of proteins. Description of super-secondary structures follows the CATH classification Orengo (1997).

3. Results

3.1. Interpretation of hydrophobicity profiles

As a real-world example we have chosen the p50 subunit of the human transcription factor NF-kappa B, bound as a homodimer to DNA – 1SVC-P domain (249–353) (Müller et al. 1995). Visual inspection of both profiles (theoretical and observed) (Fig. 2) reveals regions in which they substantially differ from each other. In addition, the DNA-binding residues have been highlighted since – as shown in Banach et al. (2013a), Marchewka et al. (2013), – the presence of a ligand might significantly affect the shape of the observed hydrophobicity distribution profile and provide hints enabling identification of complexation areas (Alejster et al., 2013). However, interaction with DNA in the protein under consideration does not seem to cause discrepancies between both profiles.

For the 1SVC-P immunoglobulin-like domain (249–353) the corresponding RD value is 0.43, suggesting the presence of a well-defined hydrophobic core in this domain.

To what extend is it possible to find a protein where the observed hydrophobicity density distribution closely matches

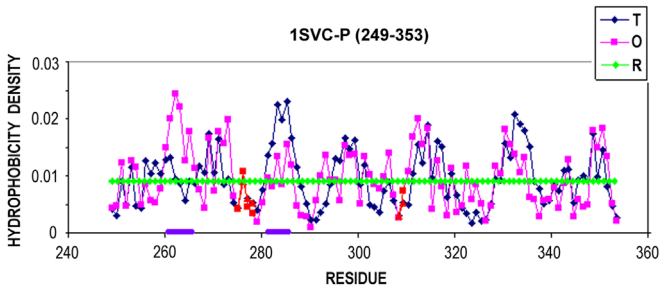

Fig. 2. Hydrophobicity density profiles: theoretical – calculated according to Eq. (2). (dark blue; T) and observed – calculated according to Eq. (1) (pink; O) for the 1SVC-P immunoglobulin-like domain (249–353). Residues marked in red are involved in binding DNA. The bars on the horizontal axis highlight fragments where both distributions significantly differ with respect to the idealized distribution. The horizontal gray line represents the unified distribution (R).

the idealized distribution? An 'ideal protein' containing a hydrophobic core in perfect accordance with the theoretical profile (that is, with all hydrophobic residues concentrated near the center and all hydrophilic residues exposed on the surface) would possess two key characteristics: unsurprisingly, it would be highly water-soluble, but it would also be incapable of interacting with other molecules, given its high preference for contact with water.

Our analysis of numerous proteins reveals that most exhibit localized discrepancies in relation to the idealized distribution of hydrophobicity (Banach et al., 2012a; Marchewka et al., 2013). Moreover, areas where such discrepancies are concentrated often mediate the protein's biological activity (Banach et al., 2012b; Marchewka et al., 2013). For example, ligand binding pockets are typically associated with hydrophobicity deficiencies, whereas protein complexation sites frequently exhibit excess hydrophobicity (Banach et al., 2012c; Marchewka et al., 2013). Nevertheless, several proteins that follow the idealized model have been found in databases – these include antifreeze (class II) proteins (Banach et al., 2012d) which counteract the growth of ice crystals and prevent them from damaging cellular structures. The biological role of antifreeze proteins requires them to be highly water-soluble. In addition, they should not form clusters but instead need to be uniformly distributed in the aqueous environment, which explains their good accordance with fuzzy oil drop model.

Another group that has been found to be well described by the fuzzy oil drop model comprises the so-called 'fast-folding' (or, downhill) proteins (Roterman et al., 2011). In this case, the experimentally observed high reversibility of the folding process suggests that the influence of water is sufficient to guide folding, ensuring aggregation of hydrophobic residues near the center of the protein along with simultaneous exposure of hydrophilic residues on its surface (Banach et al., 2012c; Roterman et al., 2011). Indeed, proteins that follow this theoretical model are readily identifiable in the Protein Data Bank, although an in-depth search has to take into account the specific structural unit for which accordance is assessed (Roterman et al., 2011). Possible units that fulfill this criterion include individual domains, entire chains and multi-protein complexes. Depending on the size of the unit, the encapsulating 'drop' must be suitably defined (chain, domain and complex). Good accordance with the model is most frequently observed in individual domains.

3.2. General characteristics of immunoglobulin-like domains

For the present analysis we have selected immunoglobulin-like protein domains from the set discussed in Prudhomme and Chomilier (2009). Table. 1 lists the O/T and O/R values for 16 domains and β-structural fragments that make up the super-secondary structures known as β-barrels.

The summary presented in Table. 1 (positions listed in bold) and the graphical presentation of the results in Fig. 3 (blue fragments) suggest that, among the analyzed proteins, only 1TIT (Rico et al., 2013) exhibits a hydrophobicity density distribution that is in accordance with the corresponding idealized distribution. 1TLK (Holden et al., 1992) contains the lowest number of fragments accordant with the theoretical distribution, while the highest number is found in 1TEN (Rico et al., 2013) and 2FBJ-L (Laskowski, 2009). The second general conclusion based on the presented results is that common geometric properties (for immunoglobulin-like domains, this mainly includes the β- sandwich – 2.60.40.10; classification according to CATH (Orengo et al., 1997) do not necessarily translate into similar hydrophobic core structures. As shown in Fig. 3, different β-structural fragments (parts of β-sandwiches) can result in hydrophobicity density distributions which closely approximate theoretical values ($RD < 0.5$).

3.3. Detailed characteristics of selected proteins

Below we discuss the specific properties of each sample protein (see Table 2).

3.3.1. TIT – Immunoglobulin-like module from the titin I-band, the extensible components of muscle elasticity

The titin region, located in the sarcomere I-band, is believed to play a major role in the extensibility and passive elasticity of muscle tissue (Improta et al., 1996). Molecular dynamics simulations presented in Rico et al. (2013) reveal that the protein undergoes a two-stage melting event in the absence of external factors. These stages involve the loosening of interactions between the fragments designated A (N-terminal β-fragment) and B (19-25), and between A′ (N-terminal fragment) and G (82-87). The loosening of fragment F (69-90) has been observed in a "pulled" molecular dynamics simulation (Rico et al., 2013). Based on our fuzzy oil drop analysis (Fig. 3 and Table 1), it appears that the entire domain is highly consistent with the theoretical hydrophobicity density model. This observation can be interpreted as structural stabilization due to accordance with theoretical hydrophobic density distribution for the entire chain. Since the hydrophobic core stabilizes the tertiary structure, in this case the tertiary structure becomes highly stable. These two observations suggest that any deformation of the domain structure occurs in a progressive manner and that retaining high consistency with the fuzzy oil drop model enables 1TIT to quickly revert to its original state in the absence of external forces, as experimentally confirmed in (Perrakis et al., 1994; Lu and Schulten, 2000).

3.3.2. FAB-H and 6FAB-L Mus musculus – murine anti-p-azophenylarsonate FAB fragment.

Light chain (L) – IgG1-kappa 36-71 FAB; heavy chain (H) – IgG1-kappa 36-71 Fab (Leahy et al., 1992); 2FBJ-H and 2FBJ-L together – galactan-binding immunoglobulin j539 (Suh et al., 1986); light chain (L) – IgA-kappa j539 FAB, heavy chain (H) – IgA-kappa j539 FAB. Only V domains are discussed here. Sequence similarity between the two chains is 59% for the light chains and 42% for the heavy chains. According to our analysis (Fig. 3 and Table. 1), the status (with respect to the fuzzy oil drop model) of the so-called upper core (Ewert et al., 2004; Ewert et al., 2003) in 6FAB-L is similar to that of 2FBJ-L for all β-fragments, whereas in the so-called lower core the C and F chains differ. Similarly, the upper cores of the light chains are similar whereas the lower cores differ (especially fragment F, i.e. 92-102 in 6FAB-H and 92-100 in

Table 1
O/*T*, *O*/*R* and RD values for the proteins investigated here. The RD values for individual ß-structural fragments of these proteins (composed of specific amino acid residue regions) are given. Identification of the ß-structural fragments is consistent with the PDBSum database (Laskowski, 2009). Fragments for which the observed hydrophobicity distribution remains in accordance with the theoretical distribution are shown in bold ($O/T < O/R$ and RD < 0.5). The biological properties of different groups of proteins are indicated in subheaders.

PDB-ID	Domain ß-structural fragments			
	Residues Immunoglobulins	*O*/*T*	*O*/*R*	RD
6FAB-H	1–121	0.188	0.129	0.59
	67–73	**0.012**	**0.024**	**0.33**
	77–84	0.050	0.046	0.52
	18–25	0.158	0.111	0.58
	3–7	**0.075**	**0.079**	**0.48**
	10–12	0.103	0.040	0.72
	115–120	0.168	0.042	0.80
	105–112	0.170	0.095	0.64
	92–102	**0.106**	**0.142**	**0.42**
	33–40	**0.057**	**0.103**	**0.35**
	44–52	**0.054**	**0.131**	**0.29**
	57–61	0.014	0.006	0.70
6FAB-L	1–107	0.138	0.124	0.53
	61–67	**0.028**	**0.029**	**0.49**
	69–76	**0.058**	**0.079**	**0.42**
	19–26	0.087	0.049	0.64
	3–7	**0.041**	**0.061**	**0.40**
	9–13	0.053	0.018	0.74
	102–107	**0.071**	**0.212**	**0.25**
	97–99	**0.008**	**0.041**	**0.17**
	84–91	0.059	0.036	0.62
	32–39	0.141	0.104	0.57
	43–49	**0.060**	**0.115**	**0.34**
	54–56	0.021	0.018	0.53
1FDL-H	1–116	0.152	0.132	0.53
	3–7	0.148	0.137	0.52
	18–25	0.098	0.043	0.69
	76–82	**0.061**	**0.080**	**0.43**
	68–72	**0.033**	**0.148**	**0.18**
	10–12	0.026	0.010	0.72
	110–115	0.152	0.049	0.76
	102–107	**0.033**	**0.087**	**0.27**
	91–99	**0.057**	**0.165**	**0.26**
	33–39	**0.050**	**0.065**	**0.43**
	46–52	**0.071**	**0.110**	**0.39**
	56–60	0.027	0.017	0.61
1FDL-L	**1–108**	**0.152**	**0.137**	**0.47**
	61–67	0.058	0.020	0.74
	69–76	**0.079**	**0.117**	**0.40**
	19–26	0.099	0.051	0.66
	3–7	0.049	0.047	0.51
	9–13	0.057	0.022	0.72
	102–107	**0.100**	**0.141**	**0.41**
	84–91	0.051	0.034	0.60
	32–38	**0.058**	**0.132**	**0.30**
	45–49	**0.035**	**0.113**	**0.24**
	53–55	**0.001**	**0.002**	**0.33**
2FBJ-H	1–118	0.173	0.151	0.53
	67–73	**0.020**	**0.111**	**0.15**
	77–84	**0.032**	**0.050**	**0.39**
	10–12	0.017	0.014	0.55
	18–25	0.124	0.109	0.53
	3–7	**0.035**	**0.063**	**0.36**
	57–61	0.074	0.072	0.51
	44–52	**0.088**	**0.160**	**0.35**
	33–40	**0.042**	**0.060**	**0.41**
	92–100	0.059	0.044	0.57
	104–109	0.089	0.016	0.85
	112–117	0.229	0.086	0.72
2FBJ-L	1–107	0.159	0.121	0.57
	60–66	0.029	0.026	0.53
	68–75	**0.076**	**0.077**	**0.49**
	19–26	0.12	0.034	0.78
	3–7	0.098	0.034	0.74
	9–15	0.062	0.018	0.77
	101–106	**0.087**	**0.148**	**0.37**
	94–98	0.073	0.015	0.83
	83–91	0.101	0.035	0.74
	31–37	0.104	0.083	0.55

Table 1 (*continued*)

PDB-ID	Domain ß-structural fragments			
	Residues Immunoglobulins	*O*/*T*	*O*/*R*	RD
	44–48	**0.103**	**0.188**	**0.35**
	52–54	**0.086**	**0.115**	**0.43**
4FAB-H	1–118	0.153	0.181	0.46
	5–7	0.064	0.007	0.90
	18–24	**0.113**	**0.164**	**0.41**
	79–85	**0.036**	**0.039**	**0.48**
	73–75	**0.008**	**0.082**	**0.09**
	59–63	**0.018**	**0.082**	**0.18**
	46–52	**0.033**	**0.148**	**0.18**
	34–39	**0.020**	**0.043**	**0.32**
	94–99	0.052	0.030	0.63
	112–114	0.022	0.001	0.95
4FAB-L	1–113	0.127	0.124	0.50
	9–13	**0.036**	**0.440**	**0.44**
	107–112	**0.062**	**0.238**	**0.21**
	102–104	**0.008**	**0.023**	**0.26**
	89–96	0.026	0.025	0.51
	37–43	**0.026**	**0.038**	**0.40**
	50–54	**0.006**	**0.037**	**0.14**
	58–60	**0.020**	**0.038**	**0.34**
	66–72	**0.026**	**0.027**	**0.49**
	74–81	**0.050**	**0.087**	**0.36**
	19–25	**0.014**	**0.058**	**0.19**
	5–7	**0.004**	**0.021**	**0.14**
Enzymes				
1CTN	24–130	0.382	0.147	0.72
	94–102	0.129	0.083	0.61
	60–68	0.100	0.074	0.57
	30–32	**0.006**	**0.024**	**0.20**
	40–44	0.128	0.097	0.57
	54–57	0.338	0.121	0.74
	84–91	**0.144**	**0.198**	**0.42**
	75–82	0.081	0.048	0.63
	106–116	0.171	0.133	0.56
	119–123	0.069	0.011	0.86
	126–130	0.150	0.130	0.53
1CTN	24–44,55–130	**0.146**	**0.157**	**0.48**
	94–102	**0.069**	**0.077**	**0.47**
	60–68	0.102	0.065	0.61
	30–32	**0.003**	**0.017**	**0.14**
	40–44	**0.191**	**0.366**	**0.34**
	84–91	0.205	0.180	0.53
	75–82	**0.100**	**0.087**	**0.13**
	106–116	0.130	0.140	0.50
	119–123	0.050	0.013	0.79
	126–130	**0.112**	**0.130**	**0.46**
1CYG	397–491	0.171	0.122	0.58
	404–411	0.329	0.217	0.60
	413–421	0.206	0.088	0.70
	424–432	**0.026**	**0.034**	**0.43**
	438–440	**0.006**	**0.076**	**0.07**
	450–452	0.103	0.039	0.72
	465–467	0.137	0.083	0.61
	476–478	**0.026**	**0.085**	**0.23**
	482–488	**0.036**	**0.057**	**0.39**
DNA/RNA binding				
1ACX	1–107	0.132	0.088	0.60
	59–64	**0.018**	**0.029**	**0.38**
	16–23	**0.036**	**0.080**	**0.31**
	3–7	0.027	0.018	0.60
	49–53	0.074	0.047	0.61
	27–35	**0.098**	**0.120**	**0.45**
	89–93	0.037	0.006	0.86
	40–42	**0.051**	**0.170**	**0.23**
	36–37	0.097	0.045	0.68
	76–83	**0.007**	**0.057**	**0.11**
	66–72	0.077	0.073	0.51
1SVC-P	**249-353**	**0.188**	**0.222**	**0.46**
	252_257	0.168	0.109	0.61
	267_274	**0.159**	**0.172**	**0.48**
	310_316	**0.104**	**0.109**	**0.49**
	305_307	**0.013**	**0.067**	**0.16**

Table 1 (*continued*)

PDB-ID	Domain ß-structural fragments			
	Residues Immunoglobulins	*O*/*T*	*O*/*R*	RD
	293_299	0.144	0.062	0.70
	282_289	**0.033**	**0.217**	**0.13**
	328_335	0.094	0.056	0.63
	346_351	**0.112**	**0.171**	**0.39**
	259_263	0.134	0.089	0.60
Miscellaneous				
1TLK	**33–135**	**0.155**	**0.214**	**0.42**
	40–46	**0.071**	**0.096**	**0.42**
	59–68	**0.129**	**0.137**	**0.48**
	96–103	**0.036**	**0.063**	**0.36**
	87–93	**0.034**	**0.056**	**0.38**
	80–82	**0.17**	**0.035**	**0.32**
	72–78	**0.093**	**0.131**	**0.41**
	111–119	**0.095**	**0.132**	**0.42**
	122–133	**0.059**	**0.188**	**0.24**
	51–55	0.128	0.091	0.58
1TEN	802–891	0.276	0.164	0.63
	807–812	**0.042**	**0.056**	**0.43**
	819–824	0.103	0.061	0.63
	856–860	0.189	0.188	0.50
	847–853	**0.028**	**0.109**	**0.20**
	831–840	0.122	0.038	0.76
	867–877	**0.132**	**0.155**	**0.46**
	880–882	0.141	0.028	0.83
	885–891	0.323	0.162	0.66
1TIT	**1–89**	**0.146**	**0.235**	**0.38**
	11–13	**0.081**	**0.188**	**0.30**
	19-25	**0.080**	**0.189**	**0.30**
	32–36	**0.080**	**0.168**	**0.32**
	46–52	**0.102**	**0.211**	**0.32**
	54–61	**0.037**	**0.262**	**0.12**
	69–71	**0.036**	**0.152**	**0.19**
	73–75	**0.098**	**0.162**	**0.38**
	78–80	**0.214**	**0.322**	**0.40**
	82–87	**0.174**	**0.396**	**0.31**

2FBJ-H). Fragments G (105-112; 115-120 in 6FAB-H, and 104-109; 112-117 in 2FBJ-H respectively), and A′ (9-13 in both proteins) are of special interest due to their participation in immunological signal transduction (Ewert et al., 2003).

The status of the above mentioned fragments with respect to the fuzzy oil drop model is similar. The same is true for fragments A′ (9-12 in 6FAB-L and 9-15 in 2FBJ-L) as well as G (102-107 in 6FAB-L and 101-106 in 2FBJ-L). These two pairs (A′ and G), which link the N-terminal with C-terminal fragments, are also assumed to be responsible for immunological signal transduction (Piekarska et al., 2006).

3.3.3. *1TLK – Elokin, the C-terminal domain of myosin light chain kinase (Meleagris gallopavo)*

1TLK is a calmodulin-binding domain (Holden et al., 1992). This domain bears a striking resemblance to immunoglobulin domains, including the location of the CYS residues participating in SS bond formation; however in 1TLK such SS bond is absent. In contrast to the immunoglobulin domains analyzed here, 1TLK contains a hydrophobic core that is highly accordant with the theoretical expectations, with the exception of one short fragment (51–55) in which the hydrophobicity density distribution diverges from the model (Fig. 3 and Table. 1). This fragment, localized very close to the N-terminus, is quite similar to analogous fragments in immunoglobulin domains. It seems to link the N-terminal and C-terminal fragments. Its low stability (assumed due to discordance versus the hydrophobic core model) may allow structural changes, which may then propagate though the C-terminal fragment to other parts (domains) of the protein. Very similar characteristics can be observed in 1SVC-P (249-353) as well in 1TEN. This suggests some sort of common mechanism which induces poor stability in the N-terminal fragment and links it with C-terminal fragments.

3.3.4. *1TEN – cell adhesion protein, fibronectin type III domain from Homo sapiens*

This extracellular matrix protein (tenascin) was phased using a selenomethionyl protein (Leahy et al., 1992). Our hydrophobic core analysis indicates that in this domain, only three β-structural fragments (out of 8) diverge from the model. For the domain as a whole, RD > 0.5. The 877–879 loop is identified as participating in cell adhesion (Leahy et al., 1992), which correlates with interpretation based on our model. All β-structural fragments preceding and following this loop are recognized as not accordant with the fuzzy oil drop model. This means they are somewhat unstable, enabling the structural changes necessary for interaction with other proteins. Locally the RD for this loop (877–879) is equal to 0.857, expressing local disagreement with the model. Given the local hydrophobicity density excess, this fragment may be recognized as ready for protein complexation, what is in fact the case (Fig. 3).

3.3.5. *1CYG – cyclodextrin glucanotransferase (E.C.2.4.1.19) (CGTase) from the Geobacillus stearothermophilus bacteria (Laskowski, 2009)*

The protein belongs to the previously mentioned class of enzymes; however the domain under consideration is not involved in enzymatic

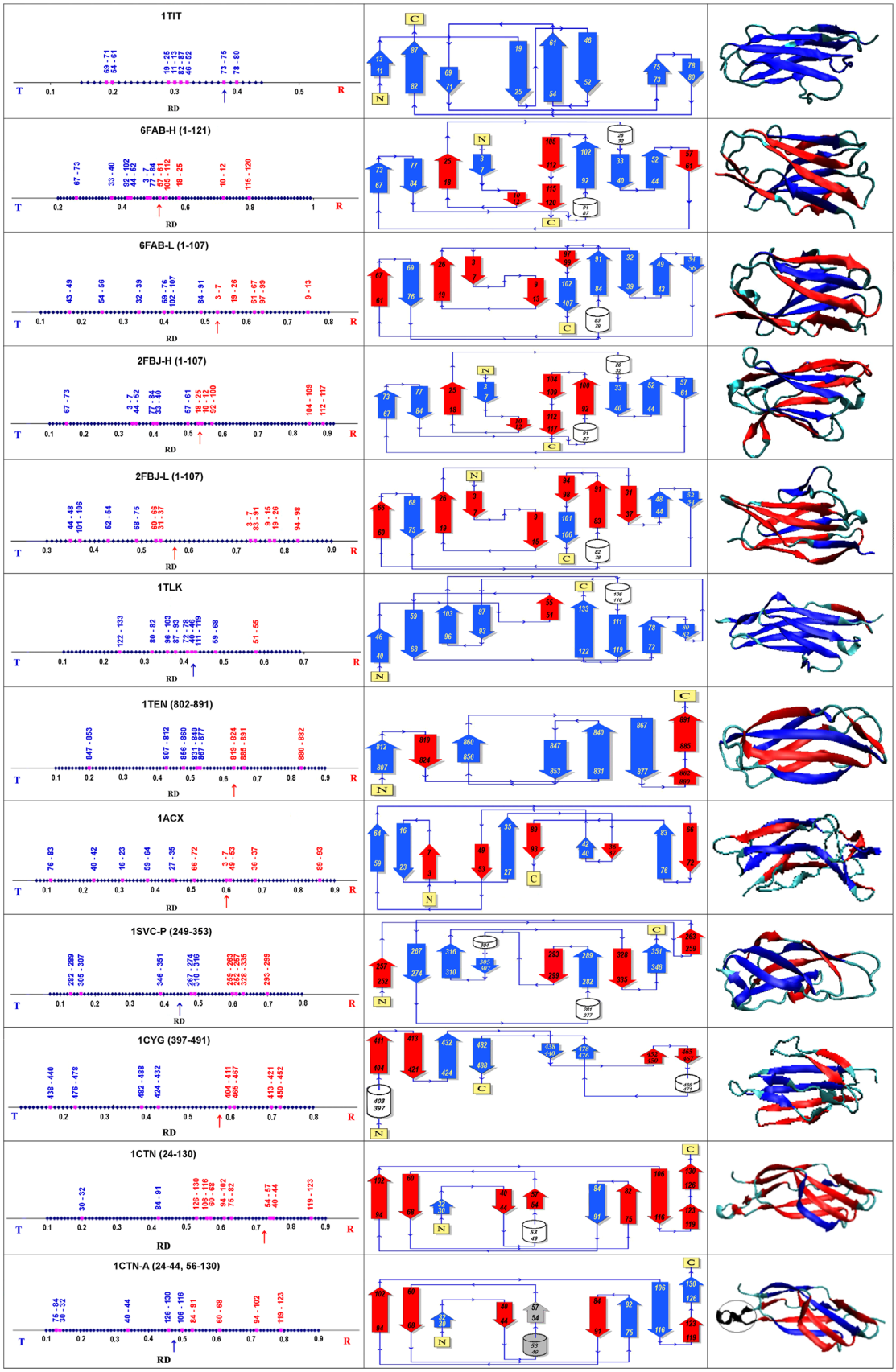

Fig. 3. Assessment of the β-structural status in selected immunoglobulin-fold domain proteins. The leftmost column shows a graphical representation of the Relative Distance (RD) values. The numbers listed above the horizontal axis indicate the residues that make up each β-fragment. A color code has been adopted, where red values represent the fragments that satisfy $O/T > O/R$ while blue values correspond to the fragments with $O/T < O/R$. Arrows below the axes indicate the status of each domain as a whole. The central column shows a schematic representation of the β-fragments using the same color code as above. Using the same color code, the right-hand column shows the 3D structure of each protein under consideration. The α-helical fragments (white cylinders) are not discussed here. The two gray fragments in the bottommost image represent parts of the protein chain that were excluded from the immunoglobulin-like domain analysis (and correspond to the black circled fragments in the right-hand image).

Table 2
List of proteins selected for detailed analysis and presentation. "HC" denotes a hydrophobic core.

PDB-ID	Name of protein/domain	Biological activity	Selection due to	Ref.
1TIT	Titin	Passive elasticity of muscles	HC entirely accordant with the model	Improta et al. (1996)
6FAB-H	Heavy chain of IgG – V domain	Antigen recognition; signal transduction	Immunoglobulin domain – not accordant	Strong et al. (1991)
6FAB-L	Light chain of IgG – V domain	Antigen recognition; signal transduction	Immunoglobulin domain – not accordant	Strong et al. (1991)
1TLK	C-terminal domain of myosin light chain kinase	Calmodulin binding	Domain accordant	Holden et al. (1992)
1TEN	Fibronectin type III	Cell adhesion protein	Domain not accordant	Leahy et al. (1992)
1SVC	NF-kappa B p50	DNA binding	Domain accordant	Müller et al. (1995)
1CYG	Cyclodextrin glucanotransferase	EC: 2.4.1.19	Domain not accordant	Laskowski (2009)
1CTN	Chitinase	Leavage of β-1-4-glycosidic	Structural role of additional loop	Perrakis et al. (1994)
6FAB-L	Light chain of IgG	Antigen recognition	Characteristics of hyper-variable loops	Strong et al. (1991)
1DVQ	transthyterin	Hormone/growth factor	Amyloidogenic	Klabunde et al. (2000)

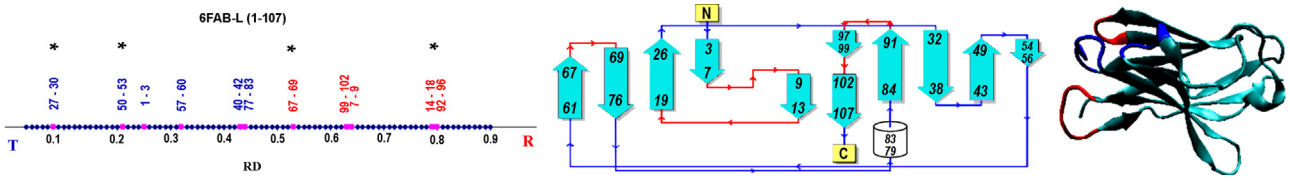

Fig. 4. The RD values for the V domain loops of 6FAB-L. Colors indicate status differences (cf. Fig. 3). Asterisks mark value ranges for CDRs. The helical fragments shown in turquoise (as well as those shown in white) are not covered by this analysis.

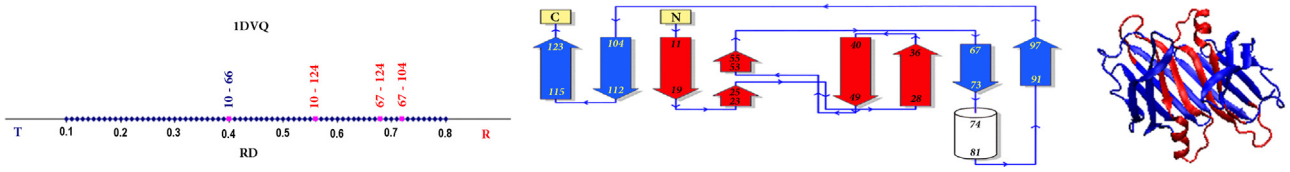

Fig. 5. Left panel: The calculated RD values for selected transthyretin fragments (1DVQ). The descriptions of all panels are the same as in Fig. 3. Center panel: The status of the β-structural fragments. The α-helical fragments (white cylinders) are not discussed here. Right panel: 3D representation of the dimeric form.

activity. Four out of its eight β-sandwich fragments diverge from the predictions of the fuzzy oil drop model.

3.3.6. *1CTN – lyase (oxo-acid), bacterial chitinase from Serratia marcescens (E.C.3.2.1.14) (Perrakis et al., 1994)*

The analysis herein focuses on its (24–130) domain which does not include the enzymatically active residues. This domain is a good example of how the RD values can depend on structural subunit definition. Although this domain is generally similar to immunoglobulins, it also possesses an "arm" – an extra loop that protrudes beyond its bounds. If the encapsulating "drop" is adjusted to cover the entire domain, the results indicate poor accordance with the model (see 1CTN panel in Fig. 3). However if the irregular fragment (protruding from the domain body) is excluded from encapsulation, the remainder of the domain is found to be highly accordant with the theoretical distribution (1CTN-A panel in Fig. 3). This issue is discussed in detail in Banach et al. (2013b).

3.4. *Antigen recognition in immunoglobulins*

The hypervariable loops of the V domains of immunoglobulins are known to recognize antigens and form stable complexes with them (Kabat, 1970; O'Nuallain et al., 2007; Wall et al., 1999). In order to account for the loops we conducted computations of RD values for the loops in the V domain in 6FAB-L (Figs. 3 and 4). Our analysis shows that two of the four hyper-variable loops (also called complementarity-determining regions; CDRs) within the V domain are consistent with the fuzzy oil drop model (Piekarska et al., 2006). The discordance of two hyper-variable loops is also accordant with our expectations since their low structural stability may facilitate contact and fitting to antigen molecules.

3.5. *Analysis of transthyretin in the context of amyloidosis*

One characteristic property of immunoglobulin domains, and of light-chain domains in particular, is their involvement in amyloidogenesis (Mizuguchi et al., 2008). The differing contribution of individual β-structural fragments to the formation of a common hydrophobic core suggests that they vary with regard to stability and indicates a possible tendency to form amyloid-like multi-protein complexes. A specific example of this phenomenon is the widely discussed transthyretin molecule (1DVQ), known for the strikingly differing amyloidogenic potentials of its N- and C-terminal fragments, which consist of approximately 50 residues each (Mizuguchi et al., 2008). In analyzing this protein we focused on its β-structural fragments as hydrophobicity density study units (central part of Fig. 6). Five out of nine β-structural fragments appeared to diverge from theoretical expectations regarding hydrophobicity density. However, the most characteristic property of thansthyretin is the large difference between its N- and C-terminal fragments with regard to their potential for amyloidogenesis. Our analysis, based on the fuzzy oil drop model, resulted in RD equal to 0.40 for the N-terminal fragment and 0.68 for the C-terminal fragment, with the C-terminal region (comprising

residues 67-104) exhibiting the greatest discordance from the model (RD=0.72) (Fig. 5 – left panel).

Calculation of the idealized and observed hydrophobicity density in OORF, using a 20 aa frame, highlights the differences between the N- and C-terminal fragments of a monomer unit of transthyretin (Fig. 6). Substantial differences between O/T and O/R values for N- and C-terminal fragments respectively can be observed, suggesting much higher stabilization (agreement between the observed and theoretical hydrophobicity density) the for N-terminal fragment. The idealized and observed hydrophobicity density profiles for the whole chain (upper panel in Fig. 6) reveal high accordance in the N-terminal fragment, however starting with residue number 62 the profiles become quite divergent.

The O/T and O/R profiles in the OORF calculation system (Fig. 6. – central panel) visualize the significant differences between both fragments under consideration with the observed distribution of the first 60 aa accordant with the assumed model.

The next 20 aa frame is much closer to the theoretical distribution. Residue number 62 begins a fragment where hydrophobicity density distribution approximates random values (i.e. no concentration of hydrophobicity in any area of the molecule). Thus, the C-terminal fragment does not appear to participate in the formation of a hydrophobic core.

The differentiation of polypeptide chain fragments in the dimeric form of transthyretin supports the possible correlation between the low stability of selected fragments and their ability to transition to amyloid-like structural forms (Fig. 5 – right panel). A possible interpretation of these observations is that the structure of the N-terminal fragment is stable while the C-terminal fragment is not stabilized by any form of hydrophobic core. This matches the known amyloidogenic properties of transthyretin.

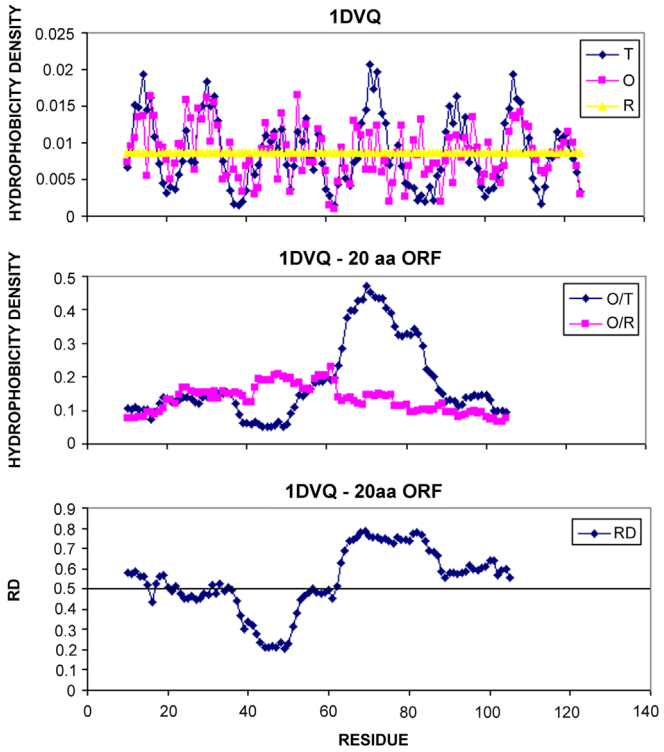

Fig. 6. Hydrophobicity density distribution in a single chain of transthyretin. Top: theoretical (T; blue rhombuses) and observed (O; pink squares) values compared to the unified (R; yellow triangles) distribution. Center: The O/T and O/R profiles calculated for 20-residue segments using the Overlapping Open Reading Frame (OORF) algorithm. Bottom: The RD profile corresponding to the O/T and O/R values calculated for the 20-residue segments using the OORF algorithm.

The RD profile (Fig. 6. – bottom panel) indicates good agreement between the observed hydrophobicity density distribution and the idealized profile for the N-terminal fragment (1–64aa) in contrast to the C-terminal fragment in which the distance versus the idealized distribution is very high, particularly for residues 65–90. This is why the input fragments for calculation of O/T and O/R values were selected on the basis of the profile (Fig. 6 – bottom panel C).

The rightmost panel in Fig. 6. presents the status (expressed as RD) of β-sheet fragments in the transthyretin dimer to show the specific localization of fragments representing significant disagreement between expectations (idealized distribution) and observation. This very strict differentiation in the monomer unit, as well as in the dimeric form of transthyretin, may be one of the factors which enable this protein to transition to an amyloid form. The dimeric form reveals the engagement of the C-terminal fragments in the inter-chain interface while the N-terminal fragment (whose structure remains accordant with the theoretical model) remains exposed to the water environment. This specific property of transthyretin will be the focus of further detailed analysis concerning easy transformation of a protein molecule into an amyloidogenic form.

4. Discussion and conclusions

The fuzzy oil drop model has been applied in the study of immunoglobulin-like structures (mainly β-sandwiches according to the CATH classification). Analysis of the agreement (or discordance) between the observed and idealized hydrophobicity density distributions enables differentiation of β-structural fragments. Local distortions in the distribution of hydrophobicity density, interpreted as structural instabilities, enable protein fragments to alter their structure as a response to external stimuli. The biological activity of V domains in immunoglobulins is typically to recognize an antigen, form a stable complex with it and permit transduction of immunological signals. The potential ability of hyper-variable loops to interact with external proteins (i.e. local excess of hydrophobicity) supports the assumption that the fuzzy oil drop model may be used as a criterion of recognition of biological activity in this type of protein molecules. β-fragments identified as unstable may be suspected of participation in immunological signal transduction. In contrast, the high stability of selected domains, as evidenced in 1TIT (titin), enables the protein to rapidly return to its native state following relaxation. Thus, the fuzzy oil drop model can be used to identify the biological function of proteins, as further explained in (Banach et al., 2012a, 2013a, 2012c; Konieczny and Roterman-Konieczna, 2012).

The differentiation of beta-structural forms may also provide clues as to the potential structural changes involved in amyloid fibril formation. This is shown on the example of tranthyretin and selected immunoglobulins.

Taking into account the results presented in this paper, as well as in other publications, one may assume that the 3D hydrophobicity density Gaussian may reflect the influence of an external force field, superimposed onto any internal force fields (which correspond to electrostatic and van der Waals interactions). An external force field of this form may direct the folding process towards the generation of a hydrophobic core, which would be quite difficult to simulate solely on the basis of traditional pairwise interactions between water and protein atoms.

Antifreeze (Banach et al., 2012d) and downhill proteins (Roterman et al., 2011) are identified as representing particularly good agreement with the theoretical hydrophobicity density distribution model, further supporting its validity. Local discrepancies between the idealized and

observed hydrophobicity density distributions seem to be related to highly specific interaction mechanisms.

The final structure of a domain (or protein) is likely to result from a balance between the internal and external force fields, including the presence of ligand molecules that actively influence structural changes in the folding process. Consequently, ligand molecules introduce local discrepancies (compared to the "ideal" hydrophobic core), and these local changes are typically related to the protein's biological function (Hagen, 2010; Chung and Eaton, 2013). In in silico folding, the multi-criteria optimization method (Konieczny and Roterman-Konieczna, 2012) has been used to strike a balance between internal and external interactions. This method has been shown to closely reflect experimental observations in a recent study (Chung and Eaton, 2013); however new questions concerning the folding process have arisen (Hagen, 2010), with particular focus on the pervasive effects of the presence of water (Chung and Eaton, 2013; Borgia et al., 2012).

Acknowledgments

We wish to thank Piotr Nowakowski for his help in authoring this manuscript, and Anna Śmietańska for technical support. The work was financially supported by the Jagiellonian University Medical College under Grant no. K/ZDS/001531.

References

Alejster, P., Banach, M., Jurkowski, W., Marchewka, D., Roterman-Konieczna, I., 2013. Comparative analysis of techniques oriented on the recognition of ligand binding area in proteins. In: Roterman-Konieczna, Irena (Ed.), Identification of Ligand Binding Site and Protein–Protein Interaction Area. Springer, Dordrecht, Heidelberg, New York, London, pp. 55–86 (2013).

Arakawa, T., Kamiya, N., Nakamura, H., Fukuda, I., 2013. Molecular dynamics simulations of double-stranded DNA in an explicit solvent model with the zero-dipole summation method. PLoS One 8 (10), e76606.

Banach, M., Konieczny, L., Roterman, I., 2013a. Can the structure of hydrophobic core determine the complexation site?. In: Roterman-Konieczna, Irena (Ed.), Identification of Ligand Binding Site and Protein–Protein Interaction Area, 2013. Springer, pp. 41–54.

Banach, M., Roterman, I., Prudhomme, N., Chomilier, J., 2013b. Hydrophobic core in domains of immunoglobulin-like fold. J. Biomol. Struct. Dyn. (PMID: 23998258).

Banach, M., Marchewka, D., Piwowar, M., Roterman, I., 2012a. The divergence entropy characterizing the internal force field in proteins, Protein folding in Silico Ed: Irena Roterman-Konieczna. Publisher: Woodhead Publishing, Oxford, Cambridge, Philadelphia, New Dehli, pp. 55–78.

Banach, M., Konieczny, L., Roterman, I., 2012b. Ligand binding site recognition. In: Roterman-Konieczna, Irena (Ed.), Protein folding In silico. β, Oxford, Cambridge, Philadelphia, New Dehli, pp. 78–94.

Banach, M., Konieczny, L., Roterman, I., 2012c. Use of the "fuzzy oil drop" model to identify the complexation area in protein homodimers. In: Roterman-Konieczna, Irena (Ed.), Protein Folding In SilicoWoodhead Publishing, Oxford, Cambridge, Philadelphia, New Dehli, pp. 95–122.

Banach, M., Prymula, K., Jurkowski, W., Konieczny, L., Roterman, I., 2012d. Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants. J. Mol. Model. 18 (1), 229–237.

Borgia, A., Wensley, B.G., Soranno, A., Nettels, D., Borgia, M.B., Hoffmann, A., Pfeil, S.H., Lipman, E.A., Clarke, J., Schuler, B., 2012. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. Nat. Commun. 3, 1195.

Brylinski, M., Konieczny, L., Roterman, I., 2007. Is the protein folding an aim-oriented process? Human haemoglobin as example. Int. J. Bioinform. Res. Appl. 3 (2), 234–260.

Chen, J., Stites, W.E., 1959. Packing is a key selection factor in the evolution of protein hydrophobic cores. Biochemistry 40 (50), 15280–15289.

Chung, H.S., Eaton, W.A., 2013. Single-molecule fluorescence probes dynamics of barrier crossing 502, 685–688Nature 502, 685–688.

Dill, K.A., MacCallum, J.L., 2012. The protein folding problem 50 years on. Science 338, 1042–1046.

Ewert, S., Honegger, A., Plückthun, A., 2004. Stability improvement of antibodies for extracellular and intracellular applications: CDR grafting to stable frameworks and structure-based framework engineering. Methods 24, 184–199.

Ewert, S., Huber, T., Honegger, A., Plückthun, A., 2003. Biophysical properties of human antibody variable domains. J. Mol. Biol. 325, 531–553.

Hagen, S.J., 2010. Solvent viscosity and friction in protein folding dynamics. Curr. Protein Pept. Sci. 11, 385–395.

Hamill, S., Steward, A., Clarke, J., 2000. The folding of an immunoglobulin like Greek key protein is defined by a common core nucleus and regions constrained by topology. J. Mol. Biol. 297, 165–178.

Holden, H.M., Ito, M., Hartshorne, D.J., Rayment, I., 1992. X-ray structure determination of telokin, the C-terminal domain of myosin light chain kinase, at 2.8 A resolution. J. Mol. Biol. 227, 840–851.

Huggins, D.J., 2012a. Benchmarking the thermodynamic analysis of water molecules around a model beta sheet. J. Comput. Chem. 33 (15), 1383–1392.

Huggins, D.J., 2012b. Correlations in liquid water for the TIP3P-Ewald, TIP4P-2005, TIP5P-Ewald, and SWM4-NDP models. J. Chem. Phys. 136 (6), 064518.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD–visual molecular dynamics. J. Mol. Graph. 14, 33–38.

Improta, S., Politou, A.S., Pastore, A., 1996. Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. Structure 4, 323–337.

Kabat, E.A., 1970. Heterogeneity and structure of antibody-combining sites. Ann. N Y Acad. Sci. 169, 43–54.

Kauzmann, W., 1959. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14, 1–63.

Klabunde, T., Petrassi, H.M., Oza, V.B., Raman, P., Kelly, J.W., Sacchettini, J.C., 2000. Rational design of potent human transthyretin amyloid disease inhibitors. Nat. Struct. Biol. 7, 312–321.

Koide, S., Huang, X., Link, K., Koide, A., Bu, Z., Engelman, D.M., 2000. Design of single-layer β-sheets without a hydrophobic core. Nature 403, 456–460.

Konieczny, L., Brylinski, M., Roterman, I., 2006. Gauss function based model of hydrophobicity density in proteins. In Silico Biol. 6 (1–2), 15–22.

Konieczny, L., Roterman-Konieczna, I., 2012. Conclussions. In: Roterman-Konieczna, Irena (Ed.), Protein folding in Silico. Woodhead Publishing, Oxford, Cambridge, Philadelphia, New Dehli191–201.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.

Laskowski, R.A., 2009. PDBsum new things. Nucleic Acids Res. 37, D355–D359.

Lazar, G.A., Handel, T.M., 1998. Hydrophobic core packing and protein design. Curr. Opin. Chem. Biol. 2, 675–679.

Leahy, D.J., Hendrickson, W.A., Aukhil, I., Erickson, H.P., 1992. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. Science 258, 987–991.

Levitt, M., 1976. A simplifed representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104, 59–107.

Li, P., Roberts, B.P., Chakravorty, D.K., Merz Jr., K.M., 2013. Rational design of particle mesh ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent. J. Chem. Theory Comput. 9 (6), 2733–2748.

Liu, Y., Palmer, J.C., Panagiotopoulos, A.Z., Debenedetti, P.G., 2012. Liquid–liquid transition in ST2 water. J. Chem. Phys. 137 (21), 214505.

Lu, H., Schulten, K., 2000. The key event in force-induced unfolding of titin's immunoglobulin domains. Biophys. J. 79, 51–65.

Marchewka, D., Jurkowski, W., Banach, M., Roterman, I., 2013. Prediction of protein–protein binding interfaces. In: Roterman-Konieczna, Irena (Ed.), Identification of Ligand Binding Site and Protein–Protein Interaction Area, 2013. Springer, pp. 105–134.

Mizuguchi, M., Hayashi, A., Takeuchi, M., Dobashi, M., Mori, Y., Shinoda, H., Lizawa, T., Lemura, M., Kawano, K., 2008. Unfolding and aggregation of transthyretin by the truncation of 50-N-terminal aminoacids. Proteins 72, 261–269.

Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., Harrison, S.C., 1995. Structure of the NF-kappa B p50 homodimer bound to DNA. Nature 373, 311–317.

Nelson, E., Grishin, N., 2006. Alternate pathways for folding in the flavodoxin fold family revealed by a nucleation growth model. J. Mol. Biol. 358, 646–653.

O'Nuallain, B., Allen, A., Kennel, S.J., Weiss, D.T., Solomon, A., Wall, J.S., 2007. Localization of a conformational epitope common to non-native and fibrillar immunoglobulin light chains. Biochemistry 46, 1240–1247.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., Thornton, J.M., 1997. CATH–a hierarchic classification of protein domain structures. Structure 5, 1093–1109.

Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S., Vorgias, C.E., 1994. Crystal structure of a bacterial chitinase at 2.3 A resolution. Structure 2, 1169–1180.

Piekarska, B., Drozd, A., Konieczny, L., Król, M., Jurkowski, W., Roterman, I., Spólnik, P., Stopa, B., Rybarska, J., 2006. The indirect generation of long-distance structural changes in antibodies upon their binding to antigen. Chem. Biol. Drug. Des. 68 (5), 276–283.

Prudhomme, N., Chomilier, J., 2009. Prediction of the protein folding core: application to the immunoglobulin fold. Biochimie 91, 1465–1474.

Rico, F., Gonzales, L., Casuso, I., Puig-Vidal, M., Scheuring, S., 2013. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulation. Science 342, 741–743.

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

Ryu, S-E., Choi, H-J., Kwon, K-S., Lee, K.N., Yu, M-H, 1996. The native strains in the hydrophobic core and flexible reactive loop of a serine protease inhibitor: crystal structure of an uncleaved α_1-antitrypsin at 2.7 å. Structure 4, 1181–1192.

Sala, J., Guàrdia, E., Masia, M., 2010. The polarizable point dipoles method with electrostatic damping: implementation on a model system. J Chem. Phys. 33 (23), 234101.

Sirk, T.W., Moore, S., Brown, E.F., 2013. Characteristics of thermal conductivity in classical water models. J. Chem. Phys. 138 (6), 064505.

Strong, R.K., Campbell, R., Rose, D.R., Petsko, G.A., Sharon, J., Margolies, M.N., 1991. Three-dimensional structure of murine anti-p-azophenylarsonate Fab 36-71. 1. X-ray crystallography, site-directed mutagenesis, and modeling of the complex with hapten. Biochemistry 30, 3739–3748.

Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikoff, S., Davies, D.R., 1986. The galactan-binding immunoglobulin Fab J539: an X-ray diffraction study at 2.6-A resolution. Proteins 1, 74–80.

Wall, J., Schell, M., Murphy, C., Hrncic, R., Stevens, F.J., Solomon, A., 1999. Thermodynamic instability of human lambda 6 light chains: correlation with fibrillogenicity. Biochemistry 38, 14101–14108.

Yang, H., Jiao, X., Li, S., 2012. Hydrophobic core–hydrophilic shell-structured catalysts: a general strategy for improving the reaction rate in water. Chem. Commun. 48, 11217–11219.