Chapter 5
Possible Mechanism of Amyloidogenesis of V Domains

Mateusz Banach, Barbara Kalinowska, Leszek Konieczny, and Irena Roterman

Abstract This chapter discusses complexation of Congo red by amyloid structures comprising immunoglobulin light chains, particularly the so-called Bence-Jones (BJ) proteins. According to the presented study, in BJ proteins the V domain is substantially less stable than the C domain. This conclusion is based on quantitative analysis of the protein's hydrophobic core, made possible by the fuzzy oil drop model. Results indicate that the V domain exhibits structural ordering characteristic of amyloid aggregates, i.e. linear propagation of local hydrophobicity peaks and troughs rather than a monocentric hydrophobic core (typically present in globular proteins). On this basis, the authors propose a hypothetical arrangement of V domains which leads to formation of an amyloid. Structural similarities between V domains in BJ proteins and other types of amyloid aggregates enable the authors to study the specific mechanism of Congo red complexation by amyloids.

The proposed Congo red complexation mechanism builds upon the authors' previous experience with bioinformatics tools. The subject should be of interest to researchers specializing in protein folding studies and misfolding diseases.

Keywords Hydrophobicity • Bence-Jones proteins • V domain of IgG • Amyloid • Force field • Congo red • Immunoglobulins • Amyloidgenesis • Supramolecular ligand • Misfolding

M. Banach (✉) • I. Roterman
Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Łazarza 16, 31-530, Krakow, Poland
e-mail: mateusz.banach@uj.edu.pl; myroterm@cyf-kr.edu.pl

B. Kalinowska
Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348, Krakow, Poland
e-mail: malijka@gmail.com

L. Konieczny
Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kopernika 7, 31-034 Krakow, Poland
e-mail: mbkoniec@cyf-kr.edu.pl

I. Roterman, L. Konieczny (eds.), *Self-assembled Molecular Structures – New Kind of Protein Ligand*, https://doi.org/10.1007/978-3-319-65639-7_5

5.1 Characteristics of Immunoglobulin Domains

Complexation of supramolecular Congo red (CR) requires considerable relaxation of the target protein's native form. Immunoglobulin domains are known for their structural instabilities (amyloidogenic properties), and also for their ability to eagerly bind CR. In line with basic biochemical knowledge, the stability of proteins is determined by two factors: disulfide bonds and hydrophobic core.

Immunoglobulin domains are β-sandwiches composed of two distinct fragments, referred to as the upper core and the lower core respectively. Each domain includes one centrally placed disulfide bond linking both cores. While the stabilizing influence of this bond allows the domain to persist in its native form, individual domains vary greatly with respect to the structure of their hydrophobic cores [1]. This diversity is evidenced by structural analysis based on the fuzzy oil drop (FOD) model. FOD is an extension of the "oil drop" hydrophobicity distribution model, which introduced a binary distinction between the outer (hydrophilic) and inner (hydrophobic) layers [2]. The name of the model alludes to the notion of an oil drop immersed in water – the hydrophobic substance attempts to minimize its contact surface, becoming spherical in the process. Similarly, the protein folding process results in internalization of hydrophobic residues and exposure of hydrophilic residues on the surface of a globular capsule [3].

As already remarked, the fuzzy oil drop model extends the binary oil drop paradigm by introducing a continuous gradient of hydrophobicity between the core and the surface. This gradient is mathematically expressed by a 3D Gaussian, which peaks at the center of the molecule and then gradually decreases, reaching near-zero values at a distance of 3σ (where σ is the coefficient of the Gaussian).

In this chapter we will apply the fuzzy oil drop model in the analysis of immunoglobulin light chain domains exemplified by Bence-Jones proteins [4].

5.2 Target Proteins

Table 5.1 provides a summary of proteins selected for analysis. They are collectively referred to as Bence-Jones (BJ) proteins [4] and are exclusively of human origin. Each protein is a homodimer comprised by two identical IgG light chains.

The sole exception is 2Q1E, where two pairs of V domain dimers have been identified in the crystal structure. The reference protein is the Fab fragment of human immunoglobulin G (4PUB), consisting of the light chain (L) and the heavy chain (H).

The fuzzy oil drop has been applied in the analysis of the following structural units: both chains in complex; paired domains – V(L)-V(L) and C(L)-C(L), each domain individually and for 4PUB the complexes of domains V(L)–V(H) and C(L)-C(H). We have also assessed the status of each domain as a structural subunit of the complete dimer.

Table 5.1 Proteins selected for analysis

Protein – ID PDB	Dimer	Chain class	References
1B6D	L-L	KAPPA	[5]
1BJM	L-L	LAMBDA	[6]
1DCL	L-L	LAMBDA	[7]
1LIL	L-L	LAMBDA	[8]
2OLD	L-L		[9]
2OMB	L-L		[9]
2OMN	L-L		[9]
2Q1E	L-L-TETRAMER	KAPPA	[10]
3BJL	L-L	LAMBDA	[6]
4PUB	H-L		[11]

5.3 The Fuzzy Oil Drop Model

The fuzzy oil drop model has been extensively described in numerous publications [3, 12, 13]. The description presented below should therefore be regarded only as a brief introduction.

The traditional notion of a "hydrophobic core" refers to a concentration of hydrophobic residues at the geometric center of the protein, along with exposure of hydrophilic residues on its surface. When dealing with globular proteins, this configuration can be described with a 3D Gaussian, where the origin of the coordinate system coincides with the geometric center of the molecule and a separate σ coefficient is defined for each principal axis, delineating an ellipsoid capsule. The molecule itself should be oriented in such a way as to align its orthogonal dimensions (longest diagonals) with axes of the coordinate system. Values of σ_x, σ_y and σ_z are calculated as 1/3 of the separation between the origin of the system and the position of the most distant atom along each axis. This is schematically depicted in Fig. 5.1.

In accordance with the three-sigma rule, over 99% of the total volume of the Gaussian is captured by applying a cutoff distance of 3σ in each principal direction. The value of the Gaussian at any point within this ellipsoid capsule is interpreted as local theoretical hydrophobicity (also referred to as the "idealized" distribution).

The theoretical hydrophobicity distribution should be confronted with the observed distribution, which depends on local interactions between each residue and its neighbors. These calculations are based on the positions of the so-called effective atoms (averaged-out positions of all atoms comprising a given residue) and the intrinsic hydrophobicity of each amino acid. Each effective atom collects interactions with its neighbors, with a cutoff distance of 9 Å.

Theoretical (T) hydrophobicity is expressed by the following formulae:

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

$\tilde{H}t_j$ is the theoretical hydrophobicity density (hence the *t* designation) at the *j*th point in the protein body. $\overline{x}, \overline{y}, \overline{z}$ correspond to the peak of the Gaussian in each of the three principal directions, while σ_x, σ_y, σ_z denote the range of arguments for each coordinate system axis. These coefficients are selected in such a way that 99% of the Gaussian's integral is confined to a range of $\overline{x} \pm 3\sigma$. Values of the distribution can be assumed to equal 0 beyond this range.

If the molecule is placed inside a capsule whose dimensions are given by $\overline{x} \pm 3\sigma_x, \overline{y} \pm 3\sigma_y, \overline{z} \pm 3\sigma_z$ then the values of the corresponding Gaussian represent the idealized hydrophobicity density distribution for the target protein. If $\sigma_x = \sigma_y = \sigma_z$, the capsule is perfectly spherical; otherwise it is an ellipsoid. The Gaussian yields hydrophobicity density values at arbitrary points in the protein body – for example at points which correspond to the placement of effective atoms (one per side chain). $\tilde{H}t_j$ is the hydrophobicity density determined for the *j*th amino acid while x, y and z indicate the placement of its corresponding effective atom.

The denominator of $\frac{1}{\tilde{H}t_{sum}}$ expresses the aggregate sum of all values given by the Gaussian for each amino acid making up the protein. This enables normalization of the distribution since $\tilde{H}t_j$ will always be equal to 1.0.

$\tilde{H}t_j$ values reflect the expected hydrophobicity density which should correspond to each amino acid in order for the hydrophobic core to match theoretical predictions with perfect accuracy, with all hydrophobic residues internalized and all hydrophilic residues exposed on the surface. The closer to the surface the lower the expected hydrophobicity density.

The position of each (*j*th) residue is represented effective atom localized at the geometric center of all atoms belonging to side chain of the residue under consideration (including Cα in the case of Gly). Protein encapsulation is presented in Fig. 5.1.

On the other hand, the actual distribution of hydrophobicity density observed (O) in a protein molecule depends on inter-chain interactions, which, in turn, depend on the intrinsic hydrophobicity of each amino acid. Intrinsic hydrophobicity can be determined by experimental studies or theoretical reasoning – our work bases on the scale published in [12] while the force of hydrophobic interactions has been calculated using algorithms proposed in [14]. For each amino acid *j* (or, more accurately, for each effective atom) the sum of interactions with its neighbors is computed and subsequently normalized by dividing it by the number of elementary interactions:

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

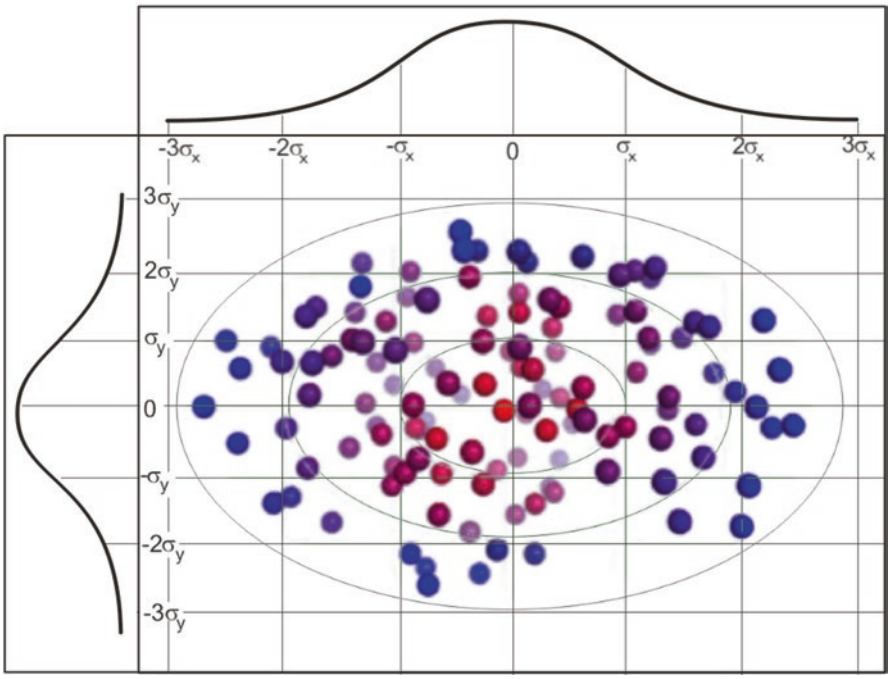

Fig. 5.1 Protein molecule placed in an ellipsoid capsule (only two dimensions are presented to preserve clarity). The adjacent one-dimensional plots present values of the Gaussian along each axis and highlight the role of σ coefficients. Color coding expresses the transition between hydrophilic (*blue*) and hydrophobic (*red*) residues. An intermediate layer is present between the core and the surface

N is the number of amino acids in the protein, $\tilde{H}_i^r$ expresses the hydrophobicity parameter of the i-th residue while r_{ij} expresses the distance between two interacting residues (jth effective atom and ith effective atom). c expresses the cutoff distance for hydrophobic interactions, which is taken as 9.0 Å (following [14]). The $\tilde{H}o_{sum}$ coefficient, representing the aggregate sum of all components, is needed to normalize the distribution which, in turn, enables meaningful comparisons between the observed and theoretical hydrophobicity density distributions.

Quantitative analysis of the differences between expected (T) and observed (O) distributions is enabled by the Kullback-Leibler entropy formula [15]:

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

The value of D_{KL} expresses the distance between the observed (p) and target (p_0) distributions, the latter of which is given by the 3D Gaussian (T). The observed distribution is referred to as O.

For the sake of simplicity, we introduce the following notation:

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

Since D_{KL} is a measure of entropy its interpretation requires a reference value. In order to facilitate meaningful comparisons, we introduce another boundary distribution (referred to as "unified" or R) which corresponds to a situation where each effective atoms represents the same hydrophobicity density (1/*N*, where *N* is the number of residues in the chain). In this type of distribution hydrophobicity density is not concentrated at any point in the protein body.

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

Comparing O|T and O|R tells us whether the given protein (O) more closely approximates the theoretical (T) or unified (R) distribution. Proteins for which O|T > O|R are regarded as lacking a prominent hydrophobic core. To further simplify matters we introduce the following relative distance (RD) criterion:

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

Here, RD < 0.5 indicates the presence of a hydrophobic core.

Figure 5.2 presents a graphical representation of RD values, restricted (for simplicity) to a one dimensional form.

D_{KL} (as well as O|T, O|R and RD) may be calculated for specific structural units (complex, single molecule, single chain, selected domain). In such cases the bounding ellipsoid is restricted to the selected fragment of the protein. It is also possible to determine the status of polypeptide chain fragments within the context of a given ellipsoid. This procedure requires prior normalization of O_i, T_i and R_i values belonging to the analyzed fragment.

The procedure described above will be consistently applied in the analysis presented in this chapter. The status of selected polypeptide chain fragment will be studied to evaluate their participation in forming a hydrophobic core. In particular, secondary folds which satisfy RD < 0.5 are thought to contribute to the molecule-wide hydrophobic core. When the opposite is true (i.e. RD > 0.5), the given fragment can be considered unstable. It appears that fragments which exhibit higher-than-expected hydrophobicity may, when exposed on the surface, be engaged in protein complexation (forming parts of the interface).

Calculations concerning fragments of the polypeptide chain requires prior normalization of T_i, O_i and R_i values belonging to the selected fragment. The results tell us whether the given fragment contributes to the molecule-wide hydrophobic core.

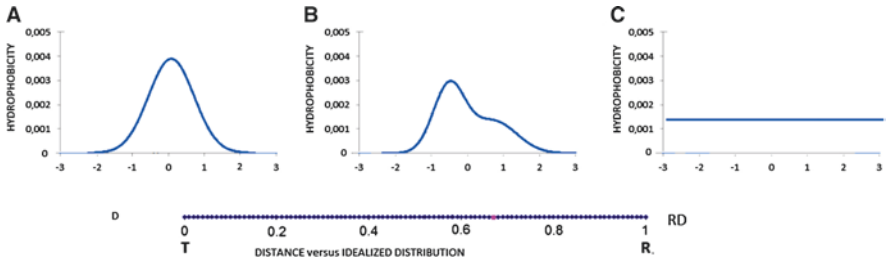

Fig. 5.2 Graphical representation of fuzzy oil drop model parameters, reduced to a single dimension. (**A**) theorized Gaussian distribution (T). (**B**) actual hydrophobicity distribution in the protein under consideration. (**C**) uniform distribution (R). (**D**) The RD parameter (equal to 0.656) marked on the horizontal axis as a *pink dot*. According to the fuzzy oil drop model this protein does not contain a well-defined hydrophobic core

Fragments can be selected for analysis on the basis of their involvement in particular secondary structures [1], supersecondary structures [16], interface areas [17], intrinsically distorted fragments [18], chameleon fragments [19] or other types of structures, depending on the research problem at hand.

A summary of sample proteins, visualizing the varied status of their hydrophobic cores, is provided in Fig. 5.3. Titin (Fig. 5.3A) is a protein which includes an immunoglobulin-like fold, exhibiting very good agreement between the theoretical and the observed hydrophobicity distribution. In this case hydrophobicity is concentrated near the center of the protein, with a hydrophilic layer present on the surface, optimizing the protein's contact with water. Such high agreement between T and O enables titin to revert to its native conformation in the absence of external forces (note that titin is found in muscle tissue and subject to frequent stretching).

The second sample protein, visualized in Fig. 5.3B, is the H chain of the human immunoglobulin Fab fragment. In this case major discrepancies between the theoretical and observed hydrophobicity distribution are observed (RD = 0.584), indicating that no monocentrichydrophobic core is present and that the H chain as a whole is only marginally stable. Further stabilization is provided by two disulfide bonds present in the Fab fragment. Notably, immunoglobulin appears to require a flexible V domain in order to align itself with the antigen.

The third protein, transthyretin, is a known amyloid precursor (Fig. 5.3C). Major differences between the N-terminal and the C-terminal fragments are evident. The protein as a whole does not follow the theoretical distribution of hydrophobicity, although the N-terminal section is a far better match for the theoretical values than its C-terminal counterpart. In such cases, it is informative to compute RD values for specific fragments of the chain, revealing the degree of their participation in the protein's hydrophobic core.

The presented work focuses on Bence-Jones complexes formed by IgG light chains [4]. Detection of such proteins in urine may indicate multiple myeloma or Waldenström's macroglobulinemia. Large deposits of B-J proteins are also encountered in kidneys and may cause amyloidogenesis [20].

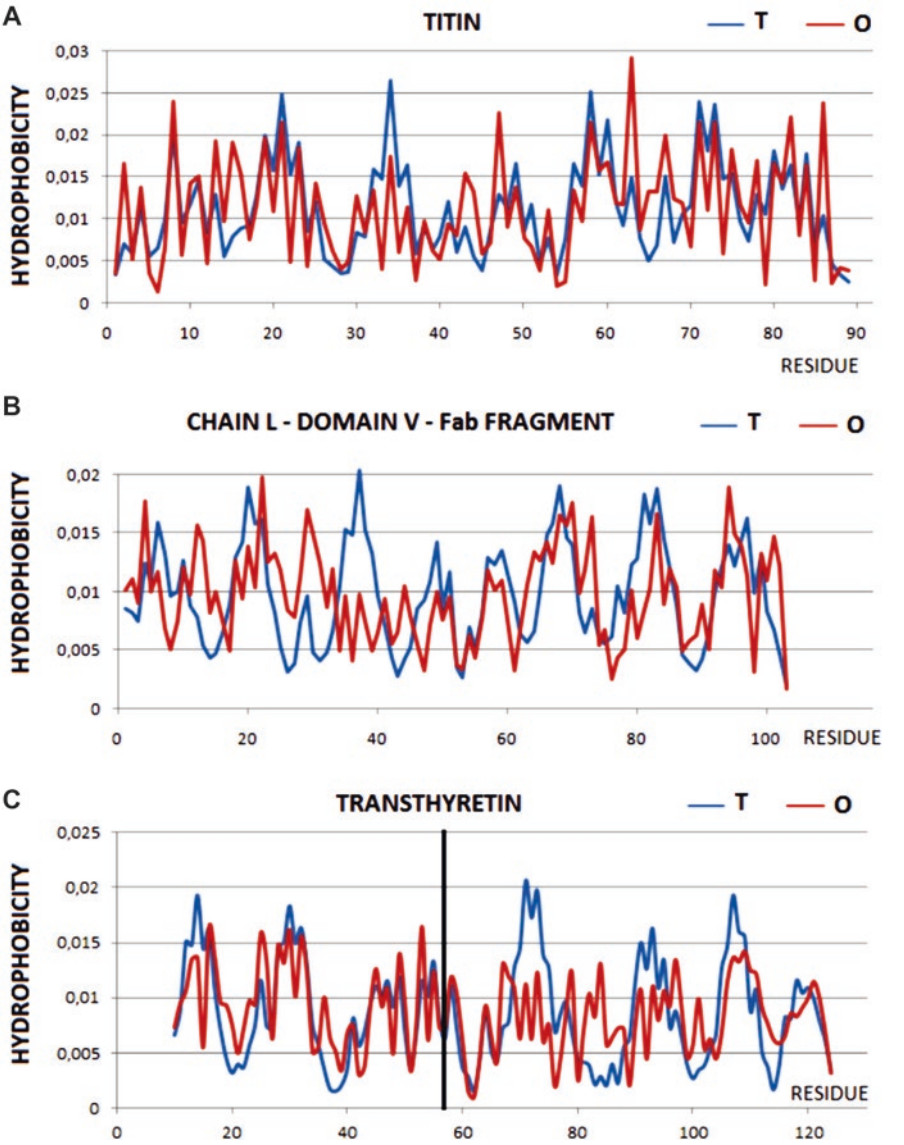

Fig. 5.3 Comparison of three proteins: (**A**) O matches T - immunoglobulin-like domain; titin; RD = 0.326. (**B**) O diverges from T – domain V chain L; human immunoglobulin Fab fragment; RD = 0.584. (**C**) O partially matches T; however areas of significant discordance are evident – transthyretin; RD = 0.562 (entire chain), 0.475 (accordant fragment at 10–57), 0.621 (discordant fragment at 51–124). The *vertical line marks* the boundary between both fragments

The reference complex is provided by the light/heavy chain dimer corresponding to the native form of immunoglobulin (4PUB).

5.4 Structure of Hydrophobic Core in B-J Proteins

The fuzzy oil drop model has been applied to identify the status of immunoglobulin domains in proteins referred to as Bence-Jones complexes. This choice of proteins is motivated by their specific properties, particularly their ability to quickly transition into amyloid forms (indicating structural instabilities) and their high affinity for supramolecular CR.

The fuzzy oil drop model reveals the status of the complete protein (complex), its individual domains as well as fragments of polypeptide chains – the presented analysis covers the status of the interface area and the N-terminal fragment which has been experimentally characterized as highly unstable [21].

5.4.1 Dimers of L-L Chains in Bence-Jones Proteins and of L-H Chains in the Fab Fragment

Table 5.2 presents a comparison of RD parameters describing full-chain dimers and V/C domains present in the complex. Analysis of results indicates that the full-chain dimer does not contain a shared hydrophobic core in the sense of the fuzzy oil drop model (with all corresponding RD values in excess of 0.5).

When analyzed as components of the complex, V domains exhibit high RD values, which suggests that they lack prominent monocentrichydrophobic cores. Two exceptions to this rule are 1LIL and 2Q1E. In the latter case, the discrepancy is due to altered composition of the V domain tetramer crystals, where each unit cell comprises two dimers with two domains per dimer. For this reason, 2Q1E will be frequently seen as an outlier in further analysis.

C domains analyzed in the context of the dimer exhibit good agreement with the theoretical hydrophobic core structure.

Interesting properties are revealed for the light/heavy chain complex in the Fab IgG fragment, where each domain is discordant in the context of the complex-wide hydrophobic core.

5.4.2 V(L)-V(L), C(L)-C(L), V(H)-V(L) and C(H)-C(L) Dimers

Since immunoglobulin chains consist of clearly distinguished paired domains, it is interesting to study the status of V(L)-V(L) and C(L)-C(L) dimers. Table 5.3 provides the corresponding quantitative characterization. In the case of the Fab fragment, we have analyzed its V(L)-V(H) and C(L)-C(H) dimers, which should be regarded as a reference.

Table 5.2 RD values describing individual domains of light chain dimers in Bence-Jones proteins

Protein – ID PDB	Dimer	Chain/domain V	RD	Chain/domain C	RD
1B6D	**0.778**	A-V (1-107)	**0.541**	A-C (108-211)	0.470
		B-V (1-107)	**0.566**	B-C (108-211)	0.417
1BJM	**0.718**	A-V (2-111)	**0.558**	A-C (112-212)	0.383
		B-V (2-111)	**0.590**	B-C (112-212)	0.331
1DCL	**0.752**	A-V (1-111)	**0.588**	A-C (112-212)	0.344
		B-V (1-111)	**0.567**	B-C (112-212)	0.343
1LIL	**0.733**	A-V (2-107)	0.473	A-C (108-211)	0.347
		B-V (2-107)	0.485	B-C (108-211)	0.336
2OLD	**0.734**	A-V (2-112)	**0.588**	A-C (113-213)	0.376
		B-V (2-112)	**0.591**	B-C (113-213)	0.310
2OMB	**0.803**	A-V (2-112)	**0.597**	A-C (113-213)	0.420
		B-V (2-112)	**0.589**	B-C (113-213)	0.303
		C-V (2-112)	**0.602**	C-C (113-213)	0.351
		D-V (2-112)	**0.600**	D-C (113-213)	0.347
2OMN	**0.750**	A-V (2-112)	**0.592**	A-C (113-213)	0.390
		B-V (2-112)	**0.581**	B-C (113-213)	0.300
2Q1E	**0.786**	A-V	0.492	A-C	0.492
		B-V	**0.507**	B-C	**0.507**
		C-V	0.489	C-C	0.489
		D-V	0.464	D-C	0.464
3BJL	**0.712**	A-V (2-111)	**0.583**	A-C (112-212)	0.375
		B-V (2-111)	**0.586**	B-C (112-212)	0.312
4PUB	**0.747**	H-V (2-122)	**0.754**	H-C (123-223)	**0.663**
		L-V (1-107)	**0.818**	L-C (108-212)	**0.728**

The reference protein (4PUB) is the Fab fragment of human immunoglobulin, composed of light (L) and heavy (H) chains

Table 5.3 RD values for V and C domain dimers

	V-V dimer			C-C dimer		
Protein	Complete	No P-P	P-P	Complete	No P-P	P-P
1B6D	**0.752**	**0.573**	**0.630**	**0.538**	**0.508**	**0.505**
1BJM	**0.683**	**0.648**	0.492	**0.529**	0.492	0.317
1DCL	**0.751**	**0.732**	**0.505**	**0.540**	**0.513**	0.291
1LIL	**0.695**	**0.663**	**0.643**	**0.502**	0.445	0.245
2OLD	**0.714**	**0.697**	**0.534**	**0.546**	**0.525**	0.300
2OMB	**0.755**	**0.731**	**0.745**	**0.542**	**0.521**	**0.535**
2OMN	**0.750**	**0.728**	**0.596**	**0.532**	**0.520**	0.283
2Q1E	**0.753**	**0.722**	0.475			
3BJL	**0.655**	**0.630**	0.385	**0.561**	**0.528**	0.291
4PUB	**0.649**	**0.603**	0.450	**0.531**	**0.515**	0.373

The "no P-P" columns characterize the status of domains following elimination of residues involved in P-P interactions, while the "P-P" columns presents the interface fragments. Values listed in boldface represent discordance

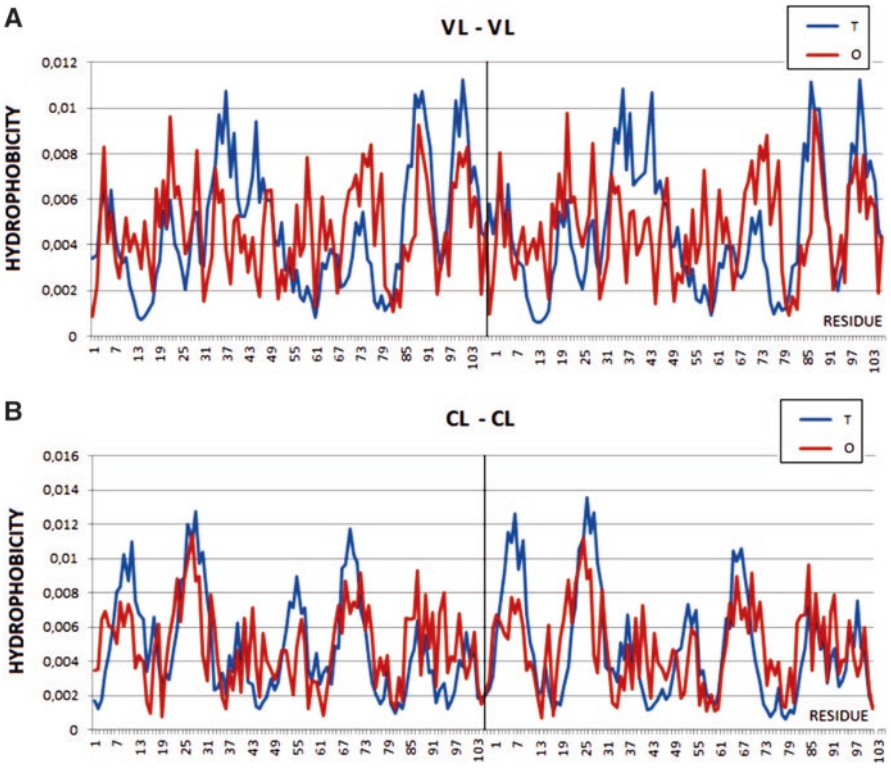

Fig. 5.4 Hydrophobicity distributions in protein domains: (**A**) V(L)-V(L), (**B**) C(L)-C(L). The divergence between T and O is greater in V domains than in the C(L)-C(L) dimer

Figure 5.4 presents the theoretical and observed distribution of hydrophobicity in 1B6D (B-J protein). Of note is the significant deviation corresponding to the V domain, with only the C-terminal complex consistent with the theoretical distribution. Comparison of V domain profiles with their C domain counterparts reveals greater stability of the C dimer (at least from the point of view of its hydrophobic core structure).

In all listed proteins, full V-V and C-C dimers deviate from the expected monocentric distributions of hydrophobicity. Elimination of residues involved in interdomain interactions produces a reduction in RD values, showing that the interface zone disrupts the structure of the hydrophobic core in each domain.

On the other hand, when analyzing the distribution of hydrophobicity in the interface itself, it turns out that a significant majority of C-C dimers match the theoretical distribution, and the same is true for the Fab fragment. Evidently, the distribution of hydrophobicity in interface residues corresponds to FOD predictions (except in 1B6D and 2OMB).

In general, the structure of C-C dimers may be interpreted as relatively stable, whereas V-V dimers are characterized by low stability.

5.4.3 *Individual Domains*

This part of the presentation focuses on individual domains regarded as standalone structural units. For each domain, a separate 3D Gaussian is plotted and the corresponding RD values calculated. Results are listed in Table 5.4.

If we base our analysis on the textbook definition of a domain (i.e. a distinct structural unit which folds on its own), the FOD properties of standalone domains should reveal their intrinsic structural stability. Analysis of results presented in Table 5.4 suggests significant differences between V and C domains. The former are generally discordant (lack hydrophobic cores), while in the latter case a prominent hydrophobic core is present for each analyzed B-J protein, with the sole exception being 1LIL. Eliminating residues involved in inter-domain interactions bring the status of V domains in line with the theoretical model – we may therefore conclude that structural instabilities are primarily due to the presence of an inter-domain interface. This theory is corroborated by the poor agreement between T and O in the interface itself.

Contrasting properties are observed in the V(L) and V(H) domains comprising the IgG Fab fragment, with both units conforming to the model. The status of Fab V(L) is similar to that of its B-J counterparts.

The observed discrepancies between the status of V and C domains in both chains (L and H) provide important clues regarding amyloidogenesis. This phenomenon is more frequently observed in B-J proteins, although L-H dimers are not immune from it. Notably, amyloidogenesis tends to involve V domains rather than C domains [22].

As already remarked, the peculiar status of 2Q1E is due to differences in its crystal structure, with a marked decrease in the quantity of residues involved in inter-domain interactions (16 compared to 29–38 in other dimers). Regarding 1LIL, its dimerization properties differ from other proteins in the study set due to structural differences in the interface zone.

5.5 Role of the N-Terminal Fragment in V Domains

The N-terminal fragment of the light chain V domain has been identified as highly unstable on the basis of experimental results [21]. This conclusion is supported by molecular dynamics simulations involving B-J proteins.

Eliminating the N-terminal fragment results in a significant decrease in RD values, proving that the fragment contributes to destabilization of the domain and disrupts its hydrophobic core (Table 5.5).

The disagreement between O and T distribution as observed in N-terminal fragment is visualized in Fig. 5.5. The status of observed hydrophobicity distribution of position 5 and fragment 11–14 can be even treated as opposite one versus the expected hydrophobicity. RD value for this fragment is equal to 0.604.

Table 5.4 RD values for individual domains comprising the presented proteins

	RD Values					
	Domain		No P-P interaction		P-P interface	
Protein	Domain V	Domain C	Domain V	Domain C	Domain V	Domain C
1B6D	**0.541/0.566**	0.470/0.417	0.195/0.183	0.471/0.414	**0.602/0.626**	0.428/0.481
1BJM	**0.558/0.590**	0.383/0.331	0.217/0.184	0.383/0.352	**0.664/0.730**	0.361/0.159
1DCL	**0.588/0.567**	0.344/0.343	0.178/0.173	0.345/0.348	**0.506/0.522**	0.287/0.234
1LIL	0.474/0.485	0.347/0.336	0.239/0.242	0.344/0.348	**0.539/0.693**	0.284/0.210
2OLD	**0.588/0.591**	0.376/0.310	0.159/0.151	0.386/0.317	**0.645/0.613**	0.145/0.177
2OMB	**0.597/0.589**	0.420/0.303	0.153/0.178	0.416/0.308	**0.664/0.731**	0.396/0.272
2OMN	**0.592/0.581**	0.390/0.301	0.154/0.164	0.389/0.313	**0.603/0.646**	0.224/0.173
2Q1E	0.492/**0.507**		0.473/0.485		**0.834/0.872**	
	0.490/0.465		0.473/0.420		**0.820/0.738**	
3BJL	**0.583/0.586**	0.375/0.312	0.189/0.193	0.374/0.327	**0.648/0.784**	0.257/0.117
4PUB (H/L)	0.473/0.460	**0.595**/0.480	**0.594**/0.465	**0.594**/0.496	0.393/0.384	**0.568**/0.323

Reference data is provided for 4PUB (Fab fragment). Calculations are performed for each domain separately

Table 5.5 RD values for complete V domains (central column) and following elimination of the N-terminal fragments (right column)

Protein	Domain V	
	Complete	No N-Terminal Fragment
1B6D	**0.541/0.566**	0.211/0.198
1BJM	**0.558/0.590**	0.219/0.189
1DCL	**0.588/0.567**	0.208/0.204
1LIL	0.474/0.485	0.244/0.241
2OLD	**0.588/0.591**	0.168/0.168
2OMB	**0.597/0.589**	0.166/0.175
2OMN	**0.592/0.581**	0.175/0.177
2Q1E	0.492/**0.507**	0.432/0.442
	0.490/0.465	0.433/0.420
3BJL	**0.583/0.586**	0.196/0.190
4PUB (H/L)	0.460	0.453

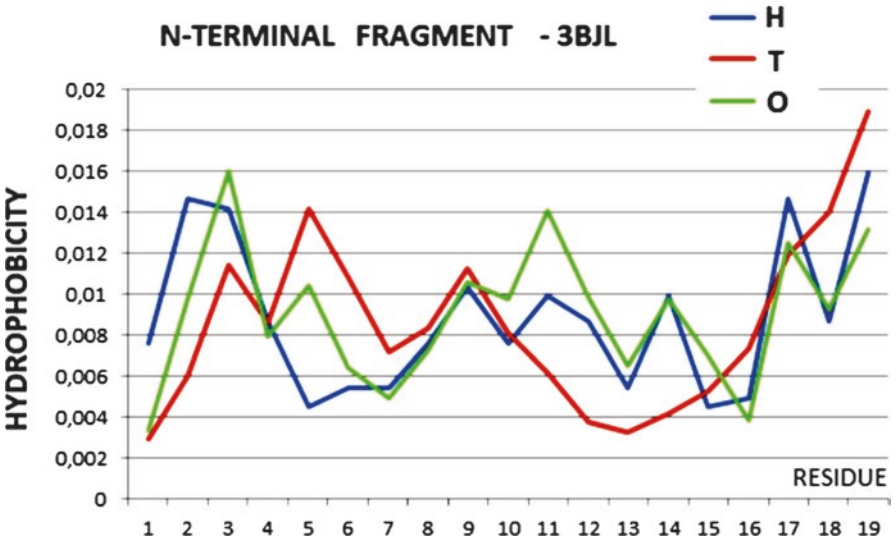

Fig. 5.5 The N-terminal fragment – profiles of T, O and H distribution

The N-terminal section in the BJ dimer occupies an exposed position, facing the environment. This renders it susceptible to structural changes – it may become uncoiled, freeing itself from the influence of the protein (in the sense of the FOD model) [21]. Figure 5.6 depicts this situation distinguishing the N-terminal fragment and fragments engaged in interface generation as the region of lower stability due to lover engagement in hydrophobic core generation.

According to fuzzy oil drop model – fragments of the polypeptide chain which do not contribute to the molecule-wide hydrophobic core are regarded as potentially unstable and susceptible to structural changes.

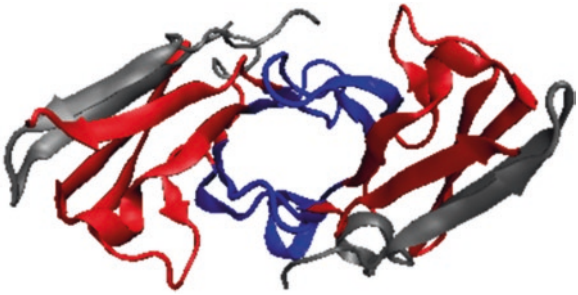

Fig. 5.6 3D visualization of the V domain in a B-J protein (3BJL), showing fragments which diverge from the theoretical hydrophobicity distribution: *gray* – N-terminal fragment; *dark blue* – interface area. VMD program was used to draw the picture [23]

5.6 Hypothetical Amyloidogenesis Mechanism Affecting the V Domain in B-J Proteins

Applying the fuzzy oil drop model to amyloid structures points to linear propagation of local hydrophobicity distributions along the long axis of the fibril. The model proposed in [24] stipulates that β-structural fragments perpendicular to the fibril's axis exhibit the following properties:

1. discordant distribution of hydrophobicity vs. theoretical values (there is no monocentric core which would ensure formation of a globular protein);
2. interspersed peaks and troughs of hydrophobicity observed along the unit β-structural fragment;
3. if the amyloid is composed of identical polypeptides, the identical local distribution is repeated for each unit peptide, with linear propagation of hydrophobicity peaks/troughs along the long fibril axis. Similar linear propagation is observed for β-structural fragments with varying sequences as long as their overall profiles remain similar.

All these conditions can be observed in the 2MVX amyloid [25].

Figure 5.7 summarizes the differences between the observed and theoretical distribution (the latter of which would ensure the formation of a centralized hydrophobic core). Each fragment (including β-structural ones) is sequentially identical to all other folds comprising the amyloid; thus the presented distribution of hydrophobicity can be repeated linearly, along with the corresponding propagation of local peaks and troughs (Fig. 5.8).

If linear propagation of local hydrophobicity profiles is taken as a criterion for identifying amyloid forms, then the presence of such arrangement should be regarded as a seed for further amyloid aggregation. Under certain conditions (e.g. shaking) linear propagation effectively "outcompetes" the standard folding process which usually produces monocentrichydrophobic cores (see titin profile on Fig. 5.3A). The folding process follows the intrinsic hydrophobicity rather than generates the common unicentric construction of hydrophobic core. The resulting amyloidogenesis may affect either the entire protein (domain), or enable multiple proteins to cluster together.

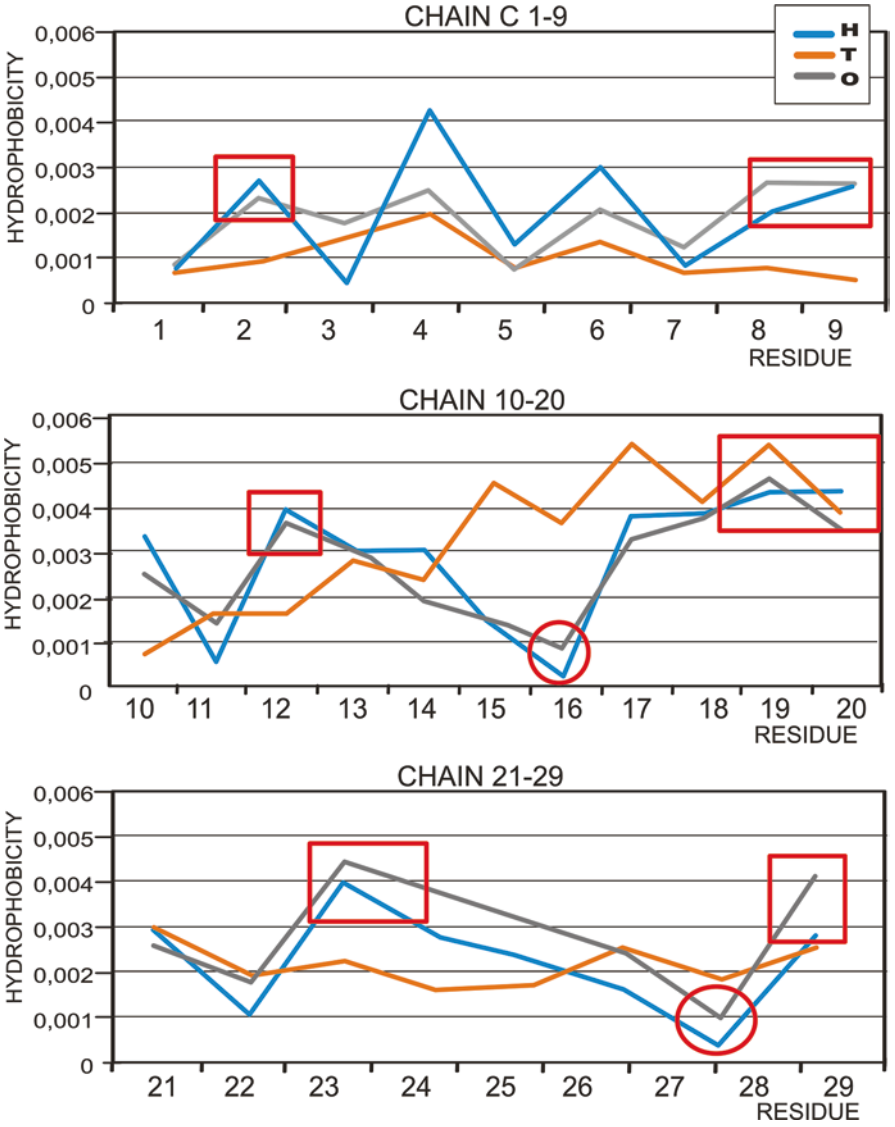

Fig. 5.7 Hydrophobicity distribution profiles in fragments as they appear in 2MVX: H – intrinsic hydrophobicity corresponding to each amino acid; T – theoretical (expected) hydrophobicity given by the FOD model; O – actual (observed) hydrophobicity resulting from inter-residue interactions. *Red* frames mark local peaks, while *red circles* denote local troughs, both of which represent deviations from the theoretical model. Selected fragments are listed above each chart. The profiles shown are identical along the long axis of the amyloidfibril due to identical sequence of peptides generating the fibril

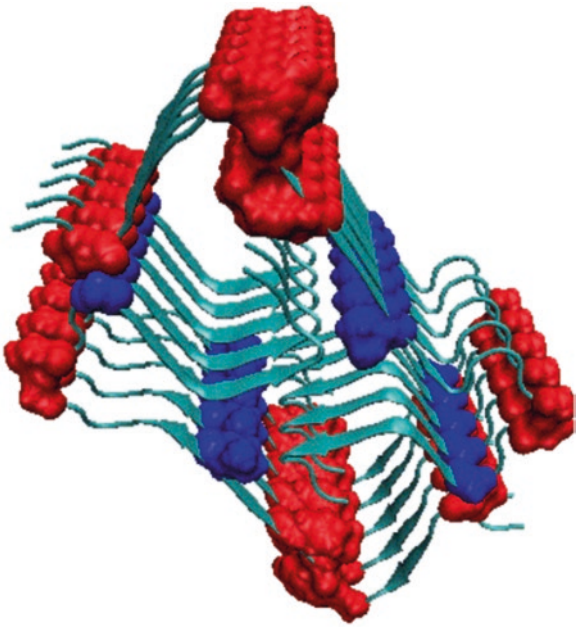

Fig. 5.8 3D presentation of linear propagation of hydrophobicity peaks (*red*) and troughs (*blue*). This arrangement is markedly different from the theoretical distribution, where hydrophobic residues are expected to cluster at the center of the molecule while hydrophilic residues remain exposed on its surface (as shown in Fig.5.7). VMD program was used to draw the picture [23]

The presented distribution is evident in the V domain of the IgG light chain (crystal structure – 3BJL), and particularly its β-structural fragments at 86–90, 40–35, 45–50 and 56–51 (note the antiparallel arrangement). Figure 5.9 presents the distribution of hydrophobicity in each of these fragments.

As shown in Fig. 5.9, some β-structural fragments deviate from the theoretical distribution of hydrophobicity (the RD value for 35–40 equals to 0.676; for 53–60 fragment equals to 0.593). Those which appear to demonstrate the accordant distribution (fragment 43–50 described by RD = 0.463; fragment 86–90 by the RD value = 0.422) in selected positions represent the hydrophobicity level locally different in respect to the expected one. What is more, the specific local profiles of adjacent β-structural fragments enable linear propagation. Figure 5.10 provides a 3D depiction of this phenomenon. Local minima are bracketed by local peaks, all of which propagate linearly and diverge from the theoretical distribution (Fig. 5.10). We can also observe a clear correspondence between the observed distribution of hydrophobicity and the intrinsic hydrophobicity of each participating amino acid. It means no tendency to create the common unicentrichydrophobic core is observed. The residues in polypeptide chain accept the conformation following its intrinsic hydrophobicity.

The 3D visualization shown on Fig. 5.10 corresponds to the profiles shown in Fig. 5.9.

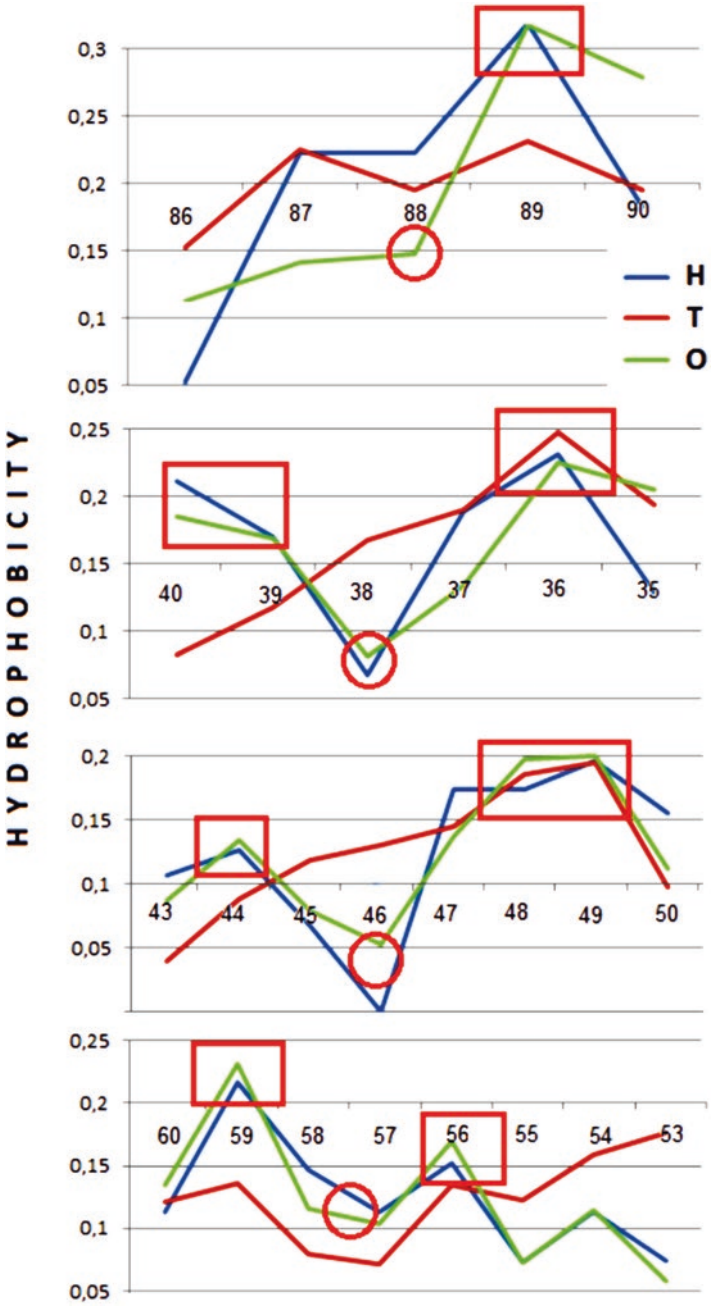

Fig. 5.9 Hydrophobicity distributions in the V domain (H – intrinsic; T – theoretical; O – observed), plotted for fragments which exhibit linear propagation of local troughs (*red circles*) and peaks (*red frames*). The order of residues in fragments accordant to anti-parallel orientation of β-structural fragments

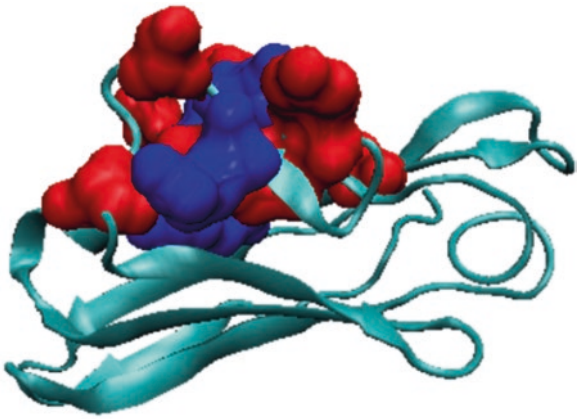

Fig. 5.10 3D presentation of the V domain. *Red* – propagation of hydrophobicity; *blue* – propagation of hydrophilicity. The hydrophilic band in the middle of the β-structural fragment, bracketed by local peaks, provides a seed for linear propagation. The resulting distribution is a poor match for theoretical values (which predict exposure of hydrophilicity on the protein surface). VMD program was used to draw the picture [23]

5.7 Complexation of Congo Red

According to the experiences with CR binding two models can be distinguished:

1. Supramolecular CR may serve as a ligand for any protein, as long as the protein contains a suitable docking cavity. The ligand micelle wedges itself between two adjacent β-structural fragments, as described in [26, 27]. Under these conditions the target protein usually retains its monocentrichydrophobic core, along with any local deviations associated with natural ligand binding capabilities [28]. The supramolecular ligand may occupy the space vacated by a displaced loop, as indeed observed in the case of IgG V domain [21]. The potential ligand binding site can be recognized using FOD calculation as local hydrophobicity deficiency [28]. See also Fig. 2.4.
2. Supramolecular CR is known for its ability to bind to amyloids. An open question concerns the manner in which the dye attaches itself to amyloid aggregations – it can be suspected that the mechanism differs from complexation of individual proteins. In an amyloid, the putative monocentrichydrophobic core is replaced with a distribution of hydrophobicity which reflects the intrinsic properties of each participating residue. If two or more β-structural fragments exhibit similar hydrophobicity profiles, the likelihood of linear aggregation is increased. When the entire V domain converts to a form dominated by linear propagation of hydrophobicity, a multidomain fibril may emerge, as schematically illustrated in Fig. 5.11. It is notable that in such situations the ribbonlike dye micelle complements the linear form of the amyloid itself (which can also be treated as ribbonlike or cylindrical micelle). The orientation of the CR micelle (linear propagation

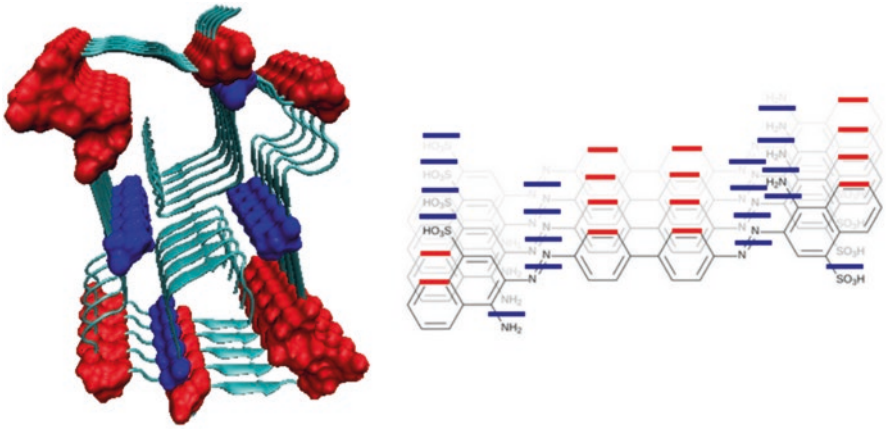

Fig. 5.11 Similarities – linear propagation – in the distribution of local hydrophobicity peaks (*red*) and troughs (*blue*) in an amyloid plaque (2MVX; a4 β-amyloid) and in the CR micelle. VMD program was used to draw the 3D picture [23] and BKChem to draw the CR formula [29]

of local hydrophobicity/hydrophilicity: aromatic rings vs. sulfonic and azo-groups) in respect to linear propagation of hydrophobicity peaks (or troughs) in amyloid fibrils (as shown in Fig. 5.8) appears to be compatible and able to align axially with each other. This observation is not invalidated by the presence of axial twists (as discussed in numerous publications), since in this respect the same property is shared by the amyloid fibril and the dye micelle making mutual adaptation still possible. The liquid crystal form of the CR micelle can align itself to a wide variety of linearized hydrophobicity distributions, which explains why the dye is capable of forming complexes with various amyloid fibrils – both protein-based and peptide-based.

5.8 Hypothetical Amyloidogenesis of V Domains

In summary, we can conclude that – at least according to the fuzzy oil drop model – stabilization of BJ dimers is mediated by C domain complexes. In contrast, the V domains appear quite unstable. RD values calculated for C-C complexes are only slightly above 0.5, while their V-V counterparts are much higher. When structural changes are expected in the dimer, the fuzzy oil drop model points to the V domain as the preferential location of such changes.

The N-terminal fragment adjacent to the interface zone (Fig. 5.5; Tables 5.4 and 5.5) is particularly prone to conformational changes resulting from its poor alignment with the theoretical hydrophobicity distribution. Such changes may provide the seed for amyloid transformation – although the issue is quite complex and requires further study [13, 24].

Figure 5.12 shows a hypothetical mechanism of multimolecular fibril generation. The status of β-structural fragments (as shown in Fig. 5.9) suggests the possible propagation of local maxima and local minima of hydrophobicity in contrast to expected distribution for these fragments. As it is shown in Fig. 5.12. the approach of two units representing similar characteristics is able to make possible propagation of linear hydrophobicity/hydrophilicity propagation. The red fragments on Fig. 5.12 are those shown in Fig. 5.9. According to 3D presentation the fragment 53–60 (distinguished as pink) is expected to fit its structure to the partner from the next unit (domain). Loose N-terminal fragments, devoid of hydrophobic stability (as confirmed by molecular dynamics simulations and experimental studies [21, 26]), may align with one another, creating a new β-interface especially due to its localization on the edge of the domain.

The crystal structure of the V-V dimer does not correspond to the actual conformation of these domains in an amyloidfibril. Certain structural changes are expected in the N-terminal fragment (Fig. 5.6), but also in the domain as a whole (see [13] for a discussion of potential changes expected for fibril formation). In light of this fact, it is difficult to speculate about the final structure of the V-V amyloid – although conformational rearrangements proposed for transthyretin [13], converting its crystal structure into an amyloid, appear equally possible in the IgG V domain (as seen in 3BJL).

The proposed supramolecular CR binding mechanism – one of many possible – is superficial in nature and does not require the dye to penetrate the amyloid. This explains why CR is able to adhere to amyloids formed by separate domains, as well as by identical β-structural fragments comprising a single domain. Due to the specific type of inter-chain interactions occurring in amyloids, intercalation of the dye is unlikely, and while CR may potentially dock in a suitable cavity (as proposed in

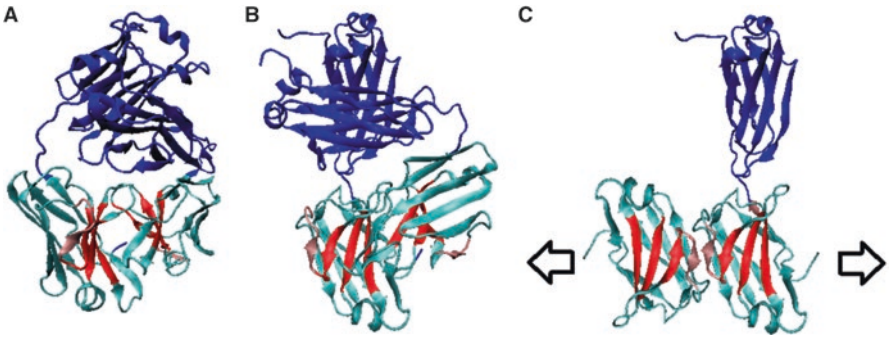

Fig. 5.12 3D presentation of possible interactions between V domains (**A** and **B**) different perspective. (**C**) Possible propagation of β-sheet in amyloid form. *Arrows* indicate possible directions of propagation. The same situation occurs in the complementary chain (second IgG light chain – V domain). *Red* – fragments consistent with the hydrophobicity distribution shown in Fig.5.8 and therefore capable of linear propagation. The *pink* fragment (53–60) requires conformational changes in order to adapt itself to linear propagation of hydrophobicity minima/maxima in β-structural fragments to the next adjacent unit. VMD program was used to draw the picture [23]

[27] see also Fig. 2.4), this mode of complexation is characteristic of individual proteins rather than amyloid fibrils. Finally, the supramolecular CR micelle may also wedge itself between parallel amyloid strains, as observed in 2MVX. Additionally both forms of supramolecular micelle (CR) as well as amyloid by itself are sensitive to external conditions reacting and adopting forms adequate to environmental factors [30].

The presented mechanism of the V domains transformation into the amyloidfibrils can be treated as possible one, obviously assuming that the proposed model of amyloidogenesis as the tendency to linear propagation of hydrophobic characteristics is acceptable. The detergent-like interaction of CR with amyloid described in this chapter is additionally supported by observation reported in [31].

Acknowledgements Work financially supported by Collegium Medicum – Jagiellonian University grant system – grant # K/ZDS/006363.

References

1. Banach M, Konieczny L, Roterman I (2014) The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. J Theor Biol 359:6–17
2. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14:1–63
3. Konieczny L, Brylinski M, Roterman I (2006) Gauss-function-Based model of hydrophobicity density in proteins. Silico Biol 6(1–2):15–22
4. Bernier GM, Putnam FW (1963) Monomer-dimer forms of Bence-Jones proteins. Nature 200:223–225
5. Roussel A, Spinelli S, Déret S, Navaza J, Aucouturier P, Cambillau C (1999) The structure of an entire noncovalent immunoglobulin kappa light-chain dimer (Bence-Jones protein) reveals a weak and unusual constant domains association. Eur J Biochem 260(1):192–199
6. Huang DB, Ainsworth CF, Stevens FJ, Schiffer M (1996) Three quaternary structures for a single protein. Proc Natl Acad Sci U S A 93(14):7017–7021
7. Ely KR, Herron JN, Harker M, Edmundson AB (1989) Three-dimensional structure of a light chain dimer crystallized in water. Conformational flexibility of a molecule in two crystal forms. J Mol Biol 210(3):601–615
8. Huang DB, Ainsworth C, Solomon A, Schiffer M (1996) Pitfalls of molecular replacement: the structure determination of an immunoglobulin light-chain dimer. Acta Crystallogr D Biol Crystallogr 52(Pt 6):1058–1066
9. Makino DL, Henschen-Edman AH, Larson SB, McPherson A (2007) Bence Jones KWR protein structures determined by X-ray crystallography. Acta Crystallogr D Biol Crystallogr 63:780–792

10. Baden EM, Owen BA, Peterson FC, Volkman BF, Ramirez-Alvarado M, Thompson JR (2008) Altered dimer interface decreases stability in an amyloidogenic protein. J Biol Chem 283(23):15853–15860
11. Kenniston JA, Faucette RR, Martik D, Comeau SR, Lindberg AP, Kopacz KJ, Conley GP, Chen J, Viswanathan M, Kastrapeli N, Cosic J, Mason S, DiLeo M, Abendroth J, Kuzmic P, Ladner RC, Edwards TE, TenHoor C, Adelman BA, Nixon AE, Sexton DJ (2014) Inhibition of plasma kallikrein by a highly specific active site blocking antibody. J Biol Chem 289(34):23596–23608
12. Kalinowska B, Banach M, Konieczny L, Roterman I (2015) Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins. Entropy 17(3):1477–1507
13. Roterman I, Banach M, Kalinowska B, Konieczny L (2016) Influence of the aqueous environment on protein structure—a plausible hypothesis concerning the mechanism of amyloidogenesis. Entropy 18(10):351
14. Levitt MA (1974) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104:59–107
15. Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22:79–86
16. Banach M, Prudhomme N, Carpentier M, Duprat E, Papandreou N, Kalinowska B, Chomilier J, Roterman I (2015) Contribution to the prediction of the fold code: application to immunoglobulin and flavodoxin cases. PLoS One 10(4):e0125098
17. Dygut J, Kalinowska B, Banach M, Piwowar M, Konieczny L, Roterman I (2016) Structural interface forms and their involvement in stabilization of multidomain proteins or protein complexes. Int J Mol Sci 17(10):E1741
18. Kalinowska B, Banach M, Konieczny L, Marchewka D, Roterman I (2014) Intrinsically disordered proteins – relation to general model expressing the active role of the water environment. Adv Protein Chem Struct Biol 94:315–346
19. Banach M, Kalinowska B, Konieczny L, Roterman I (2016) Sequence-to-structure relation in proteins-amyloidogenic proteins with chameleon sequences. J Proteomics Bioinform 9:264–275
20. Katzmann JA, Abraham RS, Dispenzieri A, Lust JA, Kyle RA (2005) Diagnostic performance of quantitative kappa and lambda free light chain assays in clinical practice. Clin Chem 51(5):878–881
21. Krol M, Roterman I, Drozd A, Konieczny L, Piekarska B, Rybarska J, Spolnik P, Stopa B (2006) The increased flexibility of CDR loops generated in antibodies by Congo red complexation favors antigen binding. J Biomol Struct Dyn 23(4):407–416
22. Schormann N, Murrell JR, Liepnieks JJ, Benson MD (1995) Tertiary structure of an amyloid immunoglobulin light chain protein: a proposed model for amyloid fibril formation. Proc Natl Acad Sci U S A 92(21):9490–9494
23. http://www.ks.uiuc.edu/Research/vmd/
24. Roterman I, Banach M, Konieczny L (2017) Application of fuzzy oil drop model describes amyloid as ribbonlike micelle. Entropy 19(4):167
25. Schütz AK, Vagt T, Huber M, Ovchinnikova OY, Cadalbert R, Wall J, Güntert P, Böckmann A, Glockshuber R, Meier BH (2015) Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. Angew Chem Int Ed Engl 54(1):331–335
26. Król M, Roterman I, Piekarska B, Konieczny L, Rybarska J, Stopa B (2003) Local and long-range structural effects caused by the removal of the N-terminal polypeptide fragment from immunoglobulin L chain lambda. Biopolymers 69(2):189–200

27. Roterman I, Rybarska J, Monieczny L, Skowronek M, Stopa B, Piekarska B, Bakalarski G (1998) Congo red bound to α-1-proteinase inhibitor as a model of supramolecular ligand and protein complex. Computer Chem 22:61–70
28. Banach M, Konieczny L, Roterman I (2013) Can the structure of hydrophobic core determine the complexation site ? In: Roterman-Konieczna I (ed) Identification of ligand binding site and protein-protein interaction area. Springer Dordrecht, Heidelberg/New York/London, pp 41–54
29. BKChem (http://bkchem.ziral.org)
30. Serpell LC (2000) Alzheimer's amyloid fibrils: structure and assembly. Biochim Biophys Acta 1502(1):16–30
31. Lendel C, Bolognesi B, Wahlstrőm A, Dobson CM, Gräslund A (2010) Detergent-like interaction of Congo red with the amyloid β-peptide. Biochemistry 49:1358–1360