Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

Article

Filamentous aggregates of tau proteins fulfil standard amyloid criteria

Dawid Dułak[1], Małgorzata Gadzała[2], Mateusz Banach[3], Magdalena Ptak[3,4], Zdzisław Wiśniowski[3], Leszek Konieczny[4] and Irena Roterman[3*]

[1] ABB Business Services Sp. z o.o. ul. Żegańska 1, 04-713 Warszawa, Poland, dawid.dulak@gmail.com

[2] ACK – Cyfronet AGH, Nawojki 11, 30-950 Krakow, Poland, currently: Schibsted Tech Polska Sp. z o. o. ul. Armii Krajowej 28, 30-150 Kraków, Poland, m.k.gadzala@gmail.com

[3] Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Łazarza 16, 31-530 Krakow, Poland, mateusz.banach@uj.edu.pl, mywisnio@cyf-kr.edu.pl, myroterm@cyf-kr.edu.pl

[4] Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11 30-348 Kraków, Poland, magdalena.ptak@uj.edu.pl

[5] Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kopernika 7, 31-034 Kraków, Poland, mbkoniec@cyf-kr.edu.pl

* correspondence: myroterm@cyf-kr.edu.pl

Abstract: Abnormal filamentous aggregates formed by tangled tau protein turn out to be classic amyloid fibrils, meeting all criteria defined under the fuzzy oil drop model in the context of amyloid characterization. The model recognizes amyloids as linear structures where local hydrophobicity minima and maxima propagate in an alternating manner along the fibril's long axis. This distribution of hydrophobicity differs greatly from the classic monocentric hydrophobic core observed in globular proteins. Rather than becoming a globule, the amyloid instead forms a ribbonlike (or cylindrical) structure, which can be thought of as a distorted spherical micelle, which in limit form appears to be the ribbon-like micelle.

Keywords: tau protein amyloids, Alzheimer's disease, tauopathy

1. Introduction

The origin of amyloid transformation has attracted scientific attention for more than 35 years – at least since being acknowledged as the cause of various neurodegenerative disorders [1]. The coexistence and mutual relations between Aβ amyloids and tau tangles, resulting in damage and destruction of synapses, is believed to provoke behavioural changes associated with cognitive impairment [2,3].

It is often stated that effective therapy should rely on arresting the formation of tau fibrils [4]. The search for genetic causes singles out APOE4 as the most important risk factor affecting Aβ pathogenesis, mediated by tau proteins [5,6]. The natural capability of tau to bind to tubulin – a hallmark of its biological activity – is due to phosphorylation. Pathological processes associated with Alzheimer's disease, evidenced by abnormal phosphorylation, result not only in

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

disorganization of microtubulin, but also in the appearance of intracellular tau aggregates, referred to as neurofibrillary tangles [7]. Tauopathies are defined as clinically, morphologically and biochemically heterogeneous neurodegenerative diseases characterized by the deposition of abnormal tau protein in the brain [8].

Comprehensive structural analysis, based on the function of shielding tensors of the Cα atoms which relates binding affinities and folding rate constants to secondary conformational preferences, may reveal common patterns of backbone density distribution in amyloidogenic regions of Alzheimer's amyloid β and tau, Parkinson's α-synuclein and prions [9]. Emergence of fibrillary structures is also thought of as the result of involvement of intrinsically disordered proteins, especially at early phases of the folding process [10].

Tau amyloid fibrils are regarded as peculiar due to the existence of two distinct superfibrillary forms: straight filament and paired helical filament. [11]. Two individual protofilaments may form different structures depending on their mutual arrangement in the dimer. The authors of [11] refer to the conformation of the protofilament as "C-shaped". The dimer (superfibril), when analyzed under cryo-EM imaging, resembles two arched C-shaped structures bound back to back. In one form, the structure is symmetrical (with the same residues in both unit molecules involved in complexation) while the other form lacks such symmetry (in this case, the central fragment of one chain makes contact with a fragment which is somewhat closer to the N terminus of the adjacent chain). The structure of this type of amyloid fibril is revealed thanks to cryo-electron microscopy [12].

Due to the unusual nature of tau amyloids, certain occurrences of Alzheimer's disease are characterized as tauopathies [13]. The cause behind these pathologies is thought to be associated with chemical processes including hyperphosphorylation, acetylation, glycation, ubiquitination, nitration and proteolytic cleavage [14]. An interesting property of protein-based fibrillary filaments, differentiating them from other structures, is the low density of β-strands (which tend to dominate in other types of amyloid fibrils, such as Aβ(42)).

According to the analysis presented in [11], dimerization of protofilaments occurs by way of hydrogen bonds forming between adjacent fragments. In its native form, tau in complex with a microtubule adopts a conformation referred to as "natively unfolded". This conformation is highly resistant to spikes in temperature and acidity. [15]

The presented here analysis focuses on microtubule-associated tau neurofibrillary tangle protein, paired helical filaments (consisting of two individual fibrils), individual proto-fibrils as well as individual chains.

Identification and characterization of amyloid structures, as presented in this work, bases on comparative analysis of the structure of globular and fibrillary proteins. The proposed fibrillarization model for globular structures is also discussed in [16, 17]. At the core of the fuzzy oil drop (FOD) model lies the concept of applying a 3D Gaussian form to express the idealized distribution of hydrophobicity in a globular protein. Such distribution has a distinct peak at the geometric center of the globule and then falls off along with distance from the center, reaching almost zero at the surface (i.e. at a distance of 3σ from the center). We have identified proteins whose structure closely corresponds to this theoretical distribution [18,19]. Any local deviations are usually associated with the protein's intended function: local excess of hydrophobicity, if occurring on the surface, usually marks a complexation interface, while hydrophobicity deficiencies tend to

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

correspond to binding cavities, capable of housing ligands (or substrates, in the case of enzymes) [20]. The universality and ubiquity of hydrophobic cores is attested to by a study of a large number of proteins with varying secondary and supersecondary structural characteristics [21]. A specific type of discordance vs. the idealized (monocentric) distribution of hydrophobicity is observed in the case of proteins which contain so-called solenoid fragments, including some antifreeze proteins [22]. Such fragments deviate from a centralized core in favor of a distribution comprising alternating bands of high and low hydrophobicity, propagating along solenoids long axis [22]. However, in addition to solenoids, biologically active proteins also include fragments whose purpose is to restrict unchecked propagation of such linear sequences (thereby preventing dimerization or polymerization), as well as to ensure solubility. The special "caps" are identified in such proteins [22]. Structures which strongly deviate from the monocentric distribution and lack suitable "caps" are prone to amyloid aggregation [16, 17]. In the context of the fuzzy oil drop model such structures can be likened to ribbonlike micelles, capable of unrestricted propagation [23].

A thorough description of the fuzzy oil drop model can be found in [24].

In the fuzzy oil drop model, the emergent structure of the hydrophobic core is though to result from interactions between the protein and the aqueous solvent, which directs hydrophobic residues towards the center of the globule and favors exposure of hydrophilic residues on the surface. As already indicated, local discordances are often associated with the presence of external factors. The fact that amyloid forms do not require mutations to emerge suggests that misfolding is not caused by factors internal to the polypeptide itself. This hypothesis is further supported by the observation that amyloids are dominated by linear patterns of hydrophobicity, themselves dependent on the intrinsic hydrophobicity of each residue, rather than by the cooperative tendency to form a hydrophobic core [16, 17, 21]. Shaking is known to promote amyloidogenesis – and it can hardly be called a chemical factor. Perhaps shaking disrupts the structure of the solvent in such a way as to prevent it from guiding "natural" conformational changes within the protein chain. Alternatively, shaking is notable for aerating the solvent. The resulting increase in the area of the liquid/gas interface may produce structural changes within the solvent itself.

In addition to analysis of the tau protein amyloid, as listed in PDB [25], this work proposes an *in silico* experiment, which involves determining alternative structures which the tau amyloid sequence may attain (using specialized protein folding software, such as Robetta [26, 27] and I-Tasser [28, 29]), and performing folding simulations based on the fuzzy oil drop model. It turns out that the sequence is indeed capable of producing a globular form with a single, monocentric hydrophobic core. Subjecting globular structures to FOD characterization enables us to track changes which result in amyloidogenesis. The work focuses on three distinct structures: 1. the superfibril (seeking the causes behind its structural variability), 2. the protofibril (identifying the characteristic properties of amyloid structures), 3. a single chain participating in the fibril. The presented research is based on observations rooted in the fuzzy oil drop model, specifically, the linear propagation of hydrophobicity in amyloids (which prevents a shared hydrophobic core from forming). As discussed in [16, 17], the presence of alternating bands of high and low hydrophobicity can be regarded as one of the principal indicators of amyloid transformation.

2. Results

3.1. Superfibril

This analysis concerns the amyloid form listed in PDB as 5O3O, 5O3L and 5O3T (pronase-treated paired helical filament in alzheimer's disease brain neurofibrillary tangle protein, paired helical filament-tau, phf-tau, *Homo sapiens*). Fragment: residues 623-695 of tau protein (306-378 according to PDB numbering) Chains A, C, E, G, I, along with their counterparts (B, D, F, H, J) make up the proto-fibrils [30]. In order to characterize individual chains in the context of the superfibril and proto-fibrils, we have singled out chains E and F. These two chains are located in the central part of the fibril and can be regarded as representative of an arbitrarily long structure. This selection also minimizes edge effects caused by the finite width of the complex.

3.1.1. Properties of superfibrils and interfaces – What is the source of different isoforms of tau aggregates?

Table 1 presents the status of amyloid tau structures in terms of RD values, revealing large discordances between T and O profiles in both models (T-O-R and T-O-H). This means that the distribution of hydrophobicity does not involve a central hydrophobic core. Further analysis will reveal that amyloids are dominated by a pattern which consists of alternating bands of high and low hydrophobicity. High values of RD further indicate that the folding process is driven by the intrinsic properties of each residue rather than by a global force field – this is also typical for amyloids [16, 17]. Regarding the hydrophobicity profile correlation coefficients, HvT and TvO lag behind HvO. This is also due to the absence of a central hydrophobic core, which is replaced by linear propagation of narrow "bands". In further sections we will specifically describe locations which exhibit these properties.

Visual comparison of T and O (Figure 1) highlights major differences between these distributions. It should be noted that the chart consists of many overlapping profiles, which means that the distribution of local minima and maxima is replicated in each adjacent chain, resulting in a set of narrow bands, as suggested above.

The fuzzy oil drop model may also be used to predict the properties of shared hydrophobic cores in protein complexes [31]. In order to properly characterize a given complex, it is important to assess the status of its interface. With regard to proto-fibrils, the distribution observed in 5O3T differs from those exhibited by the remaining structures. However the difference is limited only to the structure of interface, which is discussed in this paper. In 5O3O and 5O3L the status is similar and suggests that the superfibril emerges as a result of factors consistent with the fuzzy oil drop model, i.e. under the influence of the aqueous solvent. This interpretation is supported by high values of all correlation coefficients. We may conclude that the interface is shaped by all factors which determine the structure of the complex itself, with major involvement of water.

The picture changes, however, when dealing with 5O3T. Its high value of RD (T-O-R), coupled with negative values of HvT and TvO coefficients and a high value of the HvO coefficient, suggest that in this case the solvent does not play a significant role in complexation.

It should be noted that the status of the interface is computed by taking into account all interface residues in the entire fibril (following protein-protein contact distance criteria of PDBsum [32]). When all three correlation coefficients adopt strongly positive values, we may assume that the structure of the interface represents a compromise between all three hydrophobicity profiles (observed, intrinsic and theoretical). In contrast, negative values of HvT and TvO are understood to

mean that the interface folds "in spite of" fuzzy oil drop model and in consequence in spite of environmental effects which act upon the proto-fibril complex.

Table 1. –RD values and correlation coefficients computed for the superfibril, proto-fibrils and individual chains treated as components of the superfibril.

PDB ID	RD		Correlation coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
5O3O	0.745	0.687	0.012	0.259	0.675
5O3L	0.731	0.669	0.013	0.296	0.646
5O3T	0.724	0.641	0.008	0.301	0.716
Inter-fibril interface					
5O3O	0.401	0.550	0.666	0.754	0.950
5O3L	0.368	0.587	0.696	0.772	0.929
5O3T	0.527	0.386	-0.164	-0.003	0.850
Chains in superfibril					
5O3O					
Chain E	0.741	0.700	0.011	0.237	0.731
Chain F	0.761	0.722	0.010	0.224	0.731
Chains E+F	0.752	0.711	0.013	0.230	0.731
5O3L					
Chain E	0.728	0.675	0.017	0.274	0.696
Chain F	0.747	0.698	0.012	0.265	0.696
Chains E+F	0.741	0.690	0.014	0.268	0.606
5O3T					
Chain E	0.732	0.666	0.034	0.400	0.769
Chain F	0.728	0.666	-0.018	0.136	0.785
Chain E+F	0.730	0.666	0.008	0.273	0.777

The presented characterization concerns chains E and F, which are located centrally and therefore representative of an unrestricted fibril. The calculated values are typical for amyloid forms, and include high values of RD and HvO, along with very low (sometimes even negative) values of HvT and TvO. In 5O3T both individual chains as well as the interface fragment are shaped by intrinsic hydrophobicity rather than by the external environment, which would favor the formation of a monocentric hydrophobic core.

As the status of individual chains (in the context of the superfibril) is largely similar in all structures, we will limit their presentation to 5O3O and 5O3T (Table 2).

Table 2. Status of individual chains treated as components of the superfibril. The presentation of 5O3L is omitted since due to its similarity to 5O3O.

Chain	RD	Correlation coeff.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

	T-O-R		T-O-H		HvT		TvO		HvO	
	5O3O	5O3T	5O3O	5O3T	5O3O	5O3T	5O3O	5O3T	5O3O	5O3T
A	0.780	0.801	0.696	0.681	0.010	0.041	0.154	0.441	0.669	0.710
B	0.787	0.739	0.711	0.600	0.031	-0.022	0.303	0.115	0.636	0.723
C	0.756	0.766	0.715	0.705	0.014	0.038	0.202	0.394	0.731	0.772
D	0.779	0.728	0.741	0.665	0.022	-0.020	0.248	0.114	0.730	0.788
E	0.741	0.732	0.700	0.666	0.016	0.034	0.237	0.400	0.731	0.769
F	0.761	0.728	0.722	0.666	0.010	-0.010	0.224	0.136	0.731	0.785
G	0.733	0.700	0.690	0.628	0.016	0.028	0.265	0.394	0.731	0.770
H	0.745	0.733	0.704	0.670	0.003	-0.017	0.192	0.157	0.732	0.786
I	0.727	0.660	0.613	0.522	0.015	0.022	0.343	0.405	0.652	0.686
J	0.750	0.738	0.634	0.618	-0.019	-0.014	0.137	0.192	0.691	0.707

When analyzing individual chains as components of the superfibril, we arrive at similar RD values. In contrast, when the same chains are analyzed as components of individual proto-fibrils (5O3T), their values differ due to differences in the orientation of each proto-fibril. Negative values of correlation coefficients for HvT and TvO with high value of correlation coefficient for HvO relation suggest that 5O3T is a typical amyloid form.

Figure 1 provides a graphical representation of the superfibril and both chains (E and F) treated as components of the superfibril.

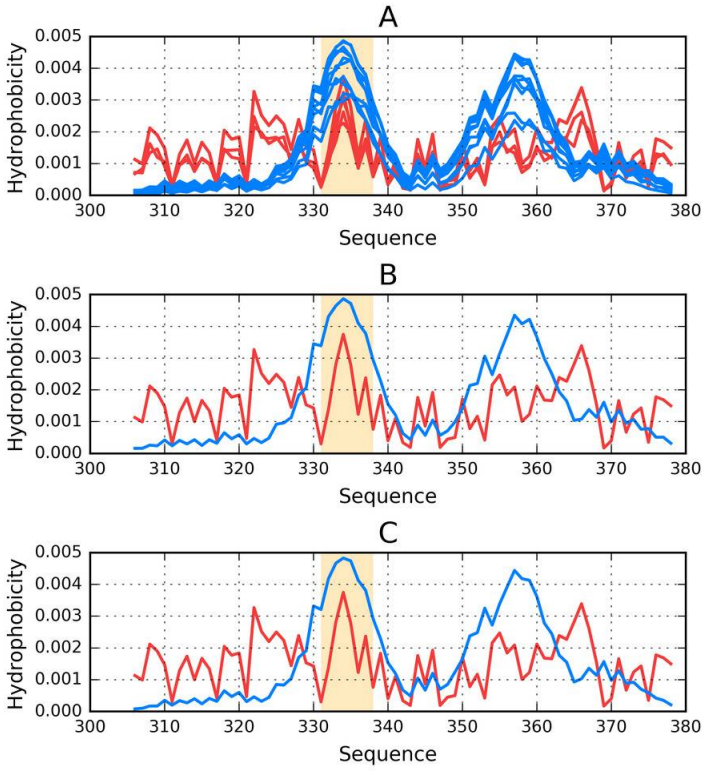

Figure 1. Theoretical (T-blue) and observed (O-red) distributions of hydrophobicity in: A – the 5O3O superfibril; B, C – chains E and F as components of the 5O3O superfibril. Orange highlighting marks residues which comprise the inter-fibril interface area (331-338).

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

The status of 5O3O – superfibril is visualized in Figure 1.A, which shows that the theoretical distribution involves two local maxima, along with hydrophilic fragments exposed on the surface. Neither maximum is evident in the observed distribution, however O includes other local maxima, located in areas where low hydrophobicity is expected. It should be noted that each of these local maxima (as well as minima) represents an entire band stretching along the fibril's long axis. The overlap is due to the repeating pattern present in each individual chain, with only the outlying chains exhibiting slightly lower hydrophobicity. On the other hand, the differences between theoretical distributions are readily apparent since this distribution predicts that hydrophobicity should decrease along with distance from the center. The degree of discordance between T and O can be analyzed by comparing theoretical charts with the observed distributions for chains E and F (which are centrally located and therefore representative for the entire fibril – see Figure 1.B, C).

The interface fragment appears to be consistent with the fuzzy oil model. Given the central location of the interface, a hydrophobicity peak is expected and – to a certain extent – present in the actual complex. Comparing O with T reveals that two outlying residues exhibit relatively low hydrophobicity, while the central section corresponds to a major spike. Consequently, we rate the interface fragment as accordant with the model.

A characteristic feature of amyloids is the presence of numerous local maxima in areas where low hydrophobicity (and vice versa) is predicted by the theoretical model. However, it is important to remember that unlike globular proteins (which may also exhibit this phenomenon), the complexed chains form here bands which stretch along the entire long axis of the fibril. These observations are confirmed by analysis of T and O for chains E and F (treated as components of the superfibril). The discordance between T and O distribution in most of proteins is of local character.

3.2. Properties of proto-fibrils

Table 3 presents the hydrophobicity parameters for each proto-fibril. In this case, each proto-fibril is treated as a distinct structural unit. This means that a separate Gaussian is constructed for each proto-fibril (in the preceding section, a shared Gaussian form was computed for the entire superfibril). Results are indicative of an amyloid form: high values of RD and HvO along with very low (even negative) values of HvT and TvO. The correlation coefficients reveal that the structure is dominated by the intrinsic properties of its component residues – an observation supported by the observed high values of RD in both models (T-O-R and T-O-H). Thus, the observed distribution is more closely aligned with R (or H) rather than T.

The status of chains E and F, treated as components of their respective proto-fibrils confirms that they adopt amyloid-like forms, although this effect is less pronounced than in the case of the superfibril. (lower values of T-O-H RD and HvO – see Table 3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2018 doi:10.20944/preprints201807.0275.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

Table 3. Status of individual chains treated as components of their respective proto-fibrils. The table lists only values obtained for the ACEGI proto-fibril. Differences with regard to the other proto-fibril are negligible.

PDB ID	RD		Correlation coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
5O3O	0.661	0.607	-0.012	0.089	0.772
Chain E	0.679	0.430	-0.027	0.095	0.548
Chain F	0.679	0.439	-0.027	0.095	0.548
5O3L	0.664	0.595	-0.022	0.082	0.767
Chain E	0.674	0.410	-0.039	0.091	0.545
Chain F	0.674	0.410	-0.039	0.091	0.545
5O3T	0.673	0.602	-0.017	0.118	0.773
Chain E	0.683	0.415	-0.033	0.098	0.550
Chain F	0.683	0.415	-0.033	0.096	0.551

Figure 2 provides a visual representation of these results, showing T and O distributions for one proto-fibril (ACEGI), and the status of the E chain within this structural unit.

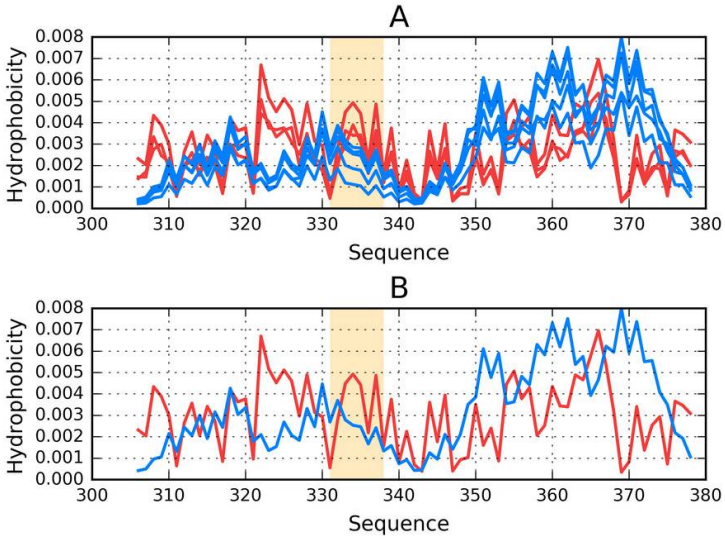

Figure 2. Theoretical (T-blue) and observed (O-red) distributions, A – as calculated for the ACEGI proto-fibril listed under 5O3O, B – for the E chain treated as a component of the proto-fibril. Orange highlighting marks residues which comprise the inter-fibril interface (331-338).

An asymmetrical distribution of local maxima is observed in the proto-fibril as a result of significant displacement of the system's central point compared to the superfibril. Numerous local maxima are present in areas where low hydrophobicity is expected. The involvement of a local maximum in the interface fragment indicates that complexation of proto-fibrils is generated as the effect of the influence of environment (according to FOD model).

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

3.3. Properties of individual chains treated as distinct structural units

Our analysis also covers individual chains treated as distinct structural units, with a separate Gaussian plotted for each chain (under the assumption that each chain folds in separation from other chains). To determine the causes of the discordance between the observed and theoretical distributions, we have singled out fragments for which this discordance is particularly evident. Note that we are not dealing with isolated deviations – in many areas, both distributions strongly oppose each other, indicating that the chain does not produce a globule and is likely insoluble due to the lack of a polar surface.

Figure 3 illustrates the status of chains E and F treated as individual structural units.

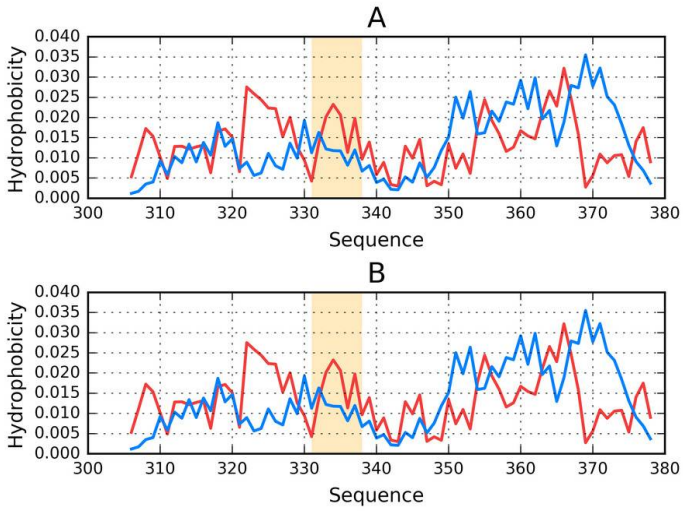

Figure 3. Theoretical (T-blue) and observed (O-red) distributions for 5O3O: A – chain E treated as a distinct structural unit; B – chain F treated as a distinct structural unit. Orange highlighting marks residues which comprise the inter-fibril interface (331-338).

It is clear that even when analyzed as distinct units, the presented chains still diverge from the theoretical distribution of hydrophobicity (Figure 3). No C-terminal maxima (predicted by T) are present in the observed distributions.

Regarding 5O3L, both distributions are similar to those calculated for 5O3O, with only the interface being somewhat different. The observed distribution, while discordant, does not resemble an amyloid (which would appear as a sinusoidal pattern consisting of similar local maxima).

Taking into account the presented distributions it is easy to pinpoint fragments where O deviates from T (see Figure 4).

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

Table 4. RD values in both models (T-O-R and T-O-H), along with HvT, TvO and HvO correlation coefficients for chains E and F analyzed as part of the superfibril, as part of a proto-fibril and on their own. The corresponding values for 5O3L and 5O3R may be found in the Supplementary Materials section. Figure 4 illustrates the division of the chain into individual fragments.

5O3O					
Chain E – superfibril					
Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.521	0.405	0.207	0.270	0.855
16-26	0.820	0.771	-0.126	-0.491	0.874
23-30	0.496	0.495	-0.350	0.435	0.580
39-49	0.575	0.544	0.040	0.147	0.971
50-61	0.851	0.802	-0.106	-0.782	0.446
61-72	0.563	0.491	-0.157	-0.024	0.700
CHAIN	0.741	0.700	0.016	0.237	0.731
Chain F – superfibril					
Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.627	0.512	0.038	0.117	0.875
16-26	0.822	0.773	-0.129	-0.495	0.875
23-30	0.496	0.495	-0.318	0.465	0.580
39-49	0.563	0.533	0.054	0.153	0.971
50-61	0.861	0.814	-0.112	-0.720	0.444
61-72	0.583	0.512	-0.153	0.006	0.699
CHAIN	0.761	0.722	0.010	0.224	0.731
Chain E – proto-fibril					
Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.668	0.553	-0.006	0.064	0.875
16-26	0.615	0.557	0.288	-0.141	0.875
23-30	0.571	0.573	0.146	-0.151	0.810
39-49	0.657	0.623	-0.139	-0.110	0.971
50-61	0.700	0.622	0.016	-0.558	0.443
61-72	0.613	0.540	-0.221	-0.200	0.700
CHAIN	0.661	0.607	-0.012	0.089	0.772
Chain F – proto-fibril					
Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

1-6	0.668	0.553	-0.005	0.064	0.875
16-26	0.615	0.558	0.286	-0.144	0.875
23-30	0.572	0.573	0.143	-0.155	0.809
39-49	0.657	0.623	-0.139	-0.110	0.971
50-61	0.700	0.621	0.016	-0.558	0.442
61-72	0.612	0.539	-0.221	-0.200	0.698
CHAIN	0.661	0.607	-0.012	0.089	0.772

Chain E – individual

Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.741	0.372	-0.093	-0.037	0.543
16-26	0.676	0.433	0.286	-0.387	0.696
23-30	0.582	0.387	0.117	-0.228	0.620
39-49	0.687	0.388	-0.114	0.017	0.888
50-61	0.700	0.489	0.057	-0.537	0.116
61-72	0.656	0.398	-0.237	-0.240	0.507
CHAIN	0.679	0.430	-0.027	0.095	0.548

Chain F - individual

Fragment	RD		Corr coeff		
	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.741	0.372	-0.093	-0.037	0.543
16-26	0.676	0.433	0.286	-0.387	0.696
23-30	0.582	0.387	0.117	-0.228	0.620
39-49	0.687	0.388	-0.114	0.017	0.888
50-61	0.700	0.489	0.057	-0.537	0.115
61-72	0.656	0.397	-0.233	-0.240	0.507
CHAIN	0.679	0.430	-0.027	0.095	0.548

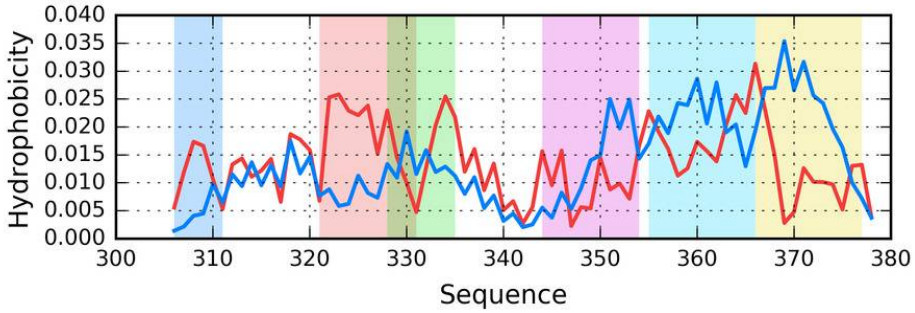

Figure 4. Theoretical (T – blue) and observed (O – red) distribution for the chain E from 5O3O (the distribution is similar in all structural units). Fragments where O significantly deviates from T are highlighted by different colors (see Table 4) that match the colors of 3D presentations on Figure 5.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

Figure 4 provides a sample set of distributions (T and O) for a single chain – E from 5O3O. As already noted, the chains differ in detail, while the overall pattern remains largely identical regardless of the structural unit in question (superfibril, proto-fibril or individual chain). The highlighted fragments have been singled out on the basis of visual inspection, supplemented with correlation coeffeicients computed for successive five-residue segments. Fragments for which HvT and TvO are negative while HvO assumes a large value will be subjected to further analysis.

Table 4 summarizes the results obtained for all structural units in 5O3O.

As shown in Table 4, the status of selected fragments is quite similar regardless of the structural unit in question – in all cases these fragments are strongly discordant vs. the theoretical distribution.

3.4. Comparative analysis involving theoretical models

As previously noted, we have carried out an *in silico* experiment which consisted of predicting the conformation of a tau protein whose sequence matches the presented amyloids. Our analysis concerned the entire molecule as well as fragments highlighted in Figure 4.

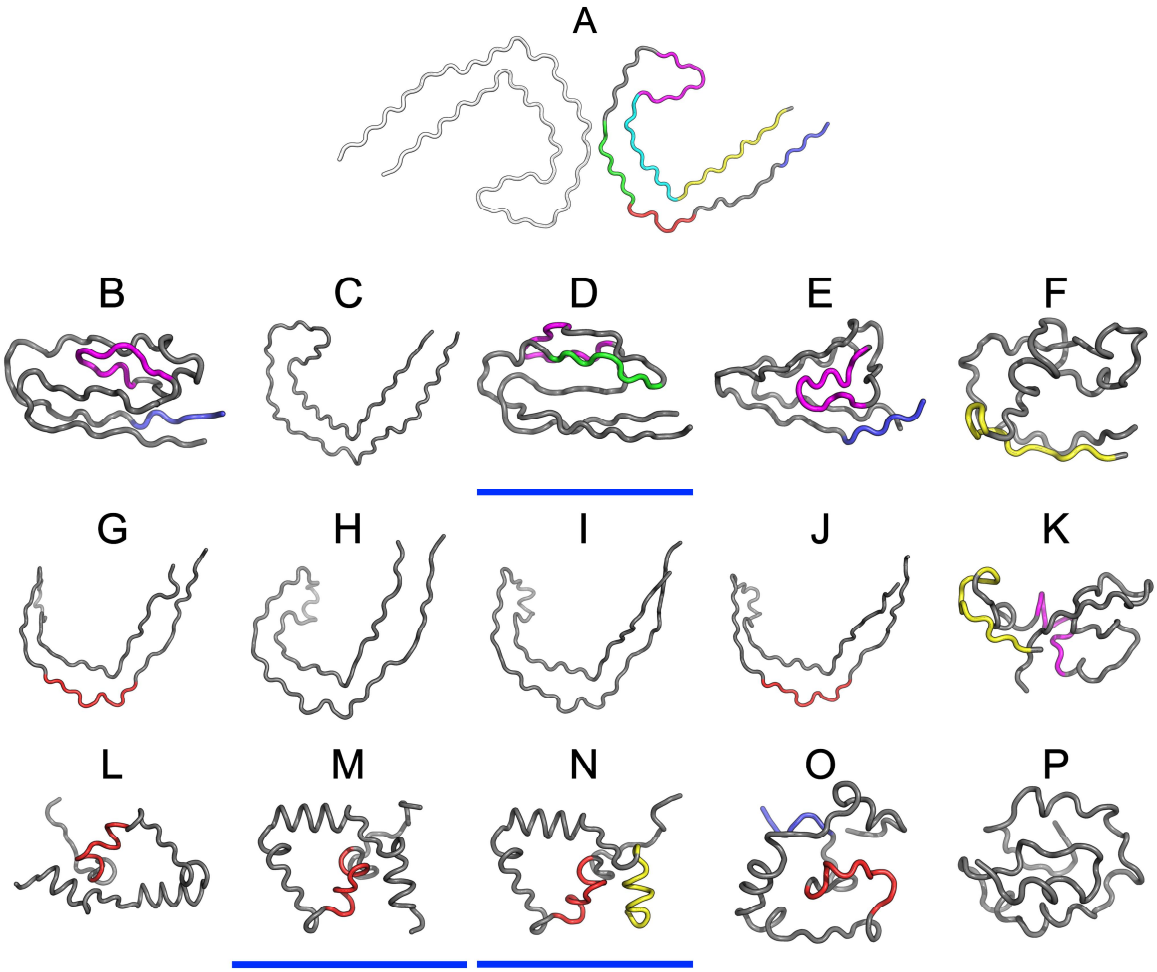

Figure 5. 3D presentation of structures obtained using I-Tasser (upper row B-F), Robetta (central row – G-K) and the fuzzy oil drop model (bottom row – L-P). The top structure (A) is the structure 5O3O as appears in PDB database. The colors used correspond to highlights on Figure 4. Models whose status is consistent with FOD predictions (RD < 0.5) have been additionally underlined in blue (D, M, N).

Figure 5 presents 3D models of tau polypeptides obtained using software packaged described in the Materials and Methods section. Visual inspection reveals the possible emergence of globular forms: I-Tasser produces four such structures (out of five input cases), while Robetta produces one (out of five). The tendency of the fuzzy oil drop model to produce globular forms should come as no surprise given the model's propensity to direct hydrophobic residues towards the center of the molecule (due to interactions with the aqueous solvent).

One of the models produced by I-Tasser appears to involve a hydrophobic core (in the sense of the fuzzy oil drop model – cf. underlined structures in Figure 5). None of the models produced by Robetta satisfies this criterion. Regarding the fuzzy oil drop model, despite its natural tendency to generate hydrophobic cores, only two 500 structures analyzed in the course of the study contain a hydrophobic core (i.e. satisfy the RD < 0.5 condition).

Table 5. RD values obtained using I-Tasser, Robetta and FOD. In addition to RD values (for T-O-R and T-O-H) the table also lists HvT, TvO and HvO correlation coefficients. The division into fragments is consistent with the one presented in Figure 4. Values in bold distinguish the status accordant with FOD model (RD < 0.5).

I-Tasser					
IT-1	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.411	0.142	0.032	0.600	0.469
16-26	0.509	0.200	0.308	0.659	0.706
23-30	0.641	0.357	0.308	0.219	0.415
39-49	0.494	0.283	0.501	0.306	0.841
50-61	0.614	0.618	0.169	0.196	0.466
61-72	0.627	0.232	0.089	0.176	0.559
CHAIN	0.519	0.266	0.200	0.473	0.574
IT-2	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.825	0.138	-0.016	0.551	0.242
16-26	0.607	0.279	0.372	0.039	0.730
23-30	0.503	0.236	0.336	0.226	0.651
39-49	0.765	0.313	-0,230	-0.358	0.849
50-61	0.667	0.548	0.027	-0.416	0.122
61-72	0.605	0.367	-0.486	-0.133	0.475
CHAIN	0.635	0.353	-0.028	0.172	0.519
IT-3	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

1-6	0.729	0.258	-0.438	-0.109	0.416
16-26	0.558	0.179	0.381	0.567	0.830
23-30	0.408	0.173	0.607	0.635	0.805
39-49	0.369	0.191	0.282	0.659	0.828
50-61	0.701	0.702	0.102	-0.196	0.587
61-72	0.545	0.285	0.212	0.162	0.638
CHAIN	**0.478**	0.262	0.270	0.524	0.684
IT-4	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.407	0.183	0.216	0.702	0.554
16-26	0.817	0.230	0.191	0.125	0.730
23-30	0.746	0.468	0.253	-0.453	0.345
39-49	0.432	0.140	0.190	0.494	0.818
50-61	0.630	0.585	0.426	0.094	0.367
61-72	0.601	0.350	0.412	0.193	0.781
CHAIN	0.605	0.315	0.229	0.252	0.581
IT-5	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.545	0.138	-0.024	0.198	0.590
16-26	0.537	0.232	0.321	0.313	0.840
23-30	0.663	0.323	0.254	0.040	0.670
39-49	0.504	0.162	0.444	0.426	0.848
50-61	0.723	0.862	0.194	0.208	0.872
61-72	0.477	0.257	0.501	0.524	0.858
CHAIN	0.513	0.327	0.260	0.418	0.659
			Robetta		
ROB-1	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.627	0.474	0.010	0.165	0.633
16-26	0.473	0.380	0.687	0.400	0.808
23-30	0.566	0.619	0.619	0.157	0.830
39-49	0.606	0.381	0.042	0.275	0.889
50-61	0.650	0.495	0.070	-0.105	0.376
61-72	0.675	0.405	-0.270	-0.182	0.594
CHAIN	0.745	0.568	-0.001	0.109	0.683
ROB-2	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.535	0.345	-0.300	0.218	0.690
16-26	0.601	0.518	0.448	0.004	0.809
23-30	0.621	0.675	0.388	-0.147	0.824
39-49	0.839	0.292	0.014	0.036	0.904
50-61	0.681	0.531	0.083	-0.297	0.404

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

61-72	0.699	0.364	-0.250	0.017	0.552
CHAIN	0.656	0.422	0.032	0.249	0.611
ROB-3	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.600	0.382	-0.495	-0.012	0.623
16-26	0.523	0.410	0.576	0.144	0.795
23-30	0.572	0.510	0.463	-0.038	0.821
39-49	0.842	0.473	-0.173	-0.189	0.955
50-61	0.646	0.512	0.127	-0.168	0.394
61-72	0.698	0.396	-0.182	0.226	0.576
CHAIN	0.682	0.444	0.074	0.351	0.650
ROB-4	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.624	0.519	-0.013	0.141	0.766
16-26	0.441	0.329	0.714	0.464	0.806
23-30	0.593	0.557	0.618	0.171	0.831
39-49	0.782	0.453	-0.120	-0.072	0.939
50-61	0.658	0.575	0.160	-0.187	0.344
61-72	0.678	0.346	-0.187	0.016	0.568
CHAIN	0.748	0.542	0.022	0.154	0.657
ROB-5	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.545	0.111	-0.045	0.571	0.558
16-26	0.701	0.341	0.201	0.166	0.916
23-30	0.748	0.314	0.251	-0.284	0.666
39-49	0.372	0,220	0.255	0.676	0.809
50-61	0.643	0.712	0.258	0.398	0.640
61-72	0.432	0.155	0.513	0.587	0.794
CHAIN	0.566	0.351	0.272	0.375	0.733
			FOD		
FOD-1	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.774	0.288	-0.555	-0.040	0.023
16-26	0.346	0.116	0.238	0.643	0.503
23-30	0.808	0.041	0.510	-0.319	0.339
39-49	0.602	0.187	-0.264	0.267	0.606
50-61	0.733	0.714	0.369	-0.291	0.457
61-72	0.568	0.217	-0.269	0.279	0.200
CHAIN	0.503	0.253	0.126	0.563	0.390
FOD – 2	RD		Correlation coefficient		
Fragment	T-O-R	T-O-H	HvT	TvO	HvO

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 July 2018 doi:10.20944/preprints201807.0275.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

1-6	0.502	0.017	-0.687	0.723	-0.013
16-26	0.315	0.110	0.165	0.770	0.527
23-30	0.675	0.261	0.249	0.105	0.521
39-49	0.543	0.271	-0.175	0.258	0.693
50-61	0.593	0.698	0.375	0.092	0.706
61-72	0.522	0.191	-0.211	0.268	0.268
CHAIN	**0.364**	0.191	0.093	0.665	0.410
FOD-3	RD			Correlation coefficient	
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.605	0.104	-0.373	0.440	0.386
16-26	0.325	0.107	0.054	0.700	0.552
23-30	0.580	0.194	0.328	0.127	0.574
39-49	0.575	0.187	-0.171	0.313	0.689
50-61	0.676	0.682	0.331	-0.282	0.554
61-72	0.338	0.080	0.135	0.603	0.395
CHAIN	**0.369**	0.175	0.152	0.635	0.414
FOD-4	RD			Correlation coefficient	
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.492	0.137	-0.349	0.443	0.231
16-26	0.387	0.136	0.224	0.601	0.546
23-30	0.705	0.470	0.301	-0.414	0.290
39-49	0.589	0.140	-0.211	0.453	0.410
50-61	0.840	0.744	-0.151	-0.366	0.252
61-72	0.556	0.319	0.038	0.092	0.357
CHAIN	0.648	0.346	0.059	0.150	0.395
FOD-5		RD		Correlation coefficient	
Fragment	T-O-R	T-O-H	HvT	TvO	HvO
1-6	0.620	0.075	-0.363	0.484	0.004
16-26	0.513	0.121	0.025	0.576	0.446
23-30	0.668	0.247	-0.113	0.031	0.070
39-49	0.806	0.205	-0.180	0.220	0.646
50-61	0.807	0.588	-0.458	-0.326	0.033
61-72	0.638	0.344	-0.154	-0.022	0.348
CHAIN	0.644	0.313	-0.049	0.200	0.357

Analysis of numerical values listed in Table 5, along with visual inspection of 3D forms reveals that some of these fragments adopt structures consistent with the Gaussian distribution. Of particular note is the fragment marked in red on Figure 5 (residues 50-61, numbered 356-357 according to PDB), which does not conform to theoretical predictions in any model. In interpreting this fact we may refer to the dominant role of the fragment whose sequence does not adapt to the centralized distribution of hydrophobicity. We can speculate that this fragment (GSLDNITHVPGG) is therefore the most amyloidogenic sequence in the presented set. This suggestion is supported by the data

shown in Table 5, particularly RD values which are either highest or second highest in the entire set.

3.5. Other fragments of tau proteins

In order for our comparative analysis to be as comprehensive as possible, we also include tau proteins (or fragments of tau amyloids) which PDB lists as capable of adopting non-amyloid conformations. 1I8H represents the 541-553 fragment of the previously described tau protein, in complex with a microtubule – specifically, pin1 ww domain complexed with human tau phosphothreonine peptide microtubule-associated protein tau. In this complex the A chain comprises residues 541-553 (so-called phf-tau), while chain B represents the ww domain (6-44) [33]. The 1I8H sequence does not fully match 5O3O (and the others), however we have included it in our analysis due to functional similarities.

Table 6. Parameters describing the 541-553 fragment (chain A) of the tau protein 1I8H in complex with the ww domain (residues 6-44).

1I8H	RD		Correlation coefficient		
	T-O-R	T-O-H	HvT	TvO	HvO
Complex	0.680	0.546	0.171	0.235	0.700
Chain A (tau) in complex	0.452	0.152	0.365	0.730	0.756
Chain A(tau) individual	0.630	0.240	0.080	0.348	0.753

Results shown in Table 6 and on Figure 6 reveal that the 1I8H complex does not conform to the fuzzy oil drop model. The status of chain A (tau), when analyzed on its own, is also discordant. On the other hand, the same chain conforms to the model when analyzed as part of the complex. This means that chain B creates suitable conditions for chain A to produce a shared hydrophobic core consistent with the 3D Gaussian.

Considering individual fragments of 1I8H, it turns out that the fragment 27-29 of chain B is the most discordant, and that eliminating this fragment from calculations lowers the RD value of the complex. This may indicate conformational alignment between chain B and chain A (given that the presence of chain A disrupts the distribution of hydrophobicity in chain B).

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

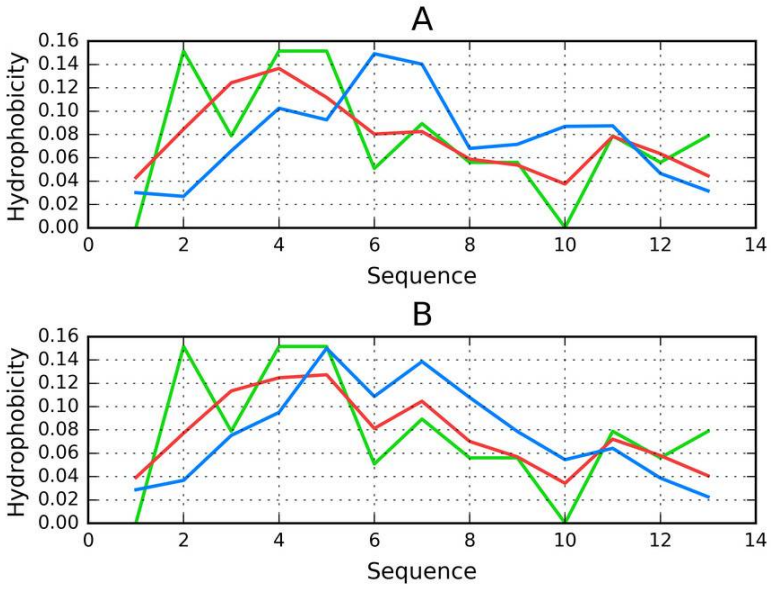

Figure 6. Distribution of hydrophobicity in chain A in the 1I8H complex (T-blue, O-red, H-green): A – when treated as a distinct structural unit; B – when treated as a component of the complex. Good alignment between O, T and H can be observed in the latter case.

Analysis of the presented results indicates the need of a "chaperone", which chain A requires to reach together a conformation consistent with the fuzzy oil drop model.

2MZ7 is another protein related to the presented tau structure. Its sequence matches the short fragment at 306-312 in tau as it is present in 5O3O. According to [34] this fragment (267-312) of tau protein is bound to microtubules.

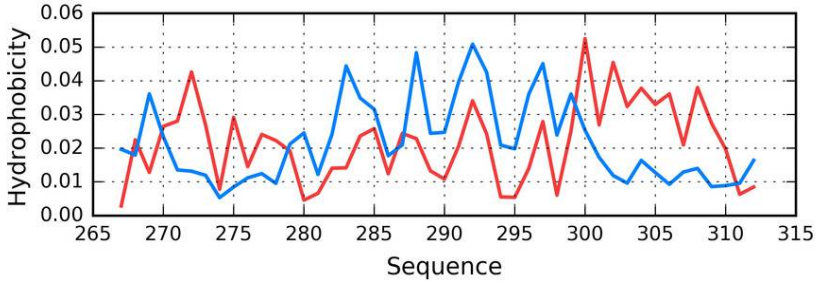

Figure 7. T (blue) and O (red) hydrophobicity distributions in 2MZ7, revealing overall strong discordance.

The status of the tau chain in 2MZ7 reveals strong discordance with regard to the theoretical model (see Figure 7), with RD=0.680 (T-O-R) and 0.527 (T-O-H). Correlation coefficients are -0.049, -0.042 and 0.673 for HvT, TvO and HvO respectively. These values indicate that the structure of the chain is dominated by the conformational tendencies of individual residues rather than by the external hydrophobic force field. Interaction with microtubules is likely the driving force behind conformational adaptation. The structure of the entire complex is not known, however information regarding the interaction of individual residues with the microtubule might

explain the discordance observed throughout the chain. Only the helical fragment at 295-299 appears accordant with the fuzzy oil drop model (with RD values and correlation coefficients of 0.264, 0.170, 0.876, 0.921 and 0.965 respectively, showing good alignment between the theoretical and observed distribution).

3.6. Peptides

To complete our study of tau-derived structures listed in PDB we also need to consider peptides capable of amyloid transformation. The possible mechanism driving this process, discussed in [16, 17], remains applicable in the presented case.

Peptides which match the tau protein sequence are mostly related to the fragment at 306-311 – the short N-terminal fragment of the tau protein present in 5O3O, 5O3L and 5O3T. A short peptide which does not produce a globular form should not, in principle, be analyzed using the fuzzy oil drop model. Nevertheless, for the sake of completeness, we will present RD (T-O-R) values calculated for such peptides – see Table 7.

Table 7. Properties of peptides which match the N-terminal fragment of the presented tau protein. The table lists PDB ID, sequences (indicating which fragments are identical to the tau chain) and RD (T-O-R) values.

PDB ID	Sequence	Position	RD
2ON9	VQIVYK	306-311	0.164
3Q9G	VQIVYK + LA	306-311 + 2 AA	0.332
3OVL	VQIVYK	306-311	0.160
4E0M	AS + VQIVYK+AEFYK	2AA + 306-311 + 5AA	0.707
4E0O	AS + VQIVYK+AEFYK	2AA + 306-311 + 5AA	0.543
4NP8	VQIVYK	306-311	0.174
5K7N	VQIVYK	306-311	0.170

The values shown in Table 7 only reveal the type of hydrophobicity distribution, with no assessment of the hydrophobic core structure. Low values indicate that hydrophobic residues are located in the central part of the chain, surrounded by N- and C-terminal hydrophilic residues. As shown, the short VQIVYK fragment, despite including an outlying Val residue, is a good match for the centralized hydrophobic core structure. When additional neighboring residues (adjacent to the 306-311 fragment) are included in analysis, the value of RD increases significantly.

Peptides identified as capable of amyloidogenesis appear to adopt amyloid-like conformations themselves. As discussed in [16, 17], the distribution of hydrophobicity in a peptide may – regardless of its accordance with the theoretical distribution – give rise to amyloid formation, as long as the environment favors linear complexation of additional peptides, with alternating bands of high and low hydrophobicity emerging along the axis of the fibril. This is visualized in Figure 8, which compares two fringe cases (in terms of RD values).

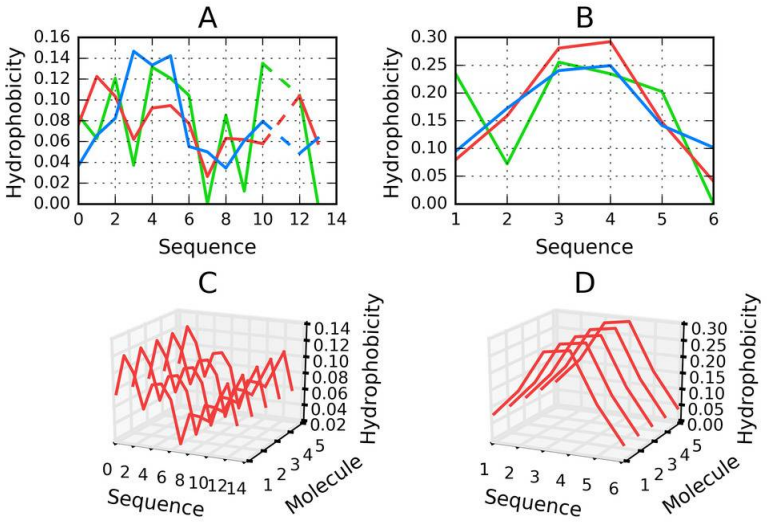

Figure 8. H (green), T (blue) and O (red) distributions for selected peptides: A – 4E0M, B – 3OVL; C (for A) and D (for B) – pseudo-3D view, presenting the observed hydrophobicity of a theorized structure of fibrillary tangles formed by linear propagation of the corresponding peptides. Dashed lines on A represent deleted residue number 11 in the PDB structure of 4E0M.

Figure 8 evidences the appearance of structures characterized by linear propagation of hydrophobicity peaks/troughs, which is a precondition of amyloid formation. As highlighted by to-date observations and interpretations, the environment must "support" the creation of such forms. It is thought that under natural conditions the structure of water does not favor formation of amyloid fibrils.

The value of RD computed for a short peptide implies how many local maxima are present. A low value indicates that hydrophobicity is concentrated in the central part of the peptide (e.g. 3OVL), while a high value suggests the presence of numerous local maxima (e.g. 4E0M).

3. Discussion

The presented comparative analysis of proteins associated with amyloid tau confirms the previously stated hypothesis concerning the structural properties of the amyloid. According to this hypothesis, the amyloid is characterized by the presence of alternating bands of variable hydrophobicity. It seems that linear propagation – which can be regarded as contrary to the emergence of a centralized hydrophobic core (as seen in globular proteins) – is a characteristic property of amyloids. A similar phenomenon can be observed in Aβ amyloids [35] as well as amyloids formed by synuclein [36]. The network of hydrogen bonds discussed in numerous studies [11, 37] favors this type of conformation and is thought to be associated with the linear properties of beta folds. In the tau amyloid, however, β-strands play a much smaller role than in other known amyloids. This suggests that while hydrogen bonds are important, their role is not necessarily linked to β-structures.

Hydrophobicity is capable of binding together proximate charged residues, however, electrostatic interactions should, in principle, prevent such clustering. Under such conditions only hydrophobic forces can result in the presented arrangement. Thus, a conformation which is driven by intrinsic

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

hydrophobicity (and does not generate a central hydrophobic core) may be regarded as both the cause and the mechanism of amyloid transformation.

The fuzzy oil drop model recognizes several possible forms for the tau superfibril. This diversity is likely caused by interactions between the solvent and the emerging amyloid. We suggest that while 5O3O and 5O3L emerge as the effect of the influence of surrounding water, in 5O3T the structure is driven by the specific band-like arrangement of hydrophobicity in the amyloid itself.

The tau protein, whose task is to mediate interaction with microtubules, must align itself to the complexed object. When the protein is subjected to folding on its own, in an independent manner, it may adopt a globular conformation and remain soluble. An open question is why the same protein undergoes complexation in a form which does not resemble a globule. As shown, a chain which is sequentially identical to the amyloid fragment of the tau chain cannot produce a globular structure. In this context, microtubules may be viewed as a „chaperone" which ensures that the protein adopts its intended conformation, required for biological activity.

Conclusions related to the process of amyloidogenesis and the role of the fuzzy oil drop model in explaining this process, all point to the need for further research into the properties of the aqueous solvent. While we possess good knowledge of the properties of ice, the corresponding "normal" (or physiological) condition of liquid water is poorly understood – for example, we are still unsure why the density of water peaks at 4 degrees C. This may explain the recent uptick in investigations which aim to explain such phenomena [37-42]. We believe that these studies may also cast a new light on the process of amyloidogenesis, which – in all likelihood – is associated with the (heretofore unknown) influence of the force field exerted by the surrounding water. This field should be modeled as a continuum rather than (as is common practice in modern molecular dynamics packages) as a collection of distinct molecules. The fuzzy oil drop model provides a good baseline for such research.

Presented analysis allows distinguishing of critical short sequences especially resistant to adopt the conformation accordant with the expected uni-centric hydrophobic distribution. This phenomenon is also observed in other amyloids especially Aβ(1-42) amyloid [35].

4.Materials and Methods

4.1. Data

The analysis concerns tau protein amyloids listed in PDB as capable of forming highly ordered superfibrils. In addition, we also consider selected fragments of the tau protein, including short peptides. Table 8 gives the full list of structures subjected to analysis.

Table 8. Set of proteins subjected to analysis, along with an indication of chain length and complexation capabilities. The rightmost column provides references.

PDB ID	Characteristics	Length	Complex	Reference
	Tau – amyloid			
5O3O	Microtubule-associated protein tau	73 aa	10 chains	[30]
5O3L	Microtubule-associated protein tau	73 aa	10 chains	[30]
5O3T	Microtubule-associated protein tau	73 aa	10 chains	[30]
	Tau – non-amyloid			

2MZ7	Tau (267-312) bound to microtubules	46 aa		[34]
1I8H	Pin1 ww domain complexed with human tau phosphothreonine peptide	14 aa	Complex	[33]
	Peptides			
2ON9	Amyloid forming peptide VQIVYK from the repeat region of tau (in tau 306-311)	6 aa		[43]
3Q9G	VQIVY segment from alzheimer's tau displayed on 42-membered macrocycle scaffold Cyclic pseudo-peptide vqiv(4bf)(orn)(hao)kl(orn)	5 aa		[44]
3OVL	Microtubule-associated protein. VQIVYK (residues 306-311)	6 aa		[45]
4E0M	SVQIVYK segment from human tau (305-311) displayed on 54-membered macrocycle scaffold (form i)	7 aa		[46]
4NP8	Structure of an amyloid forming peptide VQIVYK from the second repeat region of tau (alternate polymorph) (in tau 623-628)	7 aa		[47]
5K7N	Microed structure of tau VQIVYK peptide	6 aa		[48]

Table 8 includes tau superfibrils (5O3L, 5O3O, 5O3T), smaller structural units (including individual chains – 2MZ7), as well as complexes with other proteins (1I8H). We also consider individual peptides which are widely characterized as capable of forming amyloid structures (2ON9, 3Q9G, 3OVL, 4E0M, 4NP8, 5K7N).

All the above structures are subjected to FOD characterization in the context of the superfibril, the protofibril and the individual chain. Our analysis further extends to peptides whose composition is similar or identical to PDB sequences. The status of such molecules is determined by computing their RD coefficients. It should be noted that seeking proper hydrophobic cores in very short peptides (<15 aa) makes little sense – such peptides are characterized using FOD criteria only in order to provide a coherent platform for comparative studies. The fuzzy oil drop model provides useful information regarding the relationship between each residue's intrinsic hydrophobicity and its placement in a fully folded chain.

4.2. Folding of peptides – components of amyloid structures

Peptide sequences which form parts of the tau amyloid (e.g. 306-378, as listed under 5O3O) have been subjected to folding simulations using Robetta [26, 27] and I-Tasser [28, 29], as well as to simulations based on the fuzzy oil drop model [49]. This operation can be regarded as an *in silico* experiment whose aim is to provide alternatives to structures generated by specialized 3D structure prediction software. Our goal is to identify theoretical opportunities for alternative folds (unlike those listed under 5O3O and similar entries). The globular forms generated by the fuzzy oil drop model may provide clues regarding the discordance between the theoretical distribution od hydrophobicity and actual location of hydrophobicity maxima/minima. A ranking list of the resulting structures may be composed in order to identify factors which increase similarities between the theorized conformation and the corresponding amyloid form.

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

Robetta is a software package aimed at modeling and analysis of protein structures [26, 27]. It is a strong performer in successive editions of the CASP challenge, which focuses on predicting the 3D conformations of input residue sequences [50]. Robetta works in the following manner: the user is asked to input a sequence of amino acids comprising a given protein chain. This sequence is then subdivided into fragments (called domains) using the "Ginzu" hierarchical scanning algorithm. The algorithm searchers for fragments homologous to sequences for which the preferred secondary conformation has been established on the basis of experimental studies. Such homogolous areas are detected by (in the order of accuracy) BLAST, PSI-BLAST [51], FFAS03 [52] and 3D-Jury [53] taking as input the sequences produced in the preceding step. The identified domains are modeled by applying a comparative modeling protocol, while all other chain fragments are treated as linkers (if they consist of fewer than 50 residues) or assigned to structural families as defined in the Pfam-A database [54] using HMMER [55]. Fragments and sequences which have not been recognized as putative domains, are analyzed via MSA of the full-length target derived from a PSI-BLAST search against the NCBI non-redundant (NR) protein sequence database [56]. Putative domains identified through Pfam-A and MSA are modeled using *de novo* structure prediction. Finally, following assembly, side chains are modeled by applying Monte Carlo algorithms [57]. The presented description is based on [56].

I-Tasser (Iterative Threading ASSEmbly Refinement) is a software package which can predict the structure of a protein given its sequence. In this application, prediction bases on querying PDB for templates using the multiple threading approach. I-Tasser is a strong contender in CASP challenges, topping the ranking in editions 7 through 12 [58-60].

The user submits a sequence of amino acids, which is then compared (by LOMETS [61]) to template proteins with similar structural characteristics. An optimal template is then selected and overlapping fragments are assembled into an output model using replica-exchange Monte Carlo simulations [62], while differing fragments are modeled *ab initio*. If LOMETS is unable to identify a suitable template, the entire structure is subjected to *ab initio* modeling. The next step involves a search for low energy states (using SPICKER [63]) in the resulting chain via clustering simulation decoys. This is followed by reassembly of the template protein starting with SPICER cluster centroid, however this time the simulation is guided by spatial constraints provided by TM-align on the basis of LOMETS templates and PDB data. The purpose of the second iteration is to remove steric clashes as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are clustered and the lowest energy structures are selected. [64] The final step involves construction of a detailed model from the available structures via optimization of the hydrogen bond network using REMO. Further information can be found in [64].

Robetta computations were carried out using the publicly available service [27]. I-Tasser computations were carried out using [64]. Fuzzy oil drop model computations were carried out using PL-Grid software [65], a detailed description of which can be found in [66].

The fuzzy oil drop model involves two intermediate folding stages: the early-stage intermediate [67-70] and the late-stage intermediate [68-70]. The initial step, which is meant to generate a starting structure for further optimization, is omitted since the conformation listed under 5O3O is taken as the starting structure (treated as early-stage in this case). This chain is then immersed in an aqueous solvent, whose effects are modeled using a 3D Gaussian (as an external force field). In line with the fuzzy oil drop model, hydrophobic residues tend to congregate at the center of the protein body

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 2910; doi:10.3390/ijms19102910

while hydrophilic residues are exposed on its surface. The process produces a prominent hydrophobic core encapsulated by a hydrophilic "shell" (with near-zero values of hydrophobicity on the surface). Optimization of hydrophobic interactions and optimization of nonbinding internal interactions is carried out in an alternating fashion, with each step repeated several times.

Nonbonding interactions are optimized using Gromacs [71] software suite, available on the PL-Grid infrastructure at ACK Cyfronet AGH Kraków [65]. Fuzzy oil drop-based optimization aims to minimize differences between the idealized (3D Gauss function) and observed distribution of hydrophobicity in the target protein. The workflow interleaves both procedures in order to converge on the final conformation.

Folding simulations which rely on the fuzzy oil drop model are relevant since they acknowledge the effects exerted by the aqueous solvent, and treat them as a global phenomenon (i.e. external force field producing a molecule-wide hydrophobic core).

4.3. Comparative analysis

All tertiary conformations produced by the modeling algorithms, as well as structures listed in PDB, were analyzed with regard to the status of their hydrophobic cores, which is described by the RD (relative distance) coefficient. Comparing RD values brings the information about the degree of disorder in respect to ideal distribution. In consequence the approach to amyloid form can be assessed. RD expresses the degree of order present in the protein's hydrophobic core and indirectly indicates whether the protein is globular or not. Generating a globular structure with a prominent hydrophobic core (hydrophobicity peaking at the center of the molecule and decreasing along with distance from the center, becoming very low on the surface) suggests that the given amyloid peptide may, under certain circumstances, adopt a globular conformation. The presented ranking of protein structures, sorted in the order of decreasing globularity, reveals changes which cause proteins to forfeit their centralized hydrophobic cores and which may – in extreme cases – produce amyloid forms.

The RD coefficient can be computed for two independent cases: T-O-R and T-O-H. The former case expressing the relative distance between the observed distribution (O) and two boundary distributions: theoretical (T) which is given by the 3D Gaussian, and uniform (R, random) where each residue is ascribed a hydrophobicity value of 1/N (N being the number of residues in the input chain). R-distribution represents the case of uniform (absence of any local hydrophobicity concentration) distribution which is the opposite one versus the centralized distribution. In the latter case the uniform distribution is replaced by a distribution corresponding to the intrinsic hydrophobicity of each residue in the input chain (H). Comparing both values reveals factors which guide the folding process (this is particularly true in the T-O-H case). A high value of RD (T-O-H) indicates that folding is dominated by the intrinsic properties of each residue with no regard to cooperative generation of a shared hydrophobic core. When this type of distribution is repeated in successive fragment of the polypeptide, the result is a linear sequence of alternating bands of high and low hydrophobicity. This, in turn, enables unrestricted elongation of the fibril. An interpretation of this phenomenon (referred to as "ladders") can also be found in [73].

A comparative assessment of T-O-R and T-O-H coefficients in fibrillary/amyloid structures as well as the structures produced by various folding algorithms may enable us to identify the "seeds" of linear propagation. Fuzzy oil drop criteria have previously been used to assess the distribution of hydrophobicity in structures published by the CASP project [74].

Author Contributions: Conceptualization, I.R., and L.K.; Methodology, D.D., and M.G.; Software, M.B., and M.G.; Validation, D.D. and M.G.; Formal Analysis, I.R.; Investigation, I.R.; Resources, I.R., and M.B.; Writing-Original Draft Preparation, I.R.; Writing-Review & Editing, I.R., and M.B.; Visualization, M.B.; Supervision, L.K.; Project Administration, Z.W; Funding Acquisition, M.P.

Funding: This research was funded by Jagiellonian University – Medical College grant number [K/ZDS/006363 and K/ZDS/006366] and PLGrid Infrastructure – Cyfronet AGH, University of Science and Technology, 30-059 Krakow, Al. Mickiewicza 30, Poland.

Acknowledgments: We are indebted to Anna Śmietańska for technical work and to Piotr Nowakowski for translation services.

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

References

1. Wilcock, G.K., Esiri M.M. Plaques, tangles and dementia. A quantitative study. *J Neurol Sci* **1982**, *56(2-3)*, 343–356, DOI:10.1016/0022-510x(82)90155-1
2. Bloom, G.S. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol* **2014**, *71(4)*, 505–508, DOI:10.1001/jamaneurol.2013.5847
3. Rafii, M.S, Lukic, A.S, Andrews, R.D, Brewer, J., Rissman, R.A., Strother, S.C., Wernick, M.N., Pennington, C., Mobley, W.C., Ness, S., Matthews, D.C. PET Imaging of Tau Pathology and Relationship to Amyloid, Longitudinal MRI, and Cognitive Change in Down Syndrome: Results from the Down Syndrome Biomarker Initiative (DSBI). *J Alzheimers Dis* **2017**, *60(2)*, 439–450**,** DOI:10.3233/JAD-170390
4. Giacobini, E., Gold, G. Alzheimer disease therapy—moving from amyloid-β to tau. *Nat Rev Neurol* **2013**, *9(12)*, 677–686, DOI:10.1038/nrneurol.2013.223
5. Shi, Y., Yamada, K., Liddelow, S.A., Smith, S.T., Zhao, L., Luo, W., Tsai, R.M., Spina, S., Grinberg, L.T., Rojas, J.C., Gallardo, G., Wang, K., Roh, J., Robinson, G., Finn, M.B., Jiang, H., Sullivan, P.M., Baufeld, C., Wood, M.W., Sutphen, C., McCue, L., Xiong, C., Del-Aguila, J.L., Morris, J.C., Cruchaga, C.; Alzheimer's Disease Neuroimaging Initiative, Fagan, A.M., Miller, B.L., Boxer, A.L., Seeley, W.W., Butovsky, O., Barres, B.A., Paul, S.M., Holtzman, D.M. ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **2017**, *549(7673)*, 523–527, DOI:10.1038/nature24016
6. Goedert, M., Eisenberg, D.S., Crowther, R.A. Propagation of Tau Aggregates and Neurodegeneration. A*nnu Rev Neurosci* **2017**, *40(1)*, 189–210**,** DOI:10.1146/annurev-neuro-072116-031153
7. Maccioni, R.B., Muñoz, J.P., Barbeito, L. The Molecular Bases of Alzheimer's Disease and Other Neurodegenerative Disorders. *Archives of Medical Research* **2001**, *32(5)*, 367–381, DOI:10.1016/s0188-4409(01)00316-2
8. Kovacs, G.G. Invited review: Neuropathology of tauopathies: principles and practice. *Neuropathol Appl Neurobiol* **2015**, *41(1)*, 3–23, DOI:10.1111/nan.12208
9. Cieplak, A.S. Protein folding, misfolding and aggregation: The importance of two-electron stabilizing interactions. *PLoS One* **2017**, *12(9)*, e0180905, DOI:10.1371/journal.pone.0180905
10. Fichou, Y., Eschmann, N.A., Keller, T.J., Han, S. Conformation-based assay of tau protein aggregation. *Methods Cell Biol* **2017**, *141*, 89–112, DOI:10.1016/bs.mcb.2017.06.008

641 11. Eisenberg, D.S., Sawaya, M.R. Neurodegeneration: Taming tangled tau. *Nature* **2017**, *547(7662)*, 170–171,
642 DOI:10.1038/nature23094
643 12. Bai, X.C., McMullan, G., Scheres, S.H. How cryo-EM is revolutionizing structural biology. *Trends Biochem*
644 *Sci* **2015**, *40(1)*, 49–57, DOI:10.1016/j.tibs.2014.10.005
645 13. Kovacs, G.G., Ferrer, I., Grinberg, L.T., Alafuzoff, I., Attems, J., Budka, H., Cairns, N.J., Crary, J.F.,
646 Duyckaerts, C., Ghetti, B., Halliday, G.M., Ironside, J.W., Love, S., Mackenzie, I.R., Munoz, D.G., Murray,
647 M.E., Nelson, P.T., Takahashi, H., Trojanowski, J.Q., Ansorge, O., Arzberger, T., Baborie, A., Beach, T.G.,
648 Bieniek, K.F., Bigio, E.H., Bodi, I., Dugger, B.N., Feany, M., Gelpi, E., Gentleman, S.M., Giaccone, G.,
649 Hatanpaa, K.J., Heale, R., Hof, P.R., Hofer, M., Hortobágyi, T., Jellinger, K., Jicha, G.A., Ince, P., Kofler, J.,
650 Kövari, E., Kril, J.J., Mann, D.M., Matej, R., McKee, A.C., McLean, C., Milenkovic, I., Montine, T.J.,
651 Murayama, S., Lee, E.B., Rahimi, J., Rodriguez, R.D., Rozemüller, A., Schneider, J.A., Schultz, C., Seeley,
652 W., Seilhean, D., Smith, C., Tagliavini, F., Takao, M., Thal, D.R., Toledo, J.B., Tolnay, M., Troncoso, J.C.,
653 Vinters, H.V., Weis, S., Wharton, S.B., White, C.L. 3rd, Wisniewski, T., Woulfe, J.M., Yamada, M., Dickson,
654 D.W. Aging-related tau astrogliopathy (ARTAG): harmonized evaluation strategy. *Acta Neuropathol* **2016**,
655 *131(1)*, 87–102, DOI:10.1007/s00401-015-1509-x
656 14. Kolarova, M., García-Sierra, F., Bartos, A., Ricny, J., Ripova, D. Structure and pathology of tau protein in
657 Alzheimer disease. *Int J Alzheimers Dis* **2012**, DOI:10.1155/2012/731526
658 15. Cleveland, D.W., Hwo, S.Y., Kirschner, M.W. Physical and chemical properties of purified tau factor and
659 the role of tau in microtubule assembly. *J Mol Biol* **1977**, *116(2)*, 227–247, DOI:10.1016/0022-2836(77)90214-5
660 16. Roterman, I., Banach, M., Kalinowska, B., Konieczny, L. Influence of the Aqueous Environment on Protein
661 Structure—A Plausible Hypothesis Concerning the Mechanism of Amyloidogenesis. *Entropy* **2016**, *18(10)*,
662 351, DOI:10.3390/e18100351
663 17. Roterman, I., Banach, M., Konieczny, L. Application of the Fuzzy Oil Drop Model Describes Amyloid as a
664 Ribbonlike Micelle. *Entropy* **2017**, *19(4)*, 167, DOI:10.3390/e19040167
665 18. Banach, M., Konieczny, L., Roterman, I. The fuzzy oil drop model, based on hydrophobicity density
666 distribution, generalizes the influence of water environment on protein structure and function. *J Theor Biol*
667 **2014**, *359*, 6–17, DOI:10.1016/j.jtbi.2014.05.007
668 19. Banach, M., Konieczny, L., Roterman, I. Ligand-binding-site recognition. In: I. Roterman-Konieczna (Eds),
669 Protein folding In silico, Woodhead Publishing (currently Elsevier), Oxford, Cambridge, Philadelphia,
670 New Dehli, 2012, pp 79–93, ISBN 978-1-907568-17-6, DOI:10.1533/9781908818256.79
671 20. Banach, M., Konieczny, L., Roterman, I. Use of the fuzzy oil drop model to identify the complexation area
672 in protein homodimers. In: I. Roterman-Konieczna (Eds), Protein folding In silico, Woodhead Publishing
673 (currently Elsevier), Oxford, Cambridge, Philadelphia, New Dehli, 2012, pp 95–122, ISBN 978-1-907568-17-
674 6, DOI:10.1533/9781908818256.95
675 21. Kalinowska, B., Banach, M., Wiśniowski, Z., Konieczny, L., Roterman, I. Is the hydrophobic core a universal
676 structural element in proteins? *J Mol Model* **2017**, *23(7)*, 205, DOI:10.1007/s00894-017-3367-z
677 22. Banach, M., Konieczny, L., Roterman, I. Why do antifreeze proteins require a solenoid? *Biochimie* **2018**, *144*,
678 74-84, DOI:10.1016/j.biochi.2017.10.011
679 23. Banach, M., Konieczny, L., Roterman, I. Secondary and super-secondary structure of proteins in light of the
680 structure of hydrophbic core (in press)
681 24. Kalinowska, B., Banach, M., Konieczny, L., Roterman, I. Hydrophobic Core in DNA Interacting Proteins
682 *Entropy* **2015**, *17(3)*, 1477–1507, DOI:10.3390/e17031477

25 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28(1)*, 235–242, DOI:10.1093/nar/28.1.235

26. Ovchinnikov, S., Park, H., Kim, D.E., DiMaio, F., Baker, D. Protein structure prediction using Rosetta in CASP12. *Proteins* **2018**, *86*, 113–121, DOI:10.1002/prot.25390

27. http://robetta.bakerlab.org (accessed on 05-07-2018)

28. Zhang, Y. I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins* **2009**, *77* (S9) 100–113, DOI:10.1002/prot.22588

29. Yang, J., Zhang, Y. I-TASSER server: new development for protein structure and function predictions, *Nucleic Acids Res* **2015**, *43(W1)*, W174–W181, DOI:10.1093/nar/gkv342

30. Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., Scheres, S.H.W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547(7662)*, 185–190, DOI:10.1038/nature23002

31. Dygut, J., Kalinowska, B., Banach, M., Piwowar, M., Konieczny, L., Roterman, I. Structural Interface Forms and Their Involvement in Stabilization of Multidomain Proteins or Protein Complexes. *Int J Mol Sci* **2016**, *17(10)*, 1741, DOI:10.3390/ijms17101741

32. de Beer, T.A.P., Berka, K., Thornton, J.M., Laskowski, R.A. PDBsum additions. *Nucleic Acids Res* **2014**, *42(D1)*, D292–D296, DOI:10.1093/nar/gkt940

33. Wintjens, R., Wieruszeski, J.M., Drobecq, H., Rousselot-Pailley, P., Buée, L., Lippens, G., Landrieu, I. 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. *J Biol Chem* **2001**, *276(27)*, 25150–25156, DOI:10.1074/jbc.M010327200

34. Kadavath, H., Jaremko, M., Jaremko, Ł., Biernat, J., Mandelkow, E., Zweckstetter, M. Folding of the Tau Protein on Microtubules. *Angew Chem Int Ed Engl* **2015**, *54(35)*, 10347–51, DOI:10.1002/anie.201501714

35. Dułak, D., Gadzała, M., Banach, M., Ptak, M., Konieczny, L., Roterman, I. – Alpha-synuclein amyloid (in preparation)

36. Dułak, D., Gadzała, M., Banach, M., Konieczny, L., Roterman, I. Analysis of the structure of Aβ(15-40) amyloid (in preparation)

37. Chiti, F., Dobson, C.M. Protein misfolding, amyloid formation and human disease: A summary of progress overt the last decade. *Annu Rev Biochem* **2017**, *86(1)*, 27–68, DOI:10.1146/annurev-biochem-061516-045115

38. Kim, K.H., Späh, A., Pathak, H., Perakis, F., Mariedahl, D., Amann-Winkel, K., Sellberg, J.A., Lee, J.H., Kim, S., Park, J., Nam, K.H., Katayama, T., Nilsson, A. Maxima in the thermodynamic response and correlation functions of deeply supercooled water. *Science* **2017**, *358(6370)*, 1589–1593, DOI:10.1126/science.aap8269

39. Gallo, P., Stanley, H.E. Supercooled water reveals its secrets. *Science* **2017**, *358(6370)*, 1543–1544, DOI:10.1126/science.aar3575

40. Palmer, J.C., Martelli, F., Liu, Y., Car, R., Panagiotopoulos, A.Z., Debenedetti, P.G. Metastable liquid-liquid transition in a molecular model of water. *Nature* **2014**, *510(7505)*, 385–388, DOI:10.1126/science.aar3575

41. Tanaka, H. A self-consistent phase diagram for supercooled water. *Nature* **1996**, *380(6572)*, 328–330, DOI:10.1038/380328a0

42. Poole, P.H., Sciortino, F., Essmann, U., Stanley, H.E. Phase behaviour of metastable water. *Nature* **1992**, *360(6402)*, 324–328, DOI:org/10.1038/360324a0

43. Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., Madsen, A.Ø., Riekel, C., Eisenberg, D. Atomic structures of

amyloid cross-beta spines reveal varied steric zippers. *Nature* **2007**, *447(7143)*, 453–457, DOI:10.1038/nature05695

44. Liu, C., Sawaya, M.R., Cheng, P.N., Zheng, J., Nowick, J.S., Eisenberg, D. Characteristics of amyloid-related oligomers revealed by crystal structures of macrocyclic β-sheet mimics. *J Am Chem Soc* **2011**, *133(17)*, 6736-6744, DOI:10.1021/ja200222n
45. Landau, M., Sawaya, M.R., Faull, K.F., Laganowsky, A., Jiang, L., Sievers, S.A., Liu, J., Barrio, J.R., Eisenberg, D. Towards a pharmacophore for amyloid. *PLoS Biol* **2011**, *9(6)*, e1001080, DOI:10.1371/journal.pbio.1001080
46. Liu, C., Zhao, M., Jiang, L., Cheng, P.N., Park, J., Sawaya, M.R., Pensalfini, A., Gou, D., Berk, A.J., Glabe, C.G., Nowick, J., Eisenberg, D. Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc Natl Acad Sci U S A* **2012**, *109(51)*, 20913–20918, DOI:10.1073/pnas.1218792109
47. Wiltzius, J.J., Landau, M., Nelson, R., Sawaya, M.R., Apostol, M.I., Goldschmidt, L., Soriaga, A.B., Cascio, D., Rajashankar, K., Eisenberg, D. Molecular mechanisms for protein-encoded inheritance. *Nat Struct Mol Biol* **2009**, *16(9)*, 973–978, DOI:10.1038/nsmb.1643
48. de la Cruz, M.J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F.E., Sawaya, M.R., Cascio, D., Weiss, S.C., Kim, S.K., Hinck, C.S., Hinck, A.P., Calero, G., Eisenberg, D, Gonen T. Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nature Methods* **2017**, *14(4)*, 399–402, DOI:10.1038/nmeth.4178
49. Konieczny, L., Bryliński, M., Roterman, I. Gauss-function-Based model of hydrophobicity density in proteins. *In Silico Biol.* **2006**, *6(1-2)*, 15-22
50. Simons, K.T., Bonneau, R., Ruczinski, I., Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **1999**, Suppl *3*, 171–176
51. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **1997**, *25(17)*, 3389–3402, DOI:10.1093/nar/25.17.3389
52. Rychlewski, L., Jaroszewski, L., Li, W., Godzik, A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* **2000**, *9(2)*, 232–241, DOI:10.1110/ps.9.2.232
53. Ginalski, K., Elofsson, A., Fischer, D., Rychlewski, L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **2003**, *19*, 1015–1018, DOI:10.1093/bioinformatics/btg124
54. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L. The Pfam protein families database. *Nucleic Acids Res* **2002**, *30(1)*, 276–280, DOI:10.1093/nar/30.1.276
55. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763, DOI:10.1093/bioinformatics/14.9.755
56. Kim, D.E., Chivian, D., Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **2004**, *32(S2)*, W526–W531, DOI:10.1093/nar/gkh468
57. Kuhlman, B., Baker, D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* **2000**, 97(19), 10383-10388, DOI:10.1073/pnas.97.19.10383
58. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **2008**, *9(1)*, 40, DOI:10.1186/1471-2105-9-40
59. Roy, A., Kucukural, A., Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* **2010**, *5(4)*, 725–738, DOI:10.1038/nprot.2010.5

60. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* **2015**, *12(1)*, 7–8, DOI:10.1038/nmeth.3213

61. Wu, S., Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **2007**, *35(10)*, 3375–3382, DOI:10.1093/nar/gkm251

62. Swendsen, R.H., Wang, J.S. Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett* **1986**, *57(21)*, 2607–2609, DOI:10.1103/physrevlett.57.2607

63. Zhang, Y., Skolnick, J. SPICKER: A clustering approach to identify near-native protein folds. *Journal of Computational Chemistry* **2004**, *25(6)*, 865–871, DOI:10.1002/jcc.20011

64. https://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html (accessed on 05-07-2018)

65. https://docs.cyfronet.pl/pages/viewpage.action?pageId=23626756 (accessed on 05-07-2018)

66. Roterman-Konieczna, I. Protein Folding – protein folding versus protein structure prediction. Woodehead Publishing 2012, Oxford, Cambridge, Philadelphia, New Dehli (currently Elsevier) ISBN 978-1-907568-17-6.

67. Roterman, I., Konieczny, L., Banach, M., Jurkowski, W. Intermediates in the protein folding process: a computational model. *Int J Mol Sci* **2011**, *12(8)*, 4850–4860, DOI:10.3390/ijms11084850

68. Kalinowska, B., Krzykalski, A., Roterman, I. Contingency Table Browser - prediction of early stage protein structure. *Bioinformation* **2015**, *11(10)*, 486–488, DOI:10.6026/97320630011486

69. Jurkowski, W., Bryliński, M., Konieczny, L, Wiíniowski, Z., Roterman, I. Conformational subspace in simulation of early-stage protein folding. *Proteins* **2004**, *55(1)*, 115–127, DOI:10.1002/prot.20002

70. Kalinowska, B., Fabian, P., Stąpor, K., Roterman, I. Statistical dictionaries for hypothetical in silico model of the early-stage intermediate in protein folding. *J Comput Aided Mol Des* **2015**, *29(7)*, 609–618, DOI:10.1007/s10822-015-9839-2

71. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, *26(16)*, 1701–1718, DOI:10.1002/jcc.20291

72. Biancalana, M., Makabe, K., Koide, S. Minimalist design of water-soluble cross-β architecture. *Proc Natl Acad Sci U S A* **2010**, *107(8)*, 3469–3474, DOI:10.1073/pnas.0912654107

73. Gadzała, M., Kalinowska, B., Banach, M., Konieczny, L., Roterman, I. Determining protein similarity by comparing hydrophobic core structure. Heliyon **2017**, 3(2), e00235, DOI:10.1016/j.heliyon.2017.e00235