Regular paper

Internal force field in selected proteins

Magdalena Ptak-Kaczor[1,2], Mateusz Banach[1], Leszek Konieczny[3] and Irena Roterman[1✉]

[1]Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Kraków, Poland; [2]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Kraków, Poland; [3]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Kraków, Poland

The fuzzy oil drop model suggests that the tertiary conformation of a protein – particularly a globular one – can be likened to a spherical micelle. During the folding process, hydrophilic residues are exposed on the surface, while hydrophobic residues are retained inside the protein. The resulting hydrophobicity distribution can be mathematically modeled as a 3D Gaussian. The fuzzy oil drop model is strikingly effective in explaining the properties of type II antifreeze proteins and fast-folding proteins, as well as a vast majority of autonomous protein domains. This work aims to determine whether similar mechanisms apply to other types of nonbonding interactions. Our analysis indicates that electrostatic and van der Waals forces do not conform to the Gaussian pattern. The study involves a reference protein (titin) which shows a high agreement between the observed distribution of hydrophobicity and the theoretical (Gaussian) distribution, a selection of amyloid structures derived from the Protein Data Bank, as well as transthyretin – a protein known for its susceptibility to amyloid transformation.

INTRODUCTION

The protein force field includes various types of interactions: electrostatic, vdW and torsion potentials (Jo *et al.*, 2017; Vanommeslaeghe *et al.*, 2010; Brooks *et al.*, 2009; Hynninen & Crowlry, 2018; Pavlova *et al.*, 2018; Boonstra *et al.*, 2016; Mirijanian *et al.*, 2014; Miller *et al.*, 2008; Yu *et al.*, 2012; Best *et al.*, 2012; Patel & Brooks, 2004; Baker *et al.*, 2010; Hornak *et al.*, 2006; Wang *et al.*, 2017; Shao & Zhu 2018; Hoffmann *et al.*, 2018; Wang *et al.*, 2004; Wroblewska & Skolnick, 2007; Liu & Beveridge, 2002; Vymĕtal *et al.*, 2019; Skjevik *et al.*, 2015; Chen *et al.*, 2015; Malolepsza *et al.*, 2010; Rubenstein *et al.*, 2018; van der Spoel *et al.*, 2005; Lundborg & Lindahl, 2015; Aragones *et al.*, 2013, van der Spoel *et al.*, 2012;

Bjelkmar *et al.*, 2010; Abraham & Gready 2011; Hess *et al.*, 2008). In addition, most force fields recognize the effects of hydrogen bonds, while some take into account structural flexibility, where the bond length and dihedral angles may change as a result of stretching and shearing vibrations, respectively. An overview of the force fields used in simulation toolkits such as Amber, Charmm and ECEPP can be found in (Roterman *et al.*, 1989a; Roterman *et al.*, 1989b; Lundborg & Lindahl, 2015; Guvench & MacKerell, 2008).

In addition to the above, some force fields also include a separate component related to the presence of disulfides (Roterman *et al.*, 1989a).

The procedure for obtaining a 3D structure for a given input sequence relies on optimization methods that minimize internal energy (Jo *et al.*, 2017; Vanommeslaeghe *et al.*, 2010; Brooks *et al.*, 2009; Hynninen & Crowlry, 2018; Pavlova *et al.*, 2018; Boonstra *et al.*, 2016; Mirijanian *et al.*, 2014; Miller *et al.*, 2008; Yu *et al.*, 2012; Best *et al.*, 2012; Patel & Brooks, 2004; Baker *et al.*, 2010; Hornak *et al.*, 2006; Wang *et al.*, 2017; Shao & Zhu 2018; Hoffmann *et al.*, 2018; Wang *et al.*, 2004; Wroblewska & Skolnick, 2007; Liu & Beveridge, 2002; Vymĕtal *et al.*, 2019; Skjevik *et al.*, 2015; Chen *et al.*, 2015; Malolepsza *et al.*, 2010; Rubenstein *et al.*, 2018; van der Spoel *et al.*, 2005; Lundborg & Lindahl, 2015; Aragones *et al.*, 2013, van der Spoel *et al.*, 2012; Bjelkmar *et al.*, 2010; Abraham & Gready, 2011; Hess *et al.*, 2008).

Textbook knowledge says that the tertiary structure stability is ensured by the presence of a hydrophobic core, along with a system of disulfide bonds (Devlin, 2010). The latter is evolutionarily conditioned, while the former owes its properties to interactions between the protein chain and its aqueous environment (Dygut *et al.*, 2016; Konieczny *et al.*, 2006; Kalinowska *et al.*, 2015; Roterman *et al.*, 2011; Banach *et al.*, 2012; Banach *et al.*, 2018). Optimization of such interactions involves the exposure of hydrophilic residues to the surface and internalization of hydrophobic residues, ensuring entropically beneficial contact with the solvent. We may therefore consider the protein (particularly a globular one) as a *quasi*-spherical micelle. Various amino acid sequences have different propensities for generating micellar structures. In particular, fast-folding proteins (Roterman *et al.*, 2011), type II antifreeze proteins (Banach *et al.*, 2012; Banach *et al.*, 2018) and a vast majority of protein domains (treated as standalone structural units) exhibit strong accordance between the observed distribution of hydrophobicity and the corresponding model distribution, theorized for a perfect spherical micelle (Sałapa *et al.*, 2012). The latter "idealized" (or "theoretical" distribution) can be mathematically modeled as a 3D Gaussian (Konieczny *et al.*, 2006). The scope and extent of deviations from this distribution are the characteristic

properties of each protein and are often associated with its biological function. Discordance may take on one of two guises: a local excess of hydrophobicity on the surface and a local deficiency of hydrophobicity. The latter typically indicates the presence of a binding cavity, capable of housing a ligand or substrate (Prymula *et al.,* 2011), while the former suggests a potential complexation site (Brylinski *et al.,* 2007). The targeted nature of discrepancies between the observed and theoretical distributions of hydrophobicity ensures high specificity of molecular interactions.

Unlike internal energy optimization algorithms, which do not involve any preconditions regarding the structure of the polypeptide, the fuzzy oil drop model assumes that the folding process is driven by the presence of the aqueous solvent, which promotes micellization of the residue chain (Gadzała *et al.,* 2019).

The phenomenon of misfolding – particularly in the context of amyloid proteins – has attracted much scientific attention over the years. Notably, amyloid structures exhibit different hydrophobicity distribution patterns than globular proteins. In amyloids, this distribution is linear with alternating bands of high and low hydrophobicity stretching along the fibril's axis (Roterman *et al.,* 2017; Roterman *et al.,* 2016; Dułak *et al.,* 2018). Consequently, the amyloid may be described as a ribbonlike micelle.

This work aims to determine whether the distribution of hydrophobicity in polypeptide chains is mirrored by the distribution of other nonbonding interactions. To answer this question, electrostatic and vdW interactions were subjected to statistical analysis, similar to the one performed for hydrophobicity. The observed distributions were compared with two reference models: the theoretical Gaussian distribution (denoted T) and the uniform distribution, where each residue was ascribed the same value of a given interaction (denoted R). Relative Distance (RD) was then calculated as a quantitative measure of deviations from either of these distributions, with values lower than 0.5 indicating that the observed distribution reflected T, while values higher than 0.5 showing stronger affinity for R.

The process of performing the above computations has been detailed in numerous publications devoted to the applicability of the fuzzy oil drop model to various classes of proteins, including amyloids (Kalinowska *et al.,* 2015).

If our analysis reveals that nonbonding interactions are distributed in a similar way to hydrophobicity, we can conclude that internal energy optimization algorithms should recognize and reflect this property of the internal force field.

MATERIALS AND METHODS

Data. Our analysis involves one protein highly consistent with the 3D Gaussian distribution (and therefore with the fuzzy oil drop model), as well as the amyloids

Table 1. List of analyzed proteins.

Protein	PDB ID	Chain length	Reference
Titin	1TIT	89 aa	(Improta *et al.,* 1996)
Amyloid Aβ(1-40)	2MVX	40 aa	(Schűtz *et al.,* 2015)
Aβ(15-40)	2MPZ	26 aa	(Sgourakis *et al.,* 2015)
Aβ(11-42)	2MXU	32 aa	(Xiao *et al.,* 2015)
Tau	5O3L	73 aa	(Fitzpartick *et al.,* 2017)
α-synuclein	2N0A	140 aa	(Tuttle *et al.,* 2016)
Transthyretin	1DVQ	115 aa	(Klabunde *et al.,* 2000)

listed in the Protein Data Bank (Berman *et al.,* 2000). Table 1 presents the list of proteins selected for our study.

The set includes titin as a representative of a class of strongly accordant proteins (Banach *et al.,* 2014a; Banach *et al.,* 2014b), along with several amyloids whose structures can be found in PDB. The latter facilitate the analysis of the distribution of hydrophobic, electrostatic and vdW interactions in amyloids (note that the distribution of hydrophobicity in amyloids is extensively discussed in several other publications: (Dułak *et al.,* 2018a; Dułak *et al.,* 2018b)). Transthyretin was selected for this study due to its strong propensity for amyloid transformation (Lim *et al.,* 2016).

Fuzzy oil drop model. A detailed presentation of the fuzzy oil drop model can be found in numerous publications (Konieczny *et al.,* 2006; Kalinowska *et al.,* 2015; Roterman *et al.,* 2011). Therefore, we will limit ourselves to a brief recapitulation of its core concepts relevant to the presented results.

The observed distribution of hydrophobicity in a protein molecule is determined by the interactions between its constituent residues. These interactions depend both on the intrinsic hydrophobicity of each residue and on the separation between neighboring residues. Our calculation is based on the function proposed in (Levitt, 1976). Having obtained the observed distribution (O), we may compare it with the corresponding theoretical distribution (T), where each residue is assigned a hydrophobicity value (T_i) equal to the value of the 3D Gaussian at its location (which is assumed to correspond to the so-called effective atom, i.e. averaged-out values of the positions of all atoms comprising the given residue). The 3D Gaussian is obtained by encapsulating the protein body in a suitably shaped ellipsoid capsule, whose dimensions are expressed by a trio of sigma coefficients (denoted σ_x, σ_y and σ_z).

Normalization of T and O allows direct comparison of both distributions. This is done by employing the divergence entropy parameter proposed by Kullback and Leibler (Kullback & Leibler, 1951). This parameter, denoted D_{KL}, expresses the "distance" between two distributions: observed (O) and idealized (T), of which the latter is treated as a reference distribution. As a result, T describes a "perfect" globular protein, in which hydrophobicity peaks at the center and remains low on the surface. The value of D_{KL} cannot, however, be interpreted on its own, since it is a measure of entropy. This is why another reference distribution is necessary to facilitate meaningful comparisons. This second reference distribution – called R – is defined as a "polar opposite" of T. Under R, each residue is ascribed a value of $1/N$ (N being the total number of residues in the chain). Thus, R describes a situation where no concentration of hydrophobicity exists at any point within the protein body. Calculating D_{KL} for both O-T and O-R tells us whether O approximates T or R more closely, with lower values indicating greater similarity. To avoid having to deal with two separate values, a Relative Distance (RD) parameter is derived as the relation of D_{KL} (O-T) to the sum of D_{KL} (O-T) and (O-R). RD < 0.5 indicates the presence of a hydrophobic core.

RD is calculated as follows:

$$RD = \frac{D_{KL}(O-T)}{D_{KL}(O-T) + D_{KL}(O-R)}$$

Given that numerous proteins exhibit strong accordance between the observed distribution of hydropho-

Table 2. RD parameters values for each interaction.

PDB ID	RD for INTERACTIONS		
	Electrostatic OE	vdW OV	Hydrophobic OH
1TIT	0.839	0.807	0.382
2MPZ	0.749	0.809	0.541
2MXU	0.806	0.757	0.513
2MVX	0.858	0.826	0.608
2N0A*	0.870	0.789	0.506
5O3L	0.815	0.767	0.664
1DVQ**	0.850	0.864	0.652

*analysis restricted to the 30–100 fragment, **transthyretin. The values describe the status of the chains treated as part of the complex.
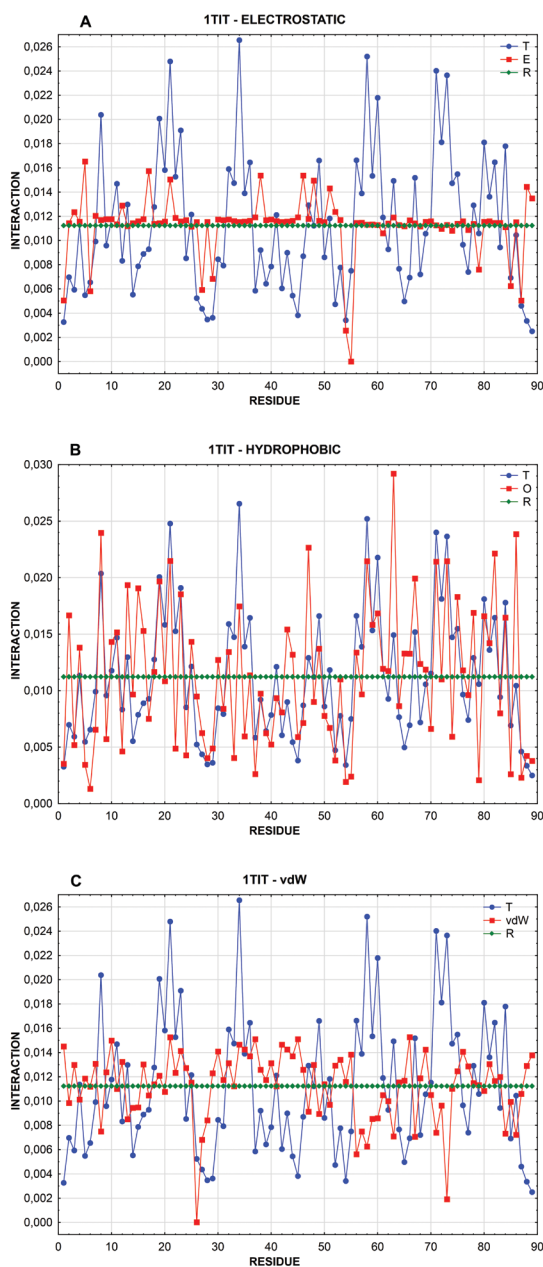

Figure 1. Distribution profiles for 1TIT (reference protein characterized by a low RD value for hydrophobic interactions):
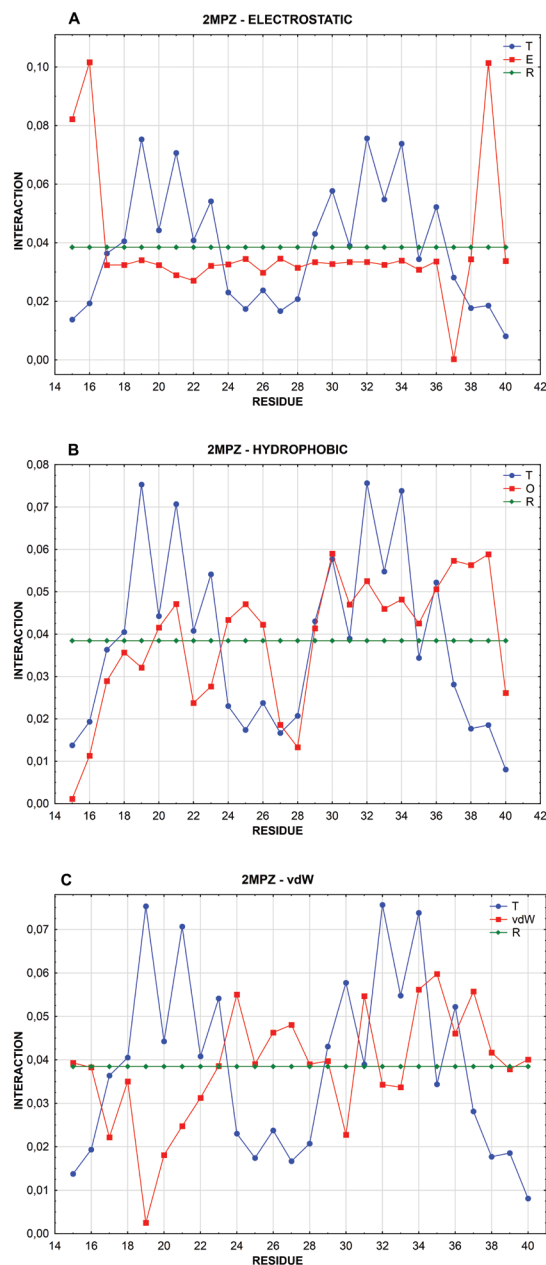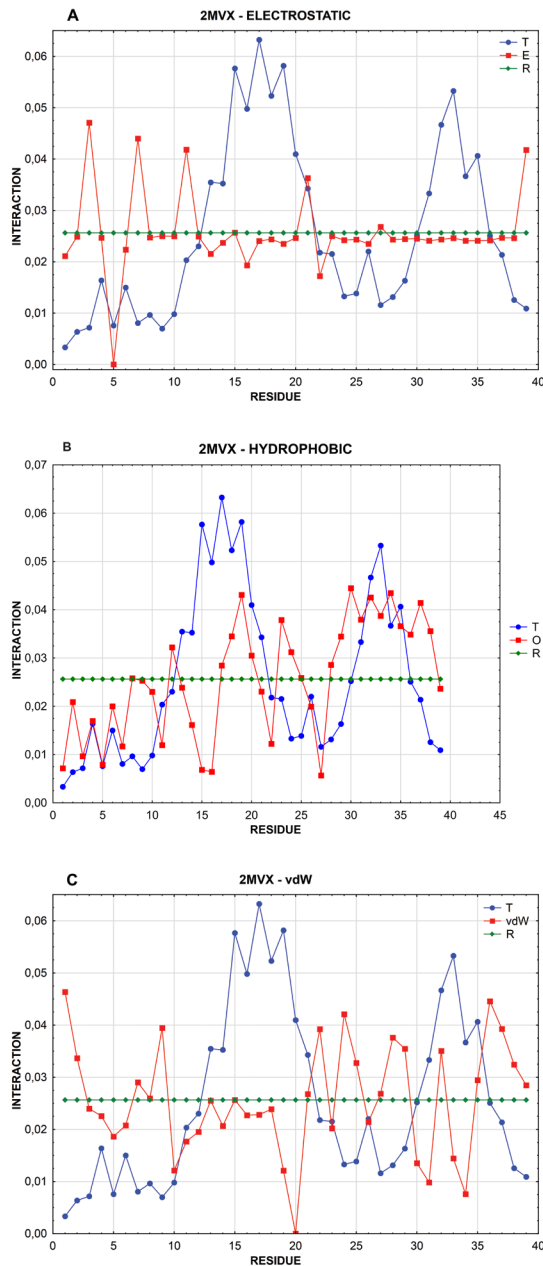A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

Figure 2. Distribution profiles for amyloid proteins: Aβ(15-40) with D23N mutation (2MPZ).
A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

bicity and the theoretical Gaussian, it is interesting to speculate whether other nonbonding interactions (particularly electrostatic and vdW potentials as computed by Gromacs (van der Spoel *et al.,* 2005)) follow a similar pattern. Consequently, we have computed E_i and V_i values for each residue, expressing (respectively) their electrostatic and van der Waals potentials. These observed values were subsequently compared with T to determine whether electrostatic and vdW interactions follow a centralized pattern, similar to hydrophobic interactions.

The Gromacs force field – electrostatic and vdW potentials. The distribution of observed electrostatic and vdW potentials was performed using Gromacs (van der Spoel *et al.,* 2005) running on computing resources supplied by Cyfronet AGH in the PL-Grid infrastructure (http://plgrid.pl/ (August 2019)). The selected molecule, whose structure had been derived
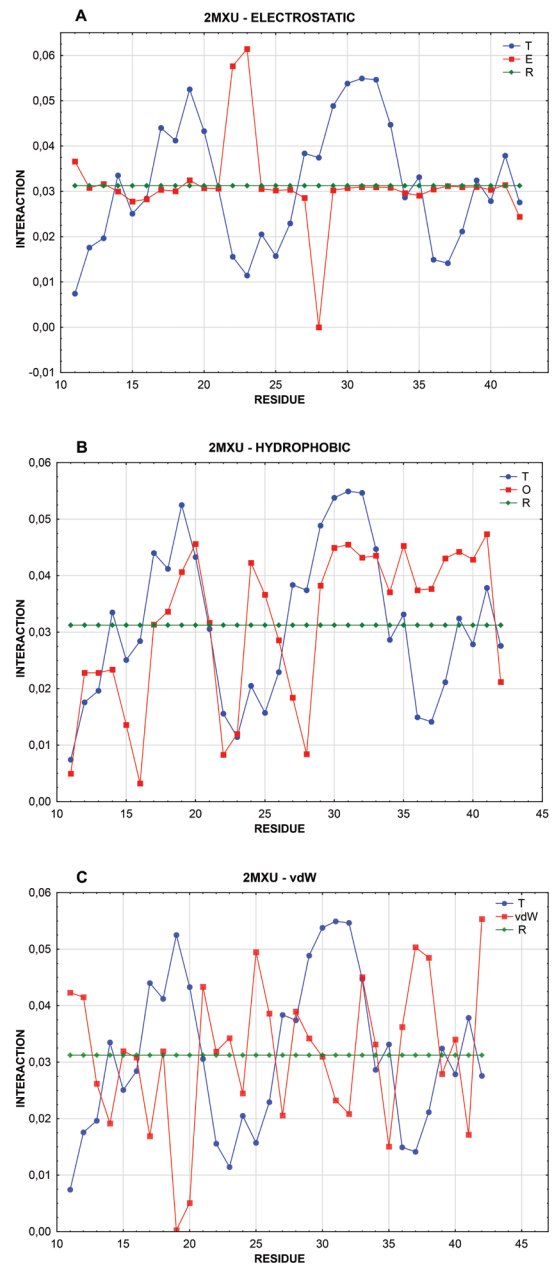
Figure 3. Distribution profiles for amyloid proteins: Aβ(1-40) (2MVX).
A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

Figure 4. Distribution profiles for amyloid proteins: Aβ(11-42) with (E22Δ mutation) (2MXU).
A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

from PDB (Berman *et al.,* 2000), was subjected to energy minimization in order to eliminate steric clashes introduced by incorporation of H atoms based on X-ray diffraction studies.

Following optimization using Gromacs software (specifically, Gromacs 5.0.7 with the OPLS-AA/L all-atom force field), we computed the strength of electrostatic and vdW interactions between each residue and its respective neighborhood. This yielded a distribution of both types of interactions throughout the protein body, mathematically expressed as a set of E_i and V_i values, which could be directly compared to our reference distributions (T and R) given by the fuzzy oil drop model (Kalinowska *et al.,* 2015).

In our previous publications, RD was used to express the relative distance between the observed distribution of hydrophobicity (O) and two reference

distributions (T and R – as described above). Here, instead of hydrophobicity, O denotes the distribution of two other types of interactions (E and V – electrostatic and vdW respectively). If the computed value of RD is greater than 0.5, we conclude that the given interaction follows the uniform pattern rather than the monocentric core pattern.

Computing the status of individual residues. In order to compare the status of individual amino acids with the corresponding reference distributions, we need to be aware of the interactions between each residue and its neighbors. Having measured these interactions, we may calculate divergence entropy (Kullback & Leibler, 1951) to determine the scope of alignment between the observed and reference distributions. The corresponding procedure (utilizing the
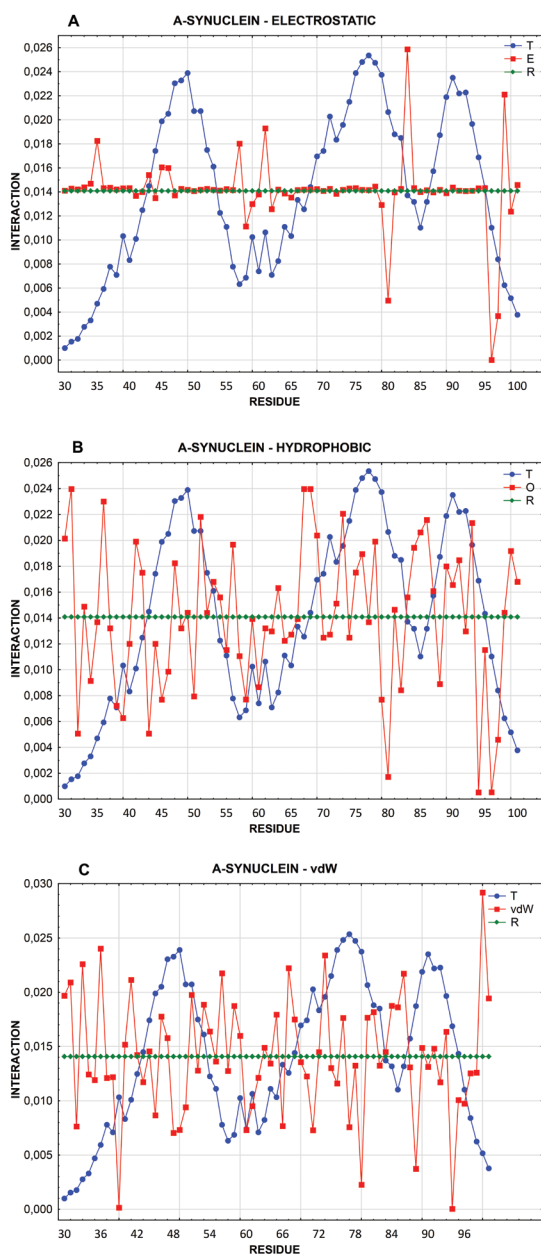
Figure 5. Distribution profiles for amyloid proteins: α-synuclein (2N0A).
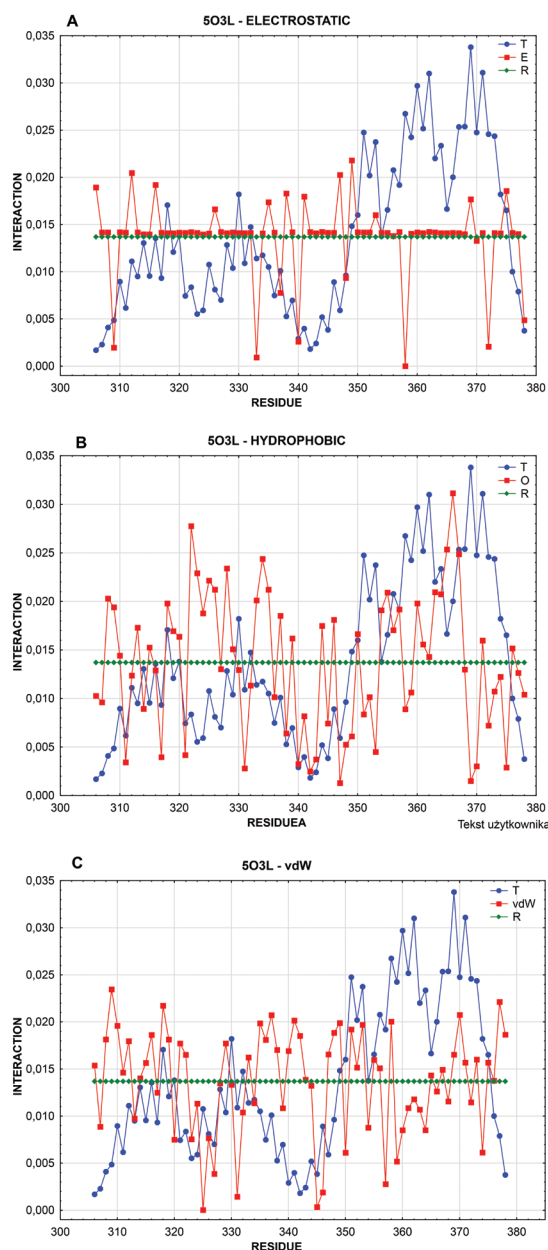A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

Gromacs suite) is presented in Supplementary Materials (at https://ojs.ptbioch.edu.pl/index.php/abp/).

RESULTS

Table 2 presents the Relative Distance (RD) values for T-O-R models for electrostatic interactions (OE), OV – vdW interactions, and OH expressing hydrophobic interactions. In each case, T (Gaussian) and R distributions were used as references.
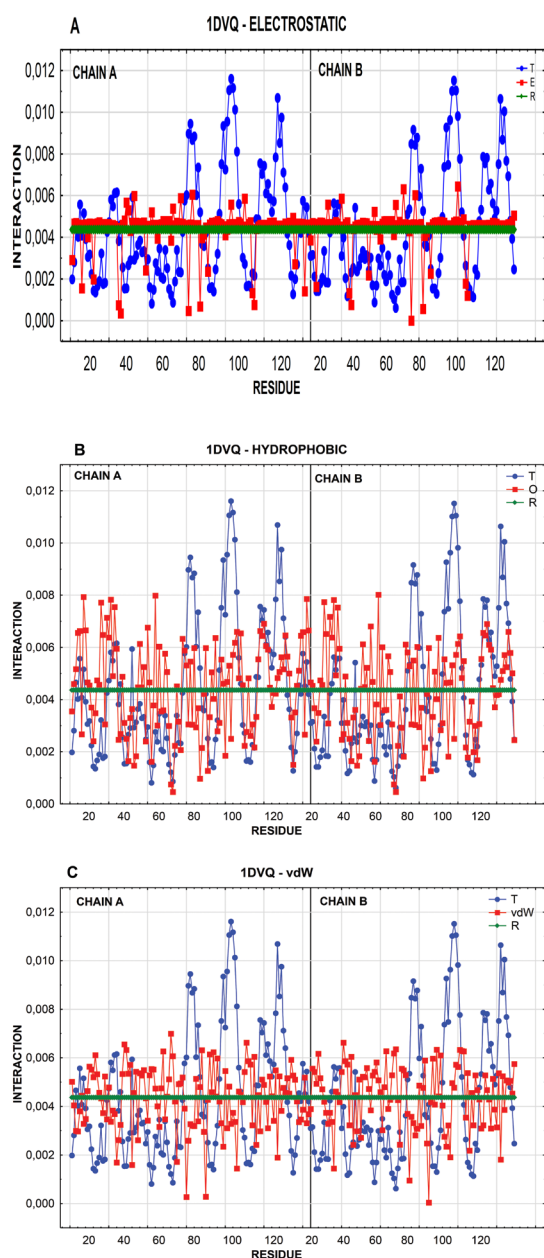
Figure 1 provides a summary of the results of the T, OH, OE and OV distribution profiles for titin. The green horizontal line represents the R distribution. The blue profile represents the idealized (T) distribution given by the 3D Gaussian. The red line represents the O distribution for the analyzed interaction. The observed profile of hydrophobic interactions is

Figure 6. Distribution profiles for amyloid proteins: tau (5O3L).
A – electrostatic (OE), B – hydrophobicity (OH), C – vdW (OV).

clearly similar to the theoretical model. The RD value for titin is very low at 0.382 (Table 1). Yet, major differences between OH and the other two observed distributions can be seen. The nonbonding interactions appear to be more closely aligned with the uniform distribution (R), and therefore evenly distributed over the entire chain length, with no noticeable peaks. The red line is aligned with the green line in both Fig. 1A and Fig. 1C, and the corresponding values of RD are very high: 0.839 for OE and 0.807 for OV. These high values reveal strong alignment of both OE and OV with R. Based on these observations, we may conclude that a hydrophobic core is clearly present in titin, while neither OE nor OV reveal any "core-like" concentration. In particular, the lack of any electrostatic and vdW concentration is observed in the central part of the molecule, despite close packing of atoms and residues at that location.

Figure 7. Distribution profiles for transthyretin (1DVQ), a protein known for its propensity for amyloid transformation.
A – electrostatic (OH), B – hydrophobicity (OE), C – vdW (OV).

highlighted differences related to distributions of hydrophobicity, identifying areas where the observed distribution opposes the theoretical model (Roterman *et al.*, 2016; Lim *et al.*, 2016; Dułak *et al.*, 2019). These fragments are regarded as "amyloid seeds" – however, in the context of electrostatic and vdW interactions, they have no special properties compared to the rest of the amyloid chain.

DISCUSSION AND CONCLUSIONS

The fuzzy oil drop model treats amyloids as ribbon-like micelles (in contrast to spherical micelles, which represent globular proteins) (Kalinowska *et al.*, 2017). The presence of short fragments, whose local distribution of hydrophobicity opposes the theoretical model, results from the systemic inability of the amyloid peptide to attain a spherical conformation (Dułak *et al.*, 2018a; Dułak *et al.*, 2018b). This locally suboptimal distribution is compensated by the isolation of discordant fragments in the structure of a ribbon-like micelle. Rapid growth of amyloid fibrils is therefore linked to the overriding need to shield hydrophobic residues by attracting and attaching additional peptides. This process results in the formation of a peculiar type of hydrophobic core in the form of a centrally located hydrophobic band.

No similar mechanism can be identified in the scope of electrostatic and vdW interactions (Marchewka *et al.*, 2011). Therefore, protein structure prediction algorithms which rely solely on the composition of the input sequence are correct in applying the isotropic model to these interactions. Anisotropic distribution appears to occur only in relation to hydrophobic forces, resulting in isolation of hydrophobic residues in the interior of a globular micelle (or in the central part of an amyloid ribbon). Despite a relatively low number of globular proteins subjected to analysis (see: Marchewka *et al.*, 2011; Roterman, 2019) for additional examples), we may conclude that the properties of electrostatic and vdW interactions in amyloids are not significantly different from those observed in globular proteins.

The practical conclusion for protein folding simulations is that non-directed optimization can be applied for energy minimization expressing electrostatic and vdW interactions, while directed optimization is better suited for hydrophobic interactions which tend to produce a hydrophobic core.

Acknowledgements

A similar situation occurs in amyloids, where both electrostatic and vdW distributions closely match R. While in terms of hydrophobic interactions, amyloids are very different from globular proteins (such as titin), similar differences in the scope of electrostatic and vdW distributions cannot be observed (Figs 2–6).

Regarding transthyretin – its distribution of hydrophobicity may be regarded as an intermediate case. The protein contains a centralized hydrophobic core but also exhibits major deviations from the Gaussian pattern. In contrast, electrostatic and vdW distributions are similar to those observed in other analyzed proteins and largely consistent with the uniform pattern (Fig. 7).

Our analysis underscores the distinction between globular proteins and amyloids. Previous publications

REFERENCES

Abraham MJ, Gready JE (2011) Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *J Comput Chem* **32**: 2031–2040. https://doi.org/10.1002/jcc.21773.

Aragones JL, Noya EG, Valeriani C, Vega C (2013) Free energy calculations for molecular solids using GROMACS. *J Chem Phys* **139**: 034104. https://doi.org/10.1063/1.4812362.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037–10041. https://doi.org/10.1073/pnas.181342398

Baker CM, Lopes PE, Zhu X, Roux B, Mackerell AD Jr (2010) Accurate calculation of hydration free energies using pair-specific lennard-jones parameters in the CHARMM drude polarizable force field. *J Chem Theory Comput* **6**: 1181–1198. https://doi.org/10.1021/ct9005773

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

Banach M, Prymula K, Jurkowski W, Konieczny L, Roterman I (2012) Fuzzy oil drop model to interpret the structure of antifreeze proteins and their mutants. *J Mol Model* **18**: 229–237. https://doi.org/10.1007/s00894-011-1033-4.

Banach M, Konieczny L, Roterman I (2014a) 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. https://doi.org/10.1016/j.jtbi.2014.05.007.

Banach M, Roterman I, Prudhomme N, Chomilier J (2014b) Hydrophobic core in domains of immunoglobulin-like fold. *J Biomol Struct Dyn* **32**: 1583–1600. https://doi.org/10.1080/07391102.2013.829756

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank *Nucleic Acids Res* **28**: 235–242. https://doi.org/10.1093/nar/28.1.235

Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J Chem Theory Comput* **8**: 3257–3273. https://doi.org/10.1021/ct300400x

Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E (2010) Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J Chem Theory Comput* **6**: 459–466. https://doi.org/10.1021/ct900549r

Boonstra S, Onck PR, Giessen Ev. (2016) CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state. *J Phys Chem B* **120**: 3692–3698. https://doi.org/10.1021/acs.jpcb.6b01316.

Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**: 1545–1614. https://doi.org/10.1002/jcc.21287.

Bryliński M, Prymula K, Jurkowski W, Kochańczyk M, Stawowczyk E, Konieczny L, Roterman I (2007) Prediction of functional sites based on the fuzzy oil drop model. *PLoS Comput Biol* **3**: e94. PMID: 17530916

Chen W, Shi C, MacKerell AD. Jr, Shen J. (2015) Conformational dynamics of two natively unfolded fragment peptides: comparison of the AMBER and CHARMM force fields. *J Phys Chem B* **119**: 7902–7910. https://doi.org/10.1021/acs.jpcb.5b02290.

Devlin TM (2010) (Editor) *Textbook of Biochemistry with Clinical Correlations*, 7th edn. ISBN: 978-0-470-28173-4

Dułak D, Gadzała M, Banach M, Ptak M, Wiśniowski Z, Konieczny L (2018a) Filamentous aggregates of tau proteins fulfil standard amyloid criteria provided by the fuzzy oil drop (FOD) *Model Int J Mol Sci* **19**: 2910. https://doi.org/10.3390/ijms19102910

Dułak D, Banach M, Gadzała M, Konieczny L, Roterman I (2018b) Structural analysis of the Aβ(15-40) amyloid fibril based on hydrophobicity distribution. *Acta Biochim Pol* **65**: 595–604. https://doi.org/10.18388/abp.2018_2647.

Dułak D, Gadzała M, Banach M, Konieczny L, Roterman I (2019) Structural analysis of the Aβ(11-42) amyloid fibril based on hydrophobicity distribution. *J Comp-Aided Mol Des*. https://doi.org/10.1007/s10822-019-00209-9.

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**. pii: E1741

Fitzpatrick AWP, Falcon B, He S, Murzin AG, Murshudov G, Garringer HJ, Crowther RA, Ghetti B, Goedert M, Scheres SHW (2017) Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**: 185–190. https://doi.org/10.1038/nature23002.

Gadzała M, Dułak D, Kalinowska B, Baster Z, Bryliński M, Konieczny L, Banach M, Roterman I (2019) The aqueous environment as an active participant in the protein folding process. *J Mol Graph Model* **87**: 227–239. https://doi.org/10.1016/j.jmgm.2018.12.008.

Guvench O, MacKerell AD Jr (2008) Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol* **443**: 63–88. https://doi.org/10.1007/978-1-59745-177-2_4

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **4**: 435–447. https://doi.org/10.1021/ct700301q.

Hoffmann F, Mulder FAA, Schäfer LV (2018) Accurate methyl group dynamics in protein simulations with AMBER force fields. *J Phys Chem B* **122**: 5038–5048. https://doi.org/10.1021/acs.jpcb.8b02769.

Hynninen AP, Crowley MF (2014) New faster CHARMM molecular dynamics engine. *J Comput Chem* **35**: 406–413. https://doi.org/10.1002/jcc.23501.

Improta S, Politou AS, Pastore A (1996) Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* **4**: 323–337. PMID: 8805538

Jo S, Cheng X, Lee J, Kim S, Park SJ, Patel DS, Beaven AH, Lee KI, Rui H, Park S, Lee HS, Roux B, MacKerell AD Jr, Klauda JB, Qi Y, Im W (2017) CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **38**: 1114–1124. https://doi.org/10.1002/jcc.24660.

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. https://doi.org/10.3390/e17031477

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. https://doi.org/10.1007/s00894-017-3367-z.

Klabunde T, Petrassi HM, Oza VB, Raman P, Kelly JW, Sacchettini JC (2000) Rational design of potent human transthyretin amyloid disease inhibitors. *Nat Struct Biol* **7**: 312–321. https://doi.org/10.1038/74082

Konieczny L, Bryliński M, Roterman I (2006) Gauss-function-Based model of hydrophobicity density in proteins. *In Silico Biol* **6**: 15–22. PMID: 17518766

Kumari R, Kumar R. Open Source Drug Discovery Consortium, Lynn A (2014) g_mmpbsa – A GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* **54**: 1951–1962. https://doi.org/10.1021/ci500020m

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

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

Lim KH, Dasari, AK, Hung I, Gan Z, Kelly JW, Wright PE, Wemmer DE (2016) Solid-state NMR studies reveal native-like β-sheet structures in transthyretin amyloid. *Biochemistry* **55**: 5272–5278. https://doi.org/10.1021/acs.biochem.6b00649.

Liu Y, Beveridge DL (2002) Exploratory studies of ab initio protein structure prediction: multiple copy simulated annealing, AMBER energy functions, and a generalized born/solvent accessibility solvation model. *Proteins* **46**: 128–146. PMID: 11746709

Lundborg M, Lindahl E (2015) Automatic GROMACS topology generation and comparisons of force fields for solvation free energy calculations. *J Phys Chem B* **119**: 810–823. https://doi.org/10.1021/jp505332p.

Malolepsza E, Strodel B, Khalili M, Trygubenko S, Fejer SN, Wales DJ (2010) Symmetrization of the AMBER and CHARMM force fields. *J Comput Chem* **31**: 1402–1409. https://doi.org/10.1002/jcc.21425.

Marchewka D, Banach M, Roterman I (2011) Internal force field in proteins seen by divergence entropy. *Bioinformation* **6**: 300–302. https://doi.org/10.6026/97320630006300

Miller BT, Singh RP, Klauda JB, Hodoscek M, Brooks BR, Woodcock HL 3rd (2008) CHARMMing: a new, flexible web portal for CHARMM. *J Chem Inf Model* **48**: 1920–1929. https://doi.org/10.1021/ci800133b.

Mirijanian DT, Mannige RV, Zuckermann RN, Whitelam S (2014) Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *J Comput Chem* **35**: 360–370. https://doi.org/10.1002/jcc.23478.

Patel S, Brooks CL 3rd (2004) CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem* **25**: 1–15. https://doi.org/10.1002/jcc.10355

Pavlova A, Parks JM, Gumbart JC (2018) Development of CHARMM-compatible force-field parameters for cobalamin and related cofactors from quantum mechanical calculations. *J Chem Theory Comput* **14**: 784–798. https://doi.org/10.1021/acs.jctc.7b01236.

Prymula K, Jadczyk T, Roterman I (2011) Catalytic residues in hydrolases: analysis of methods designed for ligand-binding site prediction. *J Comput Aided Mol Des* **25**: 117–133. https://doi.org/10.1007/s10822-010-9402-0

Roterman IK, Gibson KD, Scheraga HA (1989a) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. I. Conformational predictions for the tandemly repeated peptide (Asn-Ala-Asn-Pro)9. *J Biomol Struct Dyn* **7**: 391–419. https://doi.org/10.1080/07391102.1989.10508502

Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989b) A comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. Phi-psi maps for N-acetyl alanine N'-methyl amide: comparisons, contrasts and simple experimental tests. *J Biomol Struct Dyn* **7**: 421–453. https://doi.org/10.1080/07391102.1989.10508503

Roterman I, Konieczny L, Jurkowski W, Prymula K, Banach M (2011) Two-intermediate model to characterize the structure of fast-folding proteins. *J Theor Biol* **283**: 60–70. https://doi.org/10.1016/j.jtbi.2011.05.027.

Roterman I, Banach M, Kalinowska B, Konieczny L (2016) Influence of the aqueous environment on protein structure – a plausible hy-

pothesis concerning the mechanism of amyloidogenesis. *Entropy* **18**: 351. https://doi.org/10.3390/e18100351

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

Roterman I (2019) *From globular proteins to amyloids*. Irena Roterman-Konieczna ed, Elsevier, 2019 (in press)

Rubenstein AB, Blacklock K, Nguyen H, Case DA, Khare SD (2018) Systematic comparison of amber and rosetta energy functions for protein structure evaluation. *J Chem Theory Comput* **14**: 6015–6025. https://doi.org/10.1021/acs.jctc.8b00303.

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. https://doi.org/10.2478/bams-2012-0023

Schütz AK, Vagt T, Huber M, Ovchinnikova OY, Cadalbert R, Wall J, Güntert P, Böckmann A, Glockshuber R, Meier BH. (2015) Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew Chem Int Ed Engl* **54**: 331–335. https://doi.org/10.1002/anie.201408598.

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. https://doi.org/10.1016/j.str.2014.10.022.

Shao Q, Zhu W (2018) Assessing AMBER force fields for protein folding in an implicit solvent. *Phys Chem Phys* **20**: 7206–7216. https://doi.org/10.1039/c7cp08010g.

Skjevik ÅA, Madej BD, Dickson CJ, Teigen K, Walker RC, Gould IR (2015) All-atom lipid bilayer self-assembly with the AMBER and CHARMM lipid force fields. *Chem Commun (Camb)* **51**: 4402–4405. https://doi.org/10.1039/c4cc09584g.

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**: 1701–1718. https://doi.org/10.1002/jcc.20291

van der Spoel D, van Maaren PJ, Caleman C (2012) GROMACS molecule & liquid database. *Bioinformatics* **28**: 752–753. https://doi.org/10.1093/bioinformatics/bts020.

Tuttle MD, Comellas G, Nieuwkoop AJ, Covell DJ, Berthold DA, Kloepper KD, Courtney JM, Kim JK, Barclay AM, Kendall A, Wan W, Stubbs G, Schwieters CD, Lee VM, George JM, Rienstra CM (2016) Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat Struct Mol Biol* **23**: 409–415. https://doi.org/10.1038/nsmb.3194.

Vanommeslaeghe K, Hatcher E, Acharya, C, Kundu, S, Zhong, S, Shim, J, Darian, E, Guvench O, Lopes, P, Vorobyov, I, Mackerell AD Jr (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**: 671–690. https://doi.org/10.1002/jcc.21367.

Vymětal J, Jurásková V, Vondrášek J. (2019) AMBER and CHARMM force fields inconsistently portray the microscopic details of phosphorylation. *J Chem Theory Comput* **15**: 665–679. https://doi.org/10.1021/acs.jctc.8b00715.

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* **25**: 1157–1174.

Wang LP, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, Swope WC, Martínez TJ, Pande VS (2017) Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J Phys Chem B* **121**: 4023–4039. https://doi.org/10.1021/acs.jpcb.7b02320.

Wroblewska L, Skolnick J (2007) Can a physics-based, all-atom potential find a protein's native structure among misfolded structures? I. Large scale AMBER benchmarking. *J Comput Chem* **28**: 2059–2066. https://doi.org/10.1002/jcc.20720

Xiao Y, Ma B, McElheny D, Parthasarathy S, Long F, Hoshi M, Nussinov R, Ishii Y (2015) Aβ(1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol* **22**: 499–505. https://doi.org/10.1038/nsmb.2991.

Yu W, He X, Vanommeslaeghe K, MacKerell AD Jr (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J Comput Chem* **33**: 2451–2468. https://doi.org/10.1002/jcc.23067.