Structural analysis of the Aβ(15-40) amyloid fibril based on hydrophobicity distribution

Dawid Dułak[1], Mateusz Banach[2], Małgorzata Gadzała[3], Leszek Konieczny[4] and Irena Roterman[2,5✉]

[1]ABB Business Services Sp. z o.o., Warszawa, Poland; [2]Departament of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Kraków, Poland; [3]ACK – Cyfronet AGH, Kraków, Poland, currently: Schibsted Tech Polska Sp. z o. o., Kraków, Poland; [4]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kraków, Poland; [5]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Kraków, Poland

The Aβ42 amyloid is the causative factor behind various neurodegenerative processes. It forms elongated fibrils which cause structural devastation in brain tissue. The structure of an amyloid seems to be a contradiction of protein folding principles. Our work focuses on the Aβ(15-40) amyloid containing the D23N mutation (also known as the "Iowa mutation"), upon which an *in silico* experiment is based. Models generated using I-Tasser software as well as the fuzzy oil drop model – regarded as alternatives to the amyloid conformation – are compared in terms of their respective distributions of hydrophobicity (i.e. the existence of a hydrophobic core). In this process, fuzzy oil drop model parameters are applied in assessing the propensity of selected fragments for undergoing amyloid transformation.

INTRODUCTION

Attempts at protein structure prediction (as well as at identifying the underlying mechanisms of protein folding) have a long history (Dill *et al.*, 2012). The CASP initiative (http://predictioncenter.org) aims to develop various structure prediction methods which exemplify two competing approaches: comparative modeling (Monzon *et al.*, 2017; Zhou *et al.*, 2010; Bystroff *et al.*, 2004) and *ab initio* (new fold) models (Saunders *et al.*, 2002; Liwo *et al.*, 2005; Moult *et al.*, 2016). The former group focuses on sequential similarities under the assumption that identical sequences tend to produce similar secondary folds. This, in addition to acknowledgment of evolutionary factors, enables researchers to narrow down the search to a set of homologous proteins, which, in turn, enables the given structure to be assigned to a specific branch of the evolutionary tree. Homology-based comparative modeling provides clues regarding the properties of peptides which share sequential similarities and biological function (and which are therefore likely to adopt similar conformations). In contrast, *ab initio* (new fold) methods rely entirely on theoretical speculation with no need for comparative analysis. This approach assumes that if the model is correct, it should produce an accurate fold for any arbitrary sequence, with no need to study its evolutionary counterparts.

In to-date editions of the CASP comparative modeling techniques have generally yielded better results, although the search for *ab initio* models is by no means over (Moult *et al.*, 2016).

The longstanding dogma stating that similar sequences should produce similar structures was shaken by the discovery of chameleon sequences – with a length of up to 11 aa – which adopt drastically different conformations depending on their host protein: for example, 1D2E (β-fold) and 2C78 (helix). An entire database of chameleon fragments has since been compiled (Li *et al.*, 2015; Ghozlane *et al.*, 2009).

Existing protein structure prediction methods have not yet yielded satisfactory results, as attested to by leading experts who perceive the need for new solutions to augment their research efforts (Khoury *et al.*, 2005).

Somewhat paradoxically, misfolded proteins – including amyloids – may provide important clues regarding the folding process itself, and explain structural rearrangements which sometimes occur within fully folded proteins (Lührs *et al.*, 2005; Fu *et al.*, 2015). In the process of amyloidogenesis, structural changes may occur despite the lack of mutations. It appears that sequentially identical polypeptides may, under most circumstances, fold in an appropriate manner, producing a biologically active protein, while in some cases they may adopt conformations which differ from their respective native forms. In many cases, these alternative conformations manifest as elongated, fibrillary structures capable of unrestricted growth. In such cases the protein in question forfeits its biological function (whatever it may be), becomes insoluble and resists proteolytic enzymes.

Ever since their discovery, amyloids have been closely linked to a specific type of beta fold referred to as cross-beta (Xu, 2009). Studies which focus on structural properties of amino acids show that while some residues are more likely to form helical folds, others are more commonly found in beta fragments (Ghozlane *et al.*, 2009; Kister, 2015). Based on this observation, conformational rearrangement mechanisms are proposed that involve replication of beta folds and (in the presence

of a hydrogen bond network) may support unrestricted linear propagation. This phenomenon is regarded as the principal driving force behind amyloidogenesis (Eisenberg *et al.,* 2017).

The structure of amyloid fibrils eluded analysis for a long time (they do not crystallize, which precludes X-ray imaging, and are insoluble, preventing the application of classic NMR techniques). Recently, however, solid-state NMR (ssNMR) has been successfully applied in amyloid fibril studies (Tycko, 2011).

The presented analysis focuses on a specific type of amyloid structure referred to as Aβ(15-40) (Sgourakis *et al.,* 2015), itself part of the Aβ42 amyloid (Younkin, 1998). We apply concepts derived from the hydrophobic core theory to characterize this structure – which has already been shown to exhibit linear propagation of alternating bands of high and low hydrophobicity (Roterman *et al.,* 2017; Roterman *et al.,* 2016). Such propagation progresses along the long axis of the amyloid and is notable for lacking an arrestor which would otherwise prevent arbitrary elongation of the resulting structure. Based on these observations we propose a structural rearrangement mechanism which may cause a globular protein to transform into an elongated fibril (Roterman *et al.,* 2016; Roterman *et al.,* 2017).

Globular proteins are characterized by the presence of a monocentric hydrophobic core of hydrophobicity distribution following the 3 D Gauss function. The distribution in proteins includes local deformations. Local hydrophobicity deficiencies usually correspond to ligand binding cavities (Banach *et al.,* 2012a), while excess hydrophobicity – if found on the surface – typically marks a complexation site (Banach *et al.,* 2012b). The fuzzy oil drop model expresses the structure of the protein's hydrophobic core and is applicable to a wide variety of structurally diverse proteins (Kalinowska *et al.,* 2017).

The global discordance introduces the alternative distribution form which in amyloids adopts the linear form producing fibrils (Roterman *et al.,* 2016; Roterman *et al.,* 2017). The presence of a monocentric hydrophobic core promotes solubility – conversely, if the core is replaced by a linear pattern of alternating bands the structure becomes susceptible to unrestricted elongation. This effect may be achieved e.g. by introducing highly hydrophilic residues into the central part of the protein body. When such residues adopt a cooperative (repetitive) pattern, linear ordering becomes a strong possibility (Kalinowska *et al.,* 2017; Banach *et al.,* 2018).

The fuzzy oil drop model may be applied to assess the degree of deformations in the protein's hydrophobic core – up to and including amyloid-like conformations, as discussed in (Roterman *et al.,* 2016; Roterman *et al.,* 2017). A detailed description of the model can be found in (Kalinowska *et al.,* 2015).

This work provides a comparative analysis of Aβ(15-40) peptides capable of adopting non-amyloid conformations. In order to generate a set of starting structures we applied the I-Tasser software (Zhang, 2008); a strong contender in recent editions of the CASP challenge (Yang *et al.,* 2015). We also generated alternative structures based on fuzzy oil drop-driven folding algorithms, which – in addition to optimizing internal force fields – account for the presence of the aqueous solvent, favoring generation of a monocentric hydrophobic core (Konieczny *et al.,* 2006). Each program yielded 5 alternative structures (in accordance with CASP rules) for a total of 10 structures. These structures were compared with one another, and with the reference Aβ(15-40) amyloid. The comparative study was based on fuzzy oil drop criteria, as described in the Materials and Methods section.

The presented analysis should be viewed as a follow-up to our study of the properties of amyloid aggregates (particularly Aβ(15-40)), which acknowledge all structural forms tagged in PDB as amyloid seeds.

MATERIALS AND METHODS

The structure of Aβ(15-40). The atomic model of the Aβ amyloid which includes the so-called "Iowa mutation" (D23N) is listed in PDB under ID 2MPZ (Sgourakis *et al.,* 2015). Its structure has been determined using solid-state NMR and EM, and further confirmed by Roberta modeling (http://robetta.bakerlab.org). The 15-20 fragment of the classic Aβ42 amyloid, as presented in PDB, will be referred to as Aβ(15-20).

The listed structure appears in the form of a superfibril consisting of three individual protofibrils in a triangular configuration. Our analysis concerns the entire superfibril, an individual protofibril as well as an isolated chain (component of the protofibril). Each of these structures is characterized using the same set of parameters, facilitating meaningful comparisons regardless of the composition and size of the structure in question.

Alternatives to the ssNMR structure. In addition to the structure listed under 2MPZ we also generated alternative models for the Aβ(15-40) sequence, using I-Tasser software (https://zhanglab.ccmb.med.umich.edu/I-TASSER), which is highly ranked in the CASP competition (http://predictioncenter.org; https://www.dnastar.com/blog/structural-biology/novafold-and--i-tasser-a-winning-combination-for-protein-structure-prediction-and-analysis). While Roberta (Kim *et al.,* 2004) provides similarly accurate results, it could not be applied in the presented case due to its minimum chain length limitation, which the Aβ(15-40) fragment does not satisfy.

The input for I-Tasser computations was provided by the Aβ(15-40) sequence as listed in 2MPZ, i.e. inclusive of the D23N mutation. I-Tasser program produced five models for the target sequence. All of them are taken for analysis in this paper.

The second software package used to model the presented peptide is a tool based on the fuzzy oil drop model (Kalinowska *et al.,* 2015; Konieczny *et al.,* 2006). The model approaches the protein folding problem by introducing an external force field which accounts for the presence of water. However, rather than model the solvent as a collection of individual molecules, the FOD model treats it as a continuum, mathematically expressed by a 3D Gaussian. The effect of the environment is to direct hydrophobic residues towards the center of the emerging molecule. Folding is therefore assumed to occur inside a suitably defined 3D Gaussian capsule, where the distribution of hydrophobicity is subjected to optimization using the Gaussian as a reference. The models produced on the basis of FOD model visualize the alternative folding scenario with aim-oriented hydrophobic core generation.

The described software operates on the supercomputing resources provided by the Academic Computing Centre Cyfronet AGH (as part of the PL-Grid infrastructure) and was used to generate five alternative conformations for the Aβ(15-40) peptide. During these computations, internal force fields were optimized using the Gromacs software package (Berendsen *et al.,* 1995; http://www.gromacs.org), which is also available at Cyfronet.

FOD model delivered many structures (about 500). The models taken for analysis are those representing the extreme status: the highest and the lowest approach in respect to 3D Gauss function representing the distribution of hydrophobicity.

Comparative analysis of the obtained structures and of ssNMR Aβ(15-40). According to the amyloidosis model presented in (Roterman *et al.*, 2016; Roterman *et al.*, 2017), it is assumed that interaction between the polypeptide and the aqueous solvent plays a crucial role in ensuring correct folding and functioning of proteins. Anomalies which result in misfolded proteins may – under the fuzzy oil drop criteria – be attributed to unusual interactions between the protein and its environment (i.e. the external force field). This is why our comparative analysis bases on the fuzzy oil drop model, which provides a measure of the structural ordering of the protein's hydrophobic core. Similarities between the observed (O) and theoretical (T – given by the 3D Gaussian) distribution are quantified using Kullback-Leibler's divergence entropy formula (Kullback *et al.*, 1951). The resulting parameter provides a way to compare a variety of diverse structural forms, including alternative conformations of a specific sequence, ranked according to their accordance with the theoretical hydrophobic core structure.

The aforementioned parameter, denoted RD (Relative Distance) expresses the distance between O and two reference distributions treated as boundary cases. The first of these is the aforementioned theoretical distribution (T), while the other one, referred to as uniform (or random) ascribes a hydrophobicity value of 1/N to each residue (N being the number of residues comprising the chain). When RD<0.5, the observed distribution is regarded as more closely aligned with T, indicating the presence of a hydrophobic core. In the opposite case – RD≥0.5 – the protein is thought to lack a monocentric core. The entire model is also referred to as RD(T-O-R), which means that O (the observed distribution) is compared against T (perfect 3D Gaussian) and R (no concentration of hydrophobicity at any point within the protein body).

As shown in (Roterman *et al.*, 2016; Roterman *et al.*, 2017), amyloids exhibit a peculiar distribution of hydrophobicity which in no way resembles the monocentric core model. In order to better analyze such structures, we introduce another variant of RD, designated RD(T-O-H), where the "random" distribution is replaced with a distribution reflecting the intrinsic hydrophobicity of each residue in the input chain (denoted H). When O is more closely aligned with H than with T, we may claim that the folding process is dominated by the individual preferences of residues with no cooperative tendency to generate a shared (protein-wide) hydrophobic core. For the same reasons this type of conformation may be regarded as "selfish" – indeed, no common "policy" emerges to define the common centric hydrophobic core – as the result of cooperative participation of all residues.

The observed distribution is a result of hydrophobic interactions between neighboring residues. The force of such interactions depends on the mutual separation between residues, as well as on their intrinsic hydrophobicity, as discussed in (Levitt, 1976). In contrast, the theoretical distribution (T) can be used to derive the expected values of hydrophobicity at any point within the protein body – including at the locations of effective atoms (averaged-out positions of all atoms comprising a given residue).

In order to fully characterize the presented structural forms, we also computed three correlation coefficients (for each structure): HvT, TvO and HvO. High values of these coefficients, particularly in the case of HvO, suggest a strong influence of intrinsic hydrophobicity upon the final distribution observed in the molecule.

All the above parameters were computed separately for the whole complex (superfibril), for an individual protofibril and for an individual chain (subunit of the fibril). In each case it is assumed that the introduced parameters can perform the comparative analysis of all discussed structural forms. In addition, selected chain fragments can be studied to assess their contribution to the final structure – these computations may be useful in determining which fragments cause the fibril to emerge in the first place. Low HvT and TvO coefficients coupled with high values of HvO coefficients may be regarded as an indicator of amyloidogenic potential, where the structure is more closely aligned with H than with T. If high HvO is accompanied by negative HvT and TvO, we may conclude that the given fragment is an amyloid seed. It means that the structure is generated "against" the rules producing the globular form of protein. Such conclusions can be reinforced by high values of RD (T-O-R and particularly T-O-H).

The analysis presented in this work bases on interpreting the above parameters.

All models delivered by I-Tasser program (five of them) are taken as objects of analysis.

Structural analysis using FOD model. The characteristics of superfibril is based on the construction of 3D Gauss function encapsulating the complete superfibril. The status of protofibril is defined based on 3D Gauss function. The individual chain characteristics are concluded from the application of 3D Gauss function constructed for individual chain.

The status of chain as part of the selected form of fibril is defined via considering the chain as part of the object, which in our case is the protofibril. The reference distribution in this case is the distribution as it appears in 3D Gauss function defined for protofibril.

RESULTS

The five structures generated with I-Tasser (all structures produced by this program are present in this analysis – according to CASP rules each participant is allowed to deliver 5 models for each target) represent the full spectrum – from globules to near-unfolded forms. The same is true for the structures obtained using the FOD model. The summary of results is given in Table 1.

Structure of the Aβ(15-40) superfibril

The Aβ(15-40) structure listed in PDB, when analyzed from the fuzzy oil model perspective, may be characterized as highly discordant with respect to the monocentric distribution of hydrophobicity. This is visualized in Fig. 1A, where the expected hydrophobicity peak in the central part of the molecule is not replicated by the actual protein. Instead, we face a sinusoidal sequence of alternating peaks and troughs, resulting from symmetrical alignment of identical chain fragments. The observed flattening of peaks in the N- and C-terminal section is not caused by alignment with the monocentric core model, but rather by the fact that the listed structure consists of a finite number of polypeptides.

Structural parameters calculated for the superfibril are as follows: RD(T-O-R)=0.578; RD(T-O-H)=0.494;

Table 1. Characteristics of all the models discussed in the paper (left column),

In the main text, Af is used for description of Aβ(15-40) chain in protofibril, and Ai for status of Aβ(15-40) chain treated as individual structural unit. Parameters based on fuzzy oil drop model expressing RD(T-O-R) and RD(T-O-H) and correlation coefficients for relation are named HvT, TvO and HvO. The parameter values that suggest the status of amyloid character are given in bold. Parameters are given both for the complete chain of Aβ(15-40) and for the chain fragments.

| FORM	FRAGMENT 15-40 RD		FRAGMENT 15-40 Corr. Coeff.			FRAGMENT 14-20 RD		FRAGMENT 14-20 Corr. Coeff.			FRAGMENT 21-28 RD		FRAGMENT 21-28 Corr. Coeff.			FRAGMENT 28-31 RD		FRAGMENT 28-31 Corr. Coeff.			FRAGMENT 34-40 RD		FRAGMENT 34-40 Corr. Coeff.		
	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO	T-O-R	T-O-H	HvT	TvO	HvO
F1	0.188	0.122	0.158	0.918	0.295	0.198	0.129	0.560	0.962	0.660	0.300	0.143	-0.080	0.819	0.474	0.136	0.159	0.595	0.988	0.665	0.227	0.161	-0.443	0.948	-0.463
F2	0.215	0.073	0.208	0.856	0.254	0.232	0.058	0.527	0.825	0.543	0.224	0.315	0.798	0.922	0.903	0.124	0.033	0.278	0.971	0.324	0.304	0.187	-0.458	0.728	-0.685
F3	0.218	0.179	0.208	0.912	0.279	0.253	0.209	0.608	0.905	0.662	0.219	0.356	0.477	0.916	0.596	0.175	0.172	0.575	0.995	0.527	0.258	0.155	-0.503	0.798	-0.684
I1	0.426	0.309	0.375	0.629	0.634	0.240	0.240	0.880	0.960	0.888	0.412	0.587	0.333	0.646	0.834	0.406	0.117	0.283	0.652	0.821	0.764	0.595	0.471	0.473	-0.503
I2	0.466	0.289	0.285	0.551	0.497	0.288	0.160	0.415	0.812	0.815	0.426	0.610	0.174	0.635	0.849	0.204	0.077	0.895	0.925	0.946	0.731	0.672	0.234	0.611	0.138
I3	0.558	0.416	0.255	0.438	0.717	0.371	0.290	0.737	0.616	0.876	0.401	0.506	0.259	0.608	0.892	0.447	0.284	0.353	0.529	0.965	0.915	0.782	0.505	0.227	-0.513
CHAIN S IN FIBRYL Af	**0.567**	**0.564**	0.360	0.345	0.839	0.331	0.350	0.751	0.686	0.962	**0.748**	**0.802**	**-0.248**	**-0.172**	**0.917**	0.284	0.229	0.598	0.844	0.924	0.892	0.730	0.255	0.291	-0.339
F4	0.571	0.544	0.010	0.525	0.480	0.347	0.442	0.697	0.748	0.776	0.530	0.870	0.159	0.296	0.904	0.303	0.270	0.522	0.938	0.686	0.874	0.765	0.317	-0.125	-0.680
F5	0.591	0.553	-0.020	0.447	0.553	0.572	0.738	-0.163	-0.065	0.922	0.382	0.541	0.168	0.711	0.783	0.098	0.075	0.628	0.991	0.716	0.794	0.706	0.404	-0.101	-0.571
I4	0.620	0.387	0.183	0.442	0.552	0.363	0.177	0.523	0.659	0.856	0.295	0.180	0.494	0.844	0.755	0.400	0.171	0.055	0.581	0.690	0.834	0.802	0.171	0.316	0.068
CHAINS INDIVID Ai	0.626	0.466	0.355	0.351	0.615	0.289	0.182	0.788	0.720	0.890	**0.865**	**0.716**	**-0.260**	**-0.222**	**0.838**	0.097	0.038	0.661	0.960	0.826	0.874	0.754	0.246	0.168	-0.504
I5	0.870	0.573	-0.046	-0.006	0.460	0.829	0.188	-0.099	0.089	0.877	0.563	0.240	0.821	0.630	0.559	0.614	0.110	0.486	0.296	0.850	0.942	0.872	0.330	-0.559	-0.266
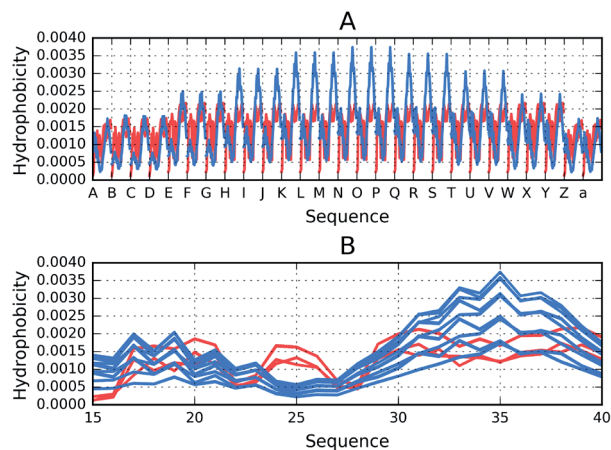
Figure 1. Distribution of hydrophobicity in the Aβ(15-40) super-fibril: T – blue, O – red.
(**A**) profile for consecutive chains as they appear in superfibril. (**B**) overlapped profiles according to residual sequence in all chains.

HvT=0.394; TvO=0.554; HvO=0.790. The RD(T-O-R) value in excess of 0.5 suggests that no central hydrophobic core is present in this structure. On the other hand, RD(T-O-H)<0.5 indicates relatively limited influence of intrinsic hydrophobicity upon the conformation of the superfibril.

We can observe the expected reduction in hydrophobicity in outlying peptides, along with an increase in hydrophobicity in the central chains (Fig. 1B). Eliminating the outlying (edge) chains reveals two distinct hydrophobicity profiles (Fig. 2A). When considering only the chains labeled G through U, a common pattern emerges (Fig. 2B), while the central chains (M, N and O) in each protofibril all share a nearly-identical distribution (Fig. 2C), which diverges from the theoretical profile.
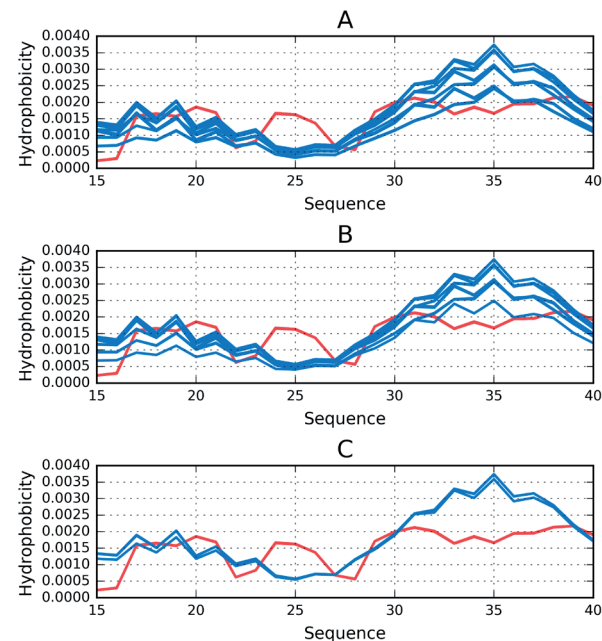
Figure 2. Distribution of hydrophobicity in the Aβ(15-40) superfibril (T – blue, O – red) presented as overlapped profiles according to residual sequence.
(**A**) following elimination of outlying chains in each protofibril: A, a, Z and B, C, D. (**B**) for the internal chains of each protofibril: G – U. (**C**) for the central chains in each protofibril: M, N and O.
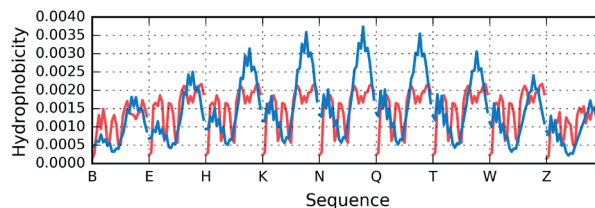
Figure 3. Theoretical (blue) and observed (red) distributions calculated for the Aβ(15-40) protofibril.
Chains are displayed in order as they appear in the protofibril.

Structure of the Aβ(15-40) protofibril

Figure 3 illustrates the distributions of hydrophobicity (T and O) in the Aβ(15-40) protofibril which consists of chains B, E, H, K, N, Q, T, W and Z. In this case, the shape of the Gaussian capsule was adjusted to encapsulate only this protofibril. The observed distribution is characteristic of amyloid structures, with no concentration of hydrophobicity observed in the central part of the fibril. A sinusoidal arrangement of alternating maxima and minima is evident instead.

Structural parameters calculated for the protofibril are as follows: RD(T-O-R)=0.615; RD(T-O-H)=0.600. Both values indicate that the distribution deviates from a monocentric model. HvT, TvO and HvO values are 0.262, 0.412 and 0.788 respectively.

Status of the chain S analyzed as a component of the protofibril

In order to identify potential amyloid seeds, we analyzed chain S as a component of the protofibril.

Distribution of hydrophobicity in chain S as part of protofibril (Fig. 4A) reveals fragments where O not only diverges from T but may even be regarded as its polar opposite. These fragments are also characterized by good alignment between H and O, which shows that their conformation is driven by the properties of individual residues. Correlation coefficients calculated for a 5 aa moving frame further confirm that for certain fragments O opposes T, preventing the formation of a globular
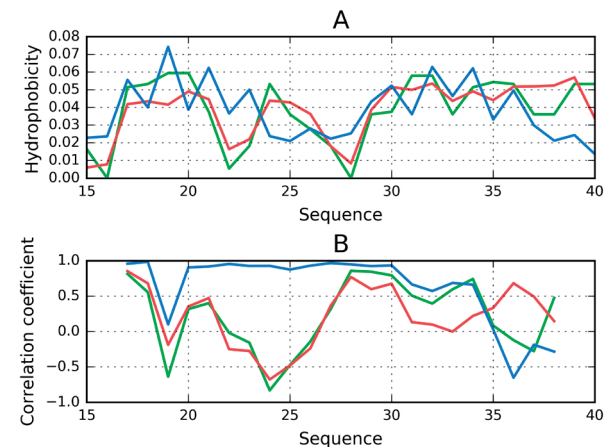
Figure 4. Chain S analyzed as a component of the protofibril.
(**A**) T (blue), O (red) and H (green) hydrophobicity distributions. (**B**) correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i.e. 20 corresponds to residue 20 in the 18-19-20-21-22 frame).
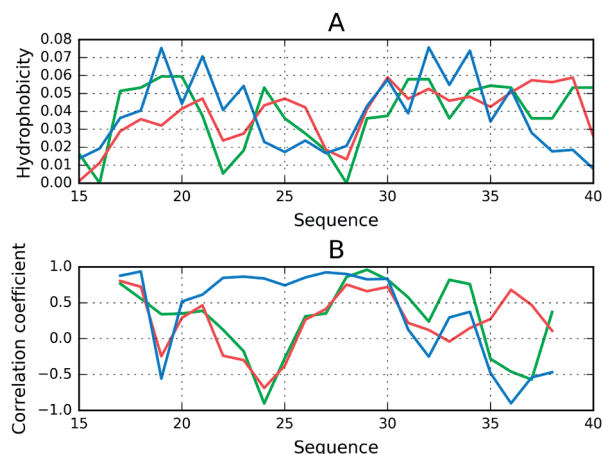
Figure 5. Chain S analyzed as an individual molecule.
(**A**) T (blue), O (red) and H (green) hydrophobicity distributions. (**B**) correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for a 5 aa moving frame (in overlapped system). The indicated position on X axis represents the central residue in a given frame (i.e. 20 corresponds to residue 20 in the 18-19-20-21-22 frame).

structure. As shown in Fig. 4B, the fragment at 21-26 may be regarded as an amyloid seed, with a very high value of HvO and negative values of HvT and TvO.

As a result, fragments 15-21, 22-26, 27-31 and 34-40 have been singled out for individual analysis concerning their adherence to the hydrophobic core model in structures generated by I-Tasser and FOD software.

## Status of the chain S analyzed as an individual molecule

Chain S treated as an individual molecule characterized by FOD-based parameters can be described as follows: RD(T-O-R)=0.626, RD(T-O-H)=0.467, correlation coefficients: HvT=0.355, TvO=0.351, HvO=0.615. The structure is dominated by intrinsic hydrophobicity despite RD(T-O-H) value below 0.5, which does not support this observation. It is the complete set of parameters that reveals that this chain is not any result of uni-centric tendency in folding process.

The visualization of the hydrophobicity distributions shown in Fig. 5A indicates the same polypeptide frag-
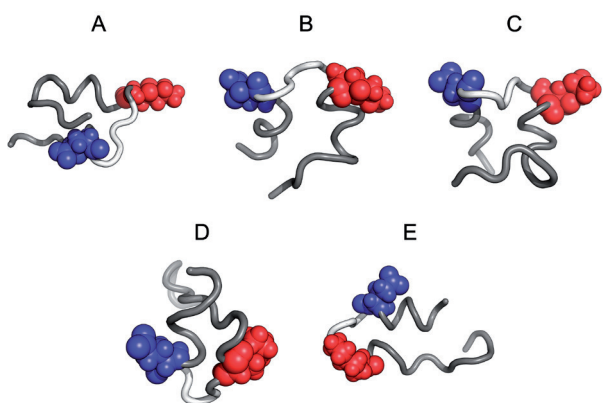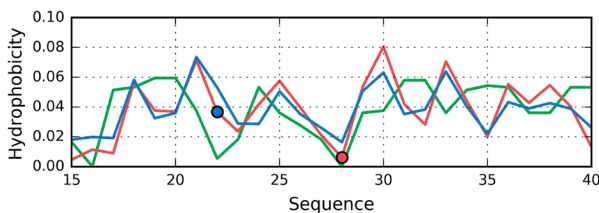
Figure 7. Intrinsic (H – green), theoretical (T – blue) and observed (O – red) distributions of hydrophobicity for structure F1 generated using the fuzzy oil drop model.
The two circles mark observed hydrophobicity values for residues 22Glu (blue) and 28Lys (red).

ments of chain S as highly discordant versus the uni-centric construction of the hydrophobic core.
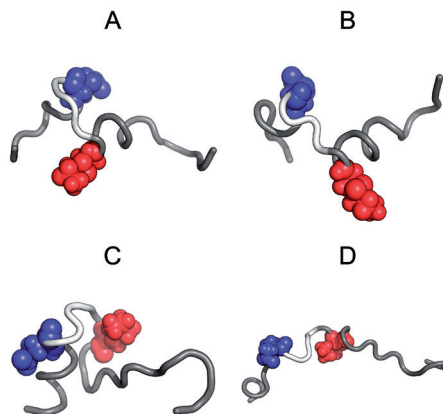
The profiles shown in Fig. 5B resemble the form of chain S treated as part of protofibril.

In summary, the fragments 15-21, 22-26, 27-31 and 34-40 have been singled out for individual analysis concerning their adherence to the hydrophobic core model in structures generated by I-Tasser and FOD software. The detailed analysis of chain fragments is given for all discussed structural forms in Table 1.

Analysis of the above results is that the fragment 22-28 represents the status discordant in respect to expected hydrophobicity distribution. The local maximum observed for this fragment is highly discordant with idealized distribution. The high values for RD and negative values of correlation coefficients for this fragment – as interpreted using the fuzzy oil drop model – classifies this fragment as representing the amyloid status. The status of this fragment is especially traced in all models discussed in this paper. This is why this fragment is distinguished in all 3D presentations of models discussed in this paper (white fragment in the presentations). The positions of residues 22Glu and 28Lys are shown in all discussed structural forms. However, the other fragments in received models appear to represent the status recognized by fuzzy oil drop model as amyloidogenetic.

Comparative analysis of conformations adopted by the Aβ(15-40) sequence

Table 1 provides a list of parameters which characterize all 10 structures generated using I-Tasser and FOD

Figure 6. 3D views of models exhibiting RD(T-O-R) value <0.5. 22Glu is shown in blue and 28Lys in red. Green backbone section corresponds to residues 23-26.
(**A**) Model F1. (**B**) Model F2. (**C**) Model F3. (**D**) Model I1. (**E**) Model I2.

Figure 8. 3D views of models exhibiting RD(T-O-R) value >0.5. 22Glu is shown in blue and 28Lys in red. Green backbone section corresponds to residues 23-26.
(**A**) Model F4. (**B**) Model F5. (**C**) Model I4. (**D**) Model I5.
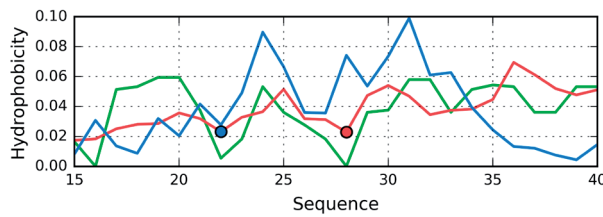
Figure 9. Intrinsic (H – green), theoretical (T – blue) and observed (O – red) distributions of hydrophobicity for a structure I5 generated using I-Tasser.
The two circles mark hydrophobicity values observed for residues 22Glu (blue) and 28Lys (red).

software while acknowledging the status of the 15-40 fragment as a component of the protofibril. The structures were sorted in the order of increasing values of the RD(T-O-R) coefficient.

It is immediately apparent that the presented polypeptide may, in fact, adopt a globular conformation, which corresponds to low values of RD(T-O-R). These low values indicate good agreement between the observed distribution and the 3D Gaussian form, which in turn suggests the presence of a hydrophobic core – as in the case of globular proteins. Figure 6 presents 3D views of some of the structures provided by I-Tasser and FOD.

Somewhat expectedly, structures which possess well-defined hydrophobic cores have been generated by an algorithm which involves optimization of hydrophobic interactions (models F1-F3), however this list also includes two I-Tasser models (I1 and I2), as revealed in Fig. 6.

Structural analysis of models characterized by low RD(T-O-R) reveals "proper" exposure of hydrophilic residues on the protein surface. This observation is further supported by plotting the full distribution of hydrophobicity in a representative structure F1 – Fig. 7.

The hydrophilic residue 28Lys occupies location on the surface, as predicted by the model (Fig. 7). Its sequential neighborhood, characterized by progressive increases of hydrophobicity, is directed towards the central part of the molecule (cf. centrally placed local maxima). We may also observe good alignment between T, O and H for the fragment at 24-30.

Despite "correct" location of the two selected hydrophilic residues (Fig. 8) (exposure on surface in F4, F5, I4 and I5 models), these proteins exhibit a major deviation from the theoretical distributions (Table 1). Figure 8 visualizes the possible different structural forms predicted for Aβ(15-40), where a globular form is obtained, how-
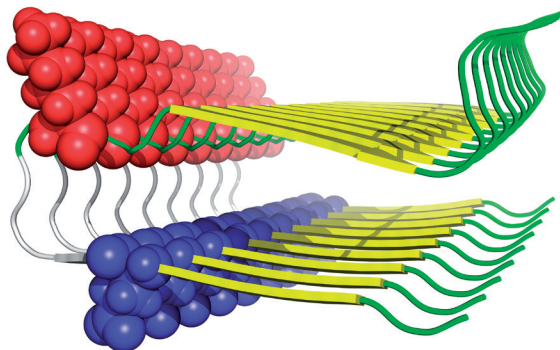
Figure 10. Location of selected residues (22Glu – blue; 28Lys – red) in the protofibril. Green backbone section corresponds to residues 23-26.
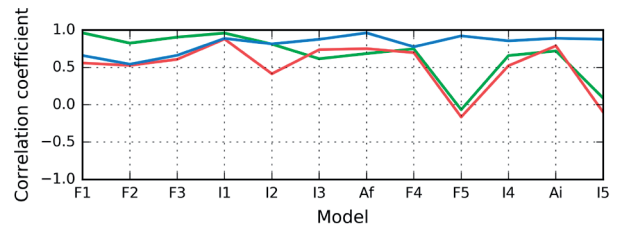

Figure 11. Correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for residues 15-20 in the successive structures as listed in Table 1.
High values of HvO coupled with negative values of HvT and TvO identify these fragments as amyloid-like in F5 and in I5.
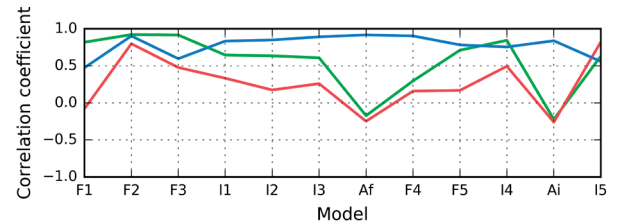
Figure 12. Correlation coefficients (HvO – blue, HvT – red, TvO – green) calculated for residues 21-26 in the successive structures as listed in Table 1.
High values of HvO coupled with negative values of HvT and TvO identify these fragments as amyloid-like in Af (chain in protofibril) and in Ai (chain as individual unit).

ever, these structures do not satisfy the condition of the order expected for soluble proteins. The solubility – according to fuzzy oil drop model – requires the presence of external layer of low hydrophobicity, which is not the case.

Model I5 provides an example of a discordant structure (Fig. 9). While residue 22Glu is correctly positioned, Lys28 is found in an area where a local maximum is expected, generating a distribution which opposes the theoretical model. These conditions may be regarded as conductive to amyloid transformation, with the 26-31
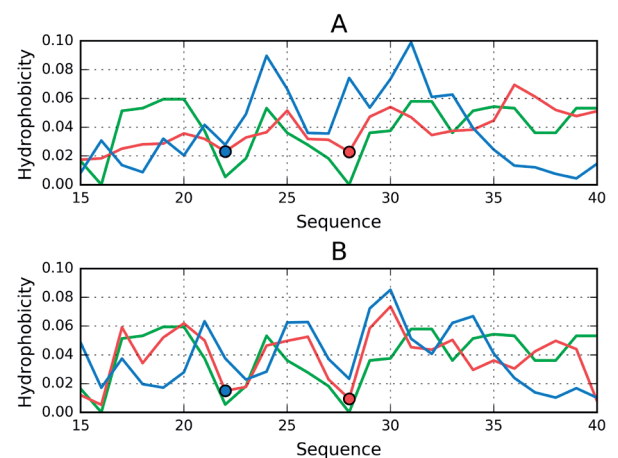
Figure 13. Theoretical (T – blue), intrinsic (H – green) and observed (O – red) distributions of hydrophobicity in models which meet amyloid seed criteria regarding the Aβ(15-40) fragments treated as part of the protofibril. The two circles mark observed hydrophobicity values for residues 22Glu (blue) and 28Lys (red).
(**A**) Model I5. (**B**) Model F5.

fragment, which is expected to form a part of the hydrophobic core, disrupted by 28Lys residue. 3D visualization (Fig. 8) reveals that this residue is located close to the central part of the molecule (more so than residue 22Glu, which is clearly exposed on the surface).

Figure 10 highlights the location of residues 22Glu and 28Lys in the protofibril, enabling visual comparison.

Figure 4A (which illustrates the distribution of hydrophobicity in chains of the protofibril) shows that residue 22Glu is locally discordant, being low in hydrophobicity, whereas the model expects this residue to be found in a highly hydrophobic environment (towards the center of the molecule). The location of residue 28Lys, as shown on Fig. 10, is consistent with the model – this residue is exposed on the surface and takes no part in interactions with the neighboring protofibril. It should be noted that residue 22Glu, while appearing exposed, is in fact internalized, since the chain is adjacent to another polypeptide where residue 22Glu is located close by. Thus, when considering the protofibril as a whole, these locations should be occupied by more hydrophobic residues than Glu.

Table 1 presents the parameters of the entire set of input structures. If we assume that high values of RD (in either model), negative HvT and TvO coefficients, and high values of the HvO coefficients are all indicative of an amyloid-like conformation, then the diagrams shown in Fig. 11 and 12 clearly identify the respective fragments as amyloid seeds.

In the final structures returned by the programs, the amino acids at position 15-20 (6 aa fragment) contains amyloid components in both I5 and F5 models, at least according to the fuzzy oil drop model (Fig. 11, Table 1).

The 21-26 fragment satisfies the criteria under which a given peptide may be regarded as amyloid-like, both in chain S in fibril (Af) and in chain S treated as individual (Ai) (Fig. 12). None of the other structures follow this pattern, which means that in their case the analyzed fragment is not an amyloid seed (although the structures F4 and I3 come close). The remaining fragments do not meet the stated criteria and are not regarded as amyloid seeds.

Figure 13 reveals a discordance between distributions in the 15-20 fragment as it is observed in I5 and F5, which, considering the presented analysis, suggests the presence of a seed of amyloidogenecity.

Structure of the superfibril interface

An additional question related to amyloid structures concerns the formation of superfibrils. Perhaps the most interesting example is the tau amyloid, which is capable of adopting two distinct conformations, differentiated by a mutual alignment of participating protofibrils (Eisenberg *et al.*, 2017).

In case of 2MPZ we observe an arrangement of three individual fibrils, adopting a highly symmetrical (equilateral triangle) form. The fuzzy oil drop model provides an explanation of the mechanisms responsible for this structure. If we calculate RD value for the residues which form part of the stabilizing interface, it turns out that the interface is consistent with the theoretical model of hydrophobicity. This implies that if the entire superfibril is encapsulated in a suitably shaped 3D Gaussian capsule, the placement of residues which comprise the interface is consistent with the fuzzy oil drop model (specifically, the following values were obtained: RD(T-O-R)=0.364;   RD(T-O-H)=0.123;   HvT=0.459; TvO=0.749; HvO=0.847). This set of values suggests

high consistency with the model – in particular, the low value of RD(T-O-H) indicates that the interface emerges as a result of cooperation between individual units and not through intrinsic hydrophobicity. Notably, the interface area stretches along the entire complex, parallel to the fibril. It meshes with the linear aggregation logic that governs the propagation of amyloid fibrils, remaining consistent with the fuzzy oil drop model at the same time. This selective adherence to the model explains the obtained low values of RD for the superfibril, listed in the Results section.

CONCLUSIONS

Summarizing the presented results, it may be useful to inquire why a polypeptide capable of adopting a globular conformation forms an elongated fibril instead. If we base our analysis on the fuzzy oil drop model, the answer is that the fibril emerges as a result of changes in the external (environmental) force field, which, under ordinary conditions, would guide the folding process to produce a globular protein (note that a vast majority of protein domains is consistent with the fuzzy oil drop model and includes a prominent hydrophobic core (Sałapa *et al.*, 2012). The natural environment for Aβ(1-42) fragment is membrane environment. This is why the peptide deprived of its permanent chaperone adopts the unusual structural form. The water environment in its standard order shall be able to direct the folding toward the centric hydrophobic core generation. The final structure of proteins reflects a consensus between internal (inter-atomic) and external forces, the latter being exerted by the aqueous solvent. It seems that the linear propagation of bands of variable hydrophobicity is caused by the alignment between the actual distribution of hydrophobicity and the intrinsic properties of individual residues, with a limited influence of the environment.

The role of the environment is reflected by the properties of superfibril, which – as suggested by its RD values – emerges as a result of the interactions between protofibrils and the aqueous solvent (Brumshtein *et al.*, 2014).

Structures generated using I-Tasser (ranging from highly accordant globules to structures strongly deviating from the theoretical model) suggests that the polypeptide – guided only by its sequence (and, as reflected in the I-Tasser algorithms – by internal interactions between constituent atoms) – may adopt highly variable final conformations. This is further confirmed by the results of the CASP challenge, in which the force field (i.e. the same algorithm) – applied by a particular participant - produced a variety of results when applied to a specific protein. Applying a force field, which directs the folding process towards the generation of a hydrophobic core, also fails to produce a uniform answer – the diversity of the resulting forms (especially regarding the properties of their hydrophobic cores) suggests that the interplay between internal and environmental forces is not accurately captured by either model.

We have discussed other forms of Aβ(1-42) (paper in preparation) and tau-amyloid (Dułak *et al.*, 2018) to analysis similar to the one presented above and we obtained results that are consistent with previous observations in respect to identification of the polypeptide chain fragments which, according to the analysis, seem to play the role of seed for amyloid transformation. The 22Glu-28Lys fragment seems to be the main candidate for amyloid transformation in Aβ(1-42). The results of

presented analyses appear consistent – it turns out that in addition to proteins highly consistent with the fuzzy oil drop model (Banach *et al.,* 2014; Dygut *et al.,* 2016) we can also identify proteins in which the hydrophobic core is locally deformed (Kalinowska *et al.,* 2017), either by local hydrophobicity deficiency (corresponding to the ligand or substrate binding cavity (Banach *et al.,* 2012a) or local excess of hydrophobicity (corresponding to the complexation interface, in the case of proteins which have a quaternary conformation (Banach *et al.,* 2012b). Of particular note are the antifreeze proteins which exhibit broad structural variability, likely associated with their biological role – i.e. disrupting the aqueous environment and thereby preventing formation of ice crystals (Banach *et al.,* 2018).

The central tenet of the fuzzy oil drop model is that the aqueous solvent generates an external force field, affecting the protein in a continuous fashion (rather than as a collection of individual water molecules). The effects can be observed in the case of membranes, which self-organize to produce flat structures consisting of identical pieces. It appears that environmental influence is critical for the formation of amyloid structures, since proper folding depends upon "normal" structuralization of water. This view is based on the observation that environmental changes alone are sufficient to transform the protein from a globule to an amyloid fibril, as observed when shaking samples. Notably, shaking increases the phase transition surface area, leading to increased aeration of the solvent. This non-chemical process alters the structure of the environment, potentially impacting the conformation of solvated proteins. There is ongoing research on the subject, which may explain the peculiar effects of the environment upon the properties of proteins (and upon life in general) (Kim *et al.,* 2017), including the observed levitation of water molecules on top of hydrophobic surfaces (Schutzius *et al.,* 2015).

Acknowledgements

REFERENCES

Banach M, Konieczny L, Roterman I (2012a) Ligand-binding site recognition. In *Protein folding in silico – Protein folding versus protein structure prediction*, Roterman-Konieczna I ed, Woodhead Publishing Oxford Cambridge Philadelphia New Delhi (Currently Elsevier), pp 79–94. doi: 10.1533/9781908818256.79

Banach M, Konieczny L, Roterman I (2012b) Use of the "fuzzy oil drop" model to identify the complexation area in protein homodimers. In *Protein folding in silico – Protein folding versus protein structure prediction*, Roterman-Konieczna I ed, Woodhead Publishing Oxford Cambridge Philadelphia New Delhi (Currently Elsevier), pp 95–122. doi: 10.1533/9781908818256.95

Banach M, Kalinowska B, Konieczny L, Roterman I (2018) Possible mechanism of amyloidogenesis of V domains. In *Self-assembled molecules – new kind of protein ligands – Supramolecular ligands*, Roterman I, Konieczny L eds, Springer Cham, pp 77–100. doi: 10.1007/978-3-319-65639-7

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. doi: 10.1016/j.jtbi.2014.05.007

Banach M, Konieczny L, Roterman I (2018) Why do antifreeze proteins require a solenoid? *Biochimie* 144: 74–84. doi: 10.1016/j.biochi.2017.10.011

Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91: 43–56. doi: 10.1016/0010-4655(95)00042-e

Brumshtein B, Esswein SR, Landau M, Ryan CM, Whitelegge JP, Phillips ML, Cascio D, Sawaya MR, Eisenberg DS (2014) Formation of amyloid fibers by monomeric light chain variable domains. *J Biol Chem* 289: 27513–27525. doi: 10.1074/jbc.M114.585638

Bystroff C, Shao Y (2004) Modeling protein folding pathways. In *Practical Bioinformatics*, Bujnicki J ed, pp. 97–122. Springer Berlin Heidelberg. doi: 10.1007/978-3-540-74268-5_5

Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338: 1042–1046. doi: 10.1126/science.1219021

Dulak D, Gadzała M, Banach M, Ptak M, Wiśniowski Z, Konieczny L, Roterman I (2018) Filamentous aggregates of tau proteins fulfil standard amyloid criteria provided by the fuzzy oil drop (FOD) model. *Int J Mol Sci* 19. pii: E2910. doi: 10.3390/ijms19102910

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: 1741. doi: 10.3390/ijms17101741

Eisenberg DS, Sawaya MR (2017) Taming tangled tau. *Nature* 547: 170–171. doi: 10.1038/nature23094

Fu Z, Aucoin D, Davis J, Van Nostrand WE, Smith SO (2015) Mechanism of Nucleated Conformational Conversion of Aβ42. *Biochemistry* 54: 4197–41207. doi: 10.1021/acs.biochem.5b00467

Ghozlane A, Joseph AP, Bornot A, de Brevern AG (2009) Analysis of protein chameleon sequence characteristics. *Bioinformation* 3: 367–369. doi: 10.6026/97320630003367

Kalinowska B, Banach M, Wiśniowski Z, Konieczny L, Roterman I (2017) Is the hydrophobic core a universal structural element in proteins? *J Mol Model* 23: 205. doi: 10.1007/s00894-017-3367-z

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: 1477–1507. doi: 10.3390/e17031477

Khoury GA, Liwo A, Khatib F, Zhou H, Chopra G, Bacardit J, Bortot LO, Faccioli RA, Deng X, He Y, Krupa P, Li J, Mozolewska MA, Sieradzan AK, Smadbeck J, Wirecki T, Cooper S, Flatten J, Xu K, Baker D, Cheng J, Delbem AC, Floudas CA, Keasar C, Levitt M, Popović Z, Scheraga HA, Skolnick J, Crivelli SN; Foldit Players (2014) WeFold: a coopetition for protein structure prediction. *Proteins* 82: 1850–1868. doi: 10.1002/prot.24538

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32(Web Server), W526–W531. doi: 10.1093/nar/gkh468

Kim KH, Späh A, Pathak H, Perakis F, Mariedahl D, Amann-Winkel K, Sellberg JA, Lee JH, Kim S, Park J, Nam KH, Katayama T, Nilsson A (2017) Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science* 358: 1589–1593. doi: 10.1126/science.aap8269

Kister A (2015) Amino acid distribution rules predict protein fold: protein grammar for beta-strand sandwich-like structures. *Biomolecules* 5: 41–59. doi: 10.3390/biom5010041

Konieczny L, Brylinski M, Roterman I (2006) Gauss-function-Based model of hydrophobicity density in proteins. *In Silico Biol* 6: 15–22

Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22 79–86. doi: 10.1214/aoms/1177729694

Levitt MA (1976) A simplified representation of protein conformations for rapid simulation of protein folding. *J Mol Biol* 104 59–107. doi: 10.1016/0022-2836(76)90004-8

Li W, Kinch LN, Karplus PA, Grishin NV (2015) ChSeq: A database of chameleon sequences. *Protein Sci* 24: 1075–1086. doi: 10.1002/pro.2689

Liwo A, Khalili M, Scheraga HA (2005) Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc Natl Acad Sci U S A* 102: 2362–2367. doi: 10.1073/pnas.0408885102

Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, Döbeli H, Schubert D, Riek R (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A* 102: 17342–17347. doi: 10.1073/pnas.0506723102

Macias-Romero C, Nahalka I, Okur HI, Roke S (2017) Optical imaging of surface chemistry and dynamics in confinement. *Science* 357: 784–788. doi: 10.1126/science.aal4346

Monzon AM, Zea DJ, Marino-Buslje C, Parisi G (2017) Homology modeling in a dynamical world. *Protein Sci* 26: 2195–2206. doi: 10.1002/pro.3274

Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins* 84: 4–14. doi: 10.1002/prot.25064

Roterman I, Banach M, Konieczny L (2017) Application of the fuzzy oil drop model describes amyloid as a ribbonlike micelle. *Entropy* **19**: 167. doi: 10.3390/e19040167

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**: 351. doi: 10.3390/e18100351

Sałapa K, Kalinowska B, Jadczyk T, Roterman I (2012) Measurement of hydrophobicity distribution in proteins – non-redundant protein Data Bank. *Bio-Algorithms and Med-Systems* **8**: 327–338. doi: 10.2478/bams-2012-0023

Saunders JA, Gibson KD, Scheraga HA (2001) Ab initio folding of multiple-chain proteins. *Biocomputing 2002* 601–612. doi: 10.1142/9789812799623_0056

Schutzius TM, Jung S, Maitra T, Graeber G, Köhme M, Poulikakos D (2015) Spontaneous droplet trampolining on rigid superhydrophobic surfaces. *Nature* **527**: 82–85. doi: 10.1038/nature15738

Sgourakis NG, Yau WM, Qiang W (2015) Modeling an in-register, parallel "iowa" aβ fibril structure using solid-state NMR data from labeled samples with rosetta. *Structure* **23**: 216–227. doi: 10.1016/j.str.2014.10.022

Tycko R (2011) Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* **62**: 279–299. doi: 10.1146/annurev-physchem-032210-103539

Xu S (2009) Cross-beta-sheet structure in amyloid fiber formation. *J Phys Chem B* **113**: 12447–12455. doi: 10.1021/jp903106x

Yang J, Zhang Y (2015) Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinformatics* **52**: 5.8.1–5.815. doi: 10.1002/0471250953.bi0508s52

Younkin SG (1998) The role of A beta 42 in Alzheimer's disease. *J Physiol Paris* **92**: 289–292. doi: 10.1016/s0928-4257(98)80035-1

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40. doi: 10.1186/1471-2105-9-40

Zhou H, Skolnick J (2010) Improving threading algorithms for remote homology modeling by combining fragment and template comparisons. *Proteins* **78**: 2041–2048. doi: 10.1002/prot.22717