Properties of monomeric Aβ42 probed by different sampling methods and force fields: Role of energy components

Cite as: J. Chem. Phys. **151**, 055101 (2019); https://doi.org/10.1063/1.5093184
Submitted: 18 February 2019 . Accepted: 10 July 2019 . Published Online: 05 August 2019

Pawel Krupa, Pham Dinh Quoc Huy, and Mai Suan Li

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Properties of monomeric Aβ42 probed by different sampling methods and force fields: Role of energy components

Cite as: J. Chem. Phys. 151, 055101 (2019); doi: 10.1063/1.5093184
Submitted: 18 February 2019 • Accepted: 10 July 2019 •
Published Online: 5 August 2019 • corrected 14 August 2019

Pawel Krupa,[1,a)] Pham Dinh Quoc Huy,[1] and Mai Suan Li[1,2,a)]

AFFILIATIONS

[1] Institute of Physics Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
[2] Institute for Computational Science and Technology, SBI Building, Quang Trung Software City, Tan Chanh Hiep Ward, District 12, Ho Chi Minh City, Vietnam

[a)] **Authors to whom correspondence should be addressed:** pkrupa@ifpan.edu.pl and masli@ifpan.edu.pl

ABSTRACT

Aggregation of intrinsically disordered proteins (IDPs), such as amyloid beta peptide, can cause serious health problems, associated with disorders including Alzheimer disease. Due to the lack of a stable structure and transient nature, such proteins and peptides are often very difficult or even impossible to study using experimental approaches. Therefore, usage of computational tools can provide valuable insight into their dynamics, structural changes, and mechanism of aggregation. Because current force fields were designed to work well for standard proteins with a well-defined native structure and high conformational stability, we examined three force fields most frequently used for studies of proteins, and two variants modified for better performance for IDPs on an example of monomeric amyloid beta 42 (Aβ42) with two sampling approaches: single 10 μs long conventional molecular dynamics (CMD) trajectories and 48-replica runs using the replica exchange MD (REMD). We found that newer force fields (Amber FF14SB and CHARMM36m) provided better results than their older versions (Amber FF99SB and CHARMM36), while the specially modified version for the IDPs (FF14SB_IDPs) yielded similar results to its parent, improving sampling using CMD simulations, hence allowing to achieve a similar level of accuracy at significantly lower computational costs. With sufficient sampling, the newer force fields provided good agreement with the available experimental data. We also analyzed the physical basis of different behaviors of force fields and sampling methods, concluding that in CHARMM interactions with water play a much more important role than in Amber force fields. This explains why, in CHARMM force fields, the monomeric Aβ42 is less stable and more hydrophilic, having a greater solvent accessible surface area.

I. INTRODUCTION

Molecular dynamics (MD)[1] is a very useful tool, which can be widely used to study dynamics and behavior of various systems, such as peptides,[2] proteins,[3] nucleic acids,[4] and other polymers.[5] To be able to observe interesting behaviors of investigated systems, reaching the sufficient time scale of the simulations is required.[6] One solution is to use coarse-grained force fields,[7] such as MARTINI,[8] UNited RESidue (UNRES),[9] Optimized Potential for Efficient structure Prediction (OPEP),[10] or CABS;[11] however, their usage can be limited by the simplifications used. On the other hand, there are various methods to enhance sampling in the existing all-atom force fields, the most popular of which is the replica exchange molecular dynamics (REMD),[12] in which N runs are executed, and after a given number of steps, exchanges are attempted based on the metropolis criterion. While more complicated variants of REMD can be run, such as Hamiltonian REMD,[13] in which more than one dimension of parameters is used, the most popular variant uses different temperatures for every replica. In such a simulation, the system is traveling from a low to a high temperature, at which it can overcome energy barriers and jump to other areas of the conformational space. The system is then cooled down, visiting different basins of the low free energy. To achieve good sampling, such a process should be repeated multiple times during the simulation; therefore, in a perfect scenario, as long as possible, REMD trajectories should be run and analyzed.[14] However, in the real world with limited resources, there are two most probable options: (i) running one or several relatively long independent MD trajectories or (ii) running

multiple (from 20 to over a 100) short runs in the replica exchange scheme.

For small model proteins in vacuum or implicit solvent or with use of coarse-grained force fields, REMD simulations are rather straightforward and provide good sampling due to the small number of interacting sites.[15] As their number is increasing (e.g., if all-atom simulations in explicit solvent are performed), the number of replicas rapidly grows to maintain a satisfactory exchange rate. However, even with a sufficient exchange rate between neighboring replicas, the investigated system must cross several dozen replicas in order to travel from low to high temperatures, which requires a very large number of exchanges and consequently, very long simulations.[14] With the development of GPGPU (general-purpose computing on graphics processing units) calculations, which can speed up the simulation by an order of magnitude,[16] performing conventional simulations instead of REMD (which are almost impossible to be run with GPUs with reasonable speed) can be rational in some cases.[17]

Despite the recent progress, results obtained by MD simulations strongly depend on the used force field[18] and water model.[19–21] In the past, development of force fields was focusing mostly on one state or type[22] of the systems (e.g., folded[23] or disordered proteins[24]), whereas currently the priority is to obtain one force field, which could be efficiently used for all protein types[25] and other biomacromolecules.[26] It was found that the presence of β-structures, understood as population of the fibril-prone structures, is strongly correlated with formation of oligomers and it is a reason why Aβ42 is more prone to aggregation than Aβ40.[27] For amyloid beta (Aβ) truncated to residues 16–22 and other short peptides, it was found that Amber FF99SB overestimated helix content and did not predict any β-structures even for dimers and trimers and that GROningen MOlecular Simulation (GROMOS) overestimated β-content, while Optimized Potential for Liquid Simulations (OPLS) provided the most reliable results.[28] In the case of dimeric Aβ40, all tested force fields overestimated collision cross section values and most of them overestimated also α-helical content in comparison with the experimental data, and only CHARMM22* and CHARMM36 provided expected collapse of the central (residues 17–21) and C-terminal (residues 30–36) hydrophobic cores.[29] Therefore, in this project, both influences of the sampling method and force field were investigated by the analysis of longer conventional MD (CMD) simulations, which were compared with shorter REMD simulations. All simulations were carried out by Amber package, using five popular force fields: Amber FF99SB,[30] FF14SB,[31] FF14SB_IDPs,[24] CHARMM36,[32] and 36m[33] for the monomeric amyloid beta 42 (Aβ42). It should be noted that the role of energy components in various force fields was not previously considered, motivating us to study this problem in detail. We have shown that peptide-solvent interaction in Chemistry at Harvard Macromolecular Mechanics (CHARMM) is more important than in Amber force fields, which led to two main observations that with the CHARMM force fields Aβ42 is less stable with lower beta content and less hydrophobic because the solvent accessible surface area (SASA) is larger.

II. METHODS

A. Simulation details

Three popular modern force fields were utilized in the simulations: Amber FF99SB, Amber FF14SB, and CHARMM36, as well as two of their variants optimized to be more suitable to study IDPs: Amber FF14SB_IDPs and CHARMM36m, using the Amber pmemd tool[34] as a computational engine. Time step in the simulations was set to 2 fs, and snapshots were saved every 0.2 ns in CMD and every 0.02 ns in REMD simulations. As an initial structure, a proposed structure of Aβ42 monomer by Yang and Teplow[35] was used (Fig. S1). Every system consisted of single uncapped Aβ42 surrounded by 7932 and 8118 water molecules, represented by transferable intermolecular potential 3P (TIP3P) model,[36,37] for Amber and CHARMM force fields, respectively, and 3 counterions (Na^+) to neutralize the charge. Minimum dimension of the periodic box was set to 63 Å, which prohibited oligomerization. CHARMM topology and coordinates files were obtained using CHARMM-GUI,[38,39] while input files for Amber simulations were generated by the tleap tool. Obtained systems were subjected to energy minimization and short MD run with NPT conditions to equilibrate molecules. Subsequently, two series of NVT simulations were run: (i) single CMD simulations of 10 μs and (ii) replica exchange molecular dynamics (REMD) simulations consisted of 48 replicas 0.6 μs each, providing 28.8 μs, in total. CMD simulations were run at 300 K, while REMD simulations were run at the range of temperatures from 278.00 to 373.77 K (temperatures of replicas were generated by the predicting tool:[40] 278.00, 279.82, 281.64, 283.47, 285.31, 287.16, 289.03, 290.90, 292.78, 294.66, 296.56, 298.45, 300.37, 302.29, 304.23, 306.17, 308.13, 310.09, 312.06, 314.04, 316.04, 318.04, 320.05, 322.07, 324.11, 326.15, 328.20, 330.27, 332.34, 334.42, 336.53, 338.64, 340.75, 342.87, 345.01, 347.16, 349.32, 351.49, 353.67, 355.85, 358.06, 360.27, 362.49, 364.72, 366.96, 369.22, 371.49, and 373.77 K) with exchanges attempted every 500 steps.

B. Analysis details

Because CMD simulations were run at 300 K, the analysis of both CMD and REMD simulation was performed at 300.00 and 300.37 K for CMD and REMD simulations, respectively. Due to the approximations used in force fields, these temperatures were treated as equal to simplify the analysis.

To check the convergence of the simulations, root-mean-square deviation (RMSD) of Cα atoms using the initial structure as a reference, radius of gyration (*R*g), and maximum distance from any Cα atom to the center of mass (*R*g-max) were calculated. These values were also used to calculate free energy maps using a binning of 0.2 Å in following equation:

$$\Delta F_n = -RT \ln \frac{I_n}{\sum_{a=1}^{N} I_a}, \quad (1)$$

where ΔF_{n} is the free energy at the *n*th bin, *R* is the universal gas constant, *T* is the absolute temperature of the thermostat, I_n is the number of structures in the nth bin, and *N* is the total number of structures.

Additionally, RMSD values to the U-shape and S-shape structures of Aβ42 were calculated. Core residues 17–42 of the first chain of 2MXU and 2BEG pdb structures were used as a reference of U-shape and S-shape conformations, respectively.

The secondary structure was calculated using the DSSP algorithm[41] implemented in the cpptraj tool in Amber package[34] in two variants: (i) using equilibrated part of the simulation and (ii) using several smaller periods from whole simulation. The second

variant was calculated to monitor changes in the secondary structure appearing during simulation. For this purpose, CMD simulation was split to 20 parts (each covering 500 ns), and REMD simulation was split to 24 parts (each covering 25 ns). Four types of secondary structures are distinguished in the analysis: α-helices, β-sheets, turns, and coils. The solvent accessible surface area was calculated for all residues using the cpptraj tool, while Collision Cross Sections (CCSs) were predicted using Mobcal software[42] for 50 snapshots for the second halves of the simulations with the Projection Approximation (PA) method, Exact Hard Sphere Scattering (EHSS) method, and Trajectory Method (TM). The latter method should produce most accurate results; however, all of them were analyzed to give broader view of the investigated systems. Additionally, to estimate expositions of the residues to the solvent, per residue and total SASA values were calculated using the Linear Combinations of terms composed from Pairwise Overlaps (LCPO) method of Weiser *et al.*[43]

Contacts between residues were calculated as normalized values for every residue. If any heavy atom in residue i is closer than 4 Å to any other heavy atom in a residue j, which is separated by at least one residue from residue i, then the contact between these residues in a given snapshot is equal to 1. If all heavy atoms in residue i are further than 4 Å from heavy atoms in residue j, then contact is equal to 0. Contact maps were calculated as averages of all snapshots from the equilibrated parts of the simulations. Additional to contacts, average correlations of the residue motion vectors were calculated over second halves of each simulation using Eq. (2),

$$AvgCorr(a,b) = \frac{\sum_{i=1}^{N} V_a(i) \cdot V_b(i)}{N}, \tag{2}$$

where $V_a(i)$ is a motion vector of residue a in frame i (comparing to the frame $i-1$) and N is the total number of analyzed snapshots.

As a special case of contacts, the number of the intrachain hydrogen bonds was calculated using both distance cutoff criterion below 3 Å for heavy atoms from the donor and acceptor and an angle cutoff of 135°.

Representative structures from the simulations were calculated by clustering ensembles of structures to 5 groups using the hierarchical agglomerative average-linkage method and determining the cluster centroid—structure closest to the average structure in the cluster. Representative structures from the largest cluster were used as a reference in further analysis.

Root-mean square fluctuation (RMSF) was used to analyze fluctuation of atoms, as in Eq. (3),

$$RMSF_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(r_{i,j} - r_{0,j}\right)^2}, \tag{3}$$

where n is the number of analyzed snapshots, $r_{i,j}$ is the position of atom j in snapshot i, and $r_{0,j}$ is the position of atom j in the reference structure, for which the representative structure of the largest cluster was used. For calculation of Hα, Cα, Cβ, and N chemical shifts, a SHIFTS[44] 5.5 software was used. The Pearson correlation coefficient (R) was calculated by comparing predicted values with experimental data.[45]

The energy decomposition was performed using AmberTools package with the same conditions as the simulations were performed, while the entropy was estimated using the normal-mode analysis included in the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) tool.

III. RESULTS AND DISCUSSION

A. Structural analysis

In the first step, all simulations were tested if they converged in matters of RMSD and *R*g (Figs. S2 and S3). As it can be seen, the initial structure of Aβ42 is usually stable for a short period of time and then it refolds and restructures. Because monomeric Aβ42 is an intrinsically disordered protein, it is not possible to say that simulation equilibrate in the meaning that, e.g., rmsd stabilized and maintained on the same level from some point in the simulation. However, the presence of repetitions of some structural changes in the protein (e.g., an increase in RMSD from around 5 Å to around 15 Å at 6.7 and 8.3 μs of the CMD simulation in Amber ff14sb_IDPs force field) indicates that the simulation is indeed converged. For that reason, second half of the simulations (5–10 and 0.3–0.6 μs for CMD and REMD simulations, respectively) was treated as equilibrated and used in most of the analysis. The range of the RMSD covered by CMD and REMD simulations is similar, but in Amber, ff99sb and ff14sb_IDPs CMD simulations reached higher RMSD values. The REMD method is known to improve conformational sampling of the system because of the possibility to overcome energy barriers at high temperature replicas.[46] However, monomeric Aβ42 is generally unstructured and only shallow energy barriers need to be overcome to change its conformation. Because REMD plot shows only one temperature, higher energy and presumably more unfolded conformations migrate to higher temperatures during the simulation. Because the time of presence of each structure at a given temperature is limited by the number of steps between successful exchanges, it is possible that this may slightly limit the occurrence of highly unfolded conformations at this temperature. Therefore, in some cases, higher RMSD values can be captured by CMD simulation, rather than REMD at the same temperature. In an ideal scenario, with perfect sampling and infinite equilibration time between exchanges, such situation would not happen, but due to the computational restrictions, the systems are not truly equilibrated in real-life simulations. Sampling of both CMD and REMD runs can be improved by using multiple starting conformations and by optimizing simulation parameters of the REMD approach. To find optimal conditions for each REMD simulation (range and distribution of temperatures, exchange rates, and time between exchanges), a set of training simulations is required, which increases the computational cost.

The analysis of the RMSD and *R*g plots (Figs. S2 and S3) shows that Amber force fields display much smaller fluctuations than CHARMM force fields. In the CMD runs, Amber ff99sb overstabilizes the Aβ42 structure the most—for most of the time, RMSD and *R*g fluctuations are small. Amber ff14sb also overstabilizes the structure but allows for larger frequent changes than ff99sb. Modified Amber ff14sb shows similar features like the nonmodified version; however, it allows for larger and much more frequent conformational changes (e.g., rapid jumps of rmsd from 5 to 15 Å). Similar behavior is observed for modification of the CHARMM36 force field (CHARMM36m), in which fluctuations are the largest and larger than in the nonmodified version.

For two standard Amber force fields: ff99sb and ff14sb, the behavior of the REMD simulation in comparison with the conventional one is consistent with a general purpose of enhanced sampling of the simulations—local fluctuation of the RMSD and *Rg* are much higher than in CMD simulation. However, for ff14sb_IDPs in REMD run, they are smaller than in the CMD one comparing to the standard Amber force fields. ff14sb_IDPs was optimized using CMD simulations, which may be the reason of the rapid changes in REMD variant observed in our simulations, which are limiting local changes and making the structure to fluctuate closely to the same conformation. In case of CHARMM force fields, both of them exhibit very large conformational changes and no quasistable structures can be detected. The presence of highly unfolded conformations of IDPs was one of the goals achieved during the CHARMM36 refinement, but this resulted in artificially high Rg values for some of benchmarked peptides and proteins[33] what was also observed in our studies.

B. Free energy maps

Maps of the free energy are confirming the general observations coming from the analysis of RMSD and *Rg* plots separately (Fig. 1). For CHARMM force field, the REMD method improves sampling of the conformational space. However, for Amber force field, this effect is not so obvious. For the ff14sb_IDPs, analysis of the REMD simulation at single temperature shows worse search of the conformational space than for the CMD runs. For ff99sb force field, the free energy minimum is larger for REMD; however, there are very few structures with RMSD above 10 Å, which are abundantly present in the CMD simulation. For ff14sb force field, the ranges of RMSD and *Rg* are similar for the REMD and CMD; however, the minimum is much larger for the conventional variant, indicating that conformations can more freely change in CMD simulation. The observed *Rg* values for the main basin of the structures with the minimum free energy in Amber force fields are in the range of the experimentally found hydrodynamics radius of 8.9–9.1 Å,[47] while in CHARMM, they are usually shifted to higher values although they are still in line with previous computational results.[48] Also, only in CHARMM force fields, highly unfolded structures ($Rg > 18$ Å) are present in both sampling methods. Such behavior is also confirmed by the Cartesian Principal Component Analysis (PCA),[49] which can be used to efficiently study protein states and dynamics.[50,51] The PCA free energy maps (Fig. S4) are much broader for the CHARMM force fields than for Amber, and deep energy minima are observed only in the latter. On the other hand, the dihedral angle PCA (dPCA) shows the presence of multiple shallow free energy minima in all force fields (Fig. S5) due to the proper separation of internal and overall motions,[52] indicating the presence of many semistable intermediate structures.

C. Representative structures

Assignment of the structures to the given clusters and sizes of the largest clusters are in agreement with the flexibility of the conformations during simulations (Table S1). The larger the flexibility is, the smaller the top cluster is. Therefore, top clusters from CMD trajectories are significantly larger than from REMD (with the exception of CHARMM36m). Representative structures are less structured for CMD than for REMD simulations in Amber [Figs. 2(e)–2(j)], while in CHARMM [Figs. 2(a)–2(d)], the tendency is opposite. The difference is especially visible for standard Amber force fields: ff99sb and ff14sb, for which significantly better sampling in REMD variant allowed to find better representative structures. In CHARMM force fields, sampling is good even in CMD runs, so further increase of the flexibility caused appearance of multiple diverse structures, which prevail smaller clusters. As a result, top cluster contained too wide variety of the conformations and cluster representative structure is less structured.

RMSD analysis of obtained representative structures (Table S2) shows that there are large differences between structures coming from different force fields and sampling methods. These differences are larger for CHARMM force fields, which is in agreement with higher flexibility. Differences between structures from CMD and REMD simulations in CHARMM force fields are comparable with differences between 5 representative structures from each sampling methods. Surprisingly, two most similar structures from all 50 representative models are clusters 1 from Amber FF14SB and FF99SB REMD simulations with a RMSD of 3.5 Å.

D. RMSF

The fluctuation of the structures [Figs. 3(a) and 3(b)] of Aβ42 in Amber force fields is similar to observations in other studies, in which Aβ9-42 variant exhibited the largest fluctuations for the 23–29 fragment.[53] Except the CHARMM force fields, for most of the simulation time, Aβ42 monomer is in a globular form with very flexible terminal parts. In CHARMM force fields, often Aβ42 is more stretched, and therefore the observed RMSF values are higher.

Higher fluctuations of the N-terminus compared to the C-terminus are in agreement with the previous studies and are most probably caused by the unique character of the last two residues, which constitute to a difference in structure and aggregation rate between Aβ42 and Aβ40 and Aβ41.[54–56]

E. Formation of fibril-prone structures

Similarities of obtained monomeric conformations to two popular shapes of Aβ42 fibril: U-shape and S-shape,[57] were investigated for all snapshots (Table S3 and Figs. S6 and S7). In all Amber force fields, such structures were not observed, while in CHARMM36, a low number of snapshots similar to the S-shape and in CHARMM36m, a higher number of snapshots similar to the U-shape (up to around 0.2% of all snapshots in REMD simulation) were found. The possibility of the formation of such structures in the CHARMM force fields is probably a result of much higher flexibility of Aβ42 and better sampling of the conformational space in these force fields. However, even a small number of fibrillike conformations for the monomeric system may indicate that in the presence of other molecules, an oligomer of the protofibrillike structure can be formed.[58,59] Similarity of up to 0.5% of the obtained structures of Aβ40 in Amber ff99sb to S-shape conformations was recently observed for simulations without copper, while in the presence of copper, even smaller presence of structures similar to the U-shape was found[60] indicating that even small changes to the system can significantly change tendency to form fibril-prone structures.

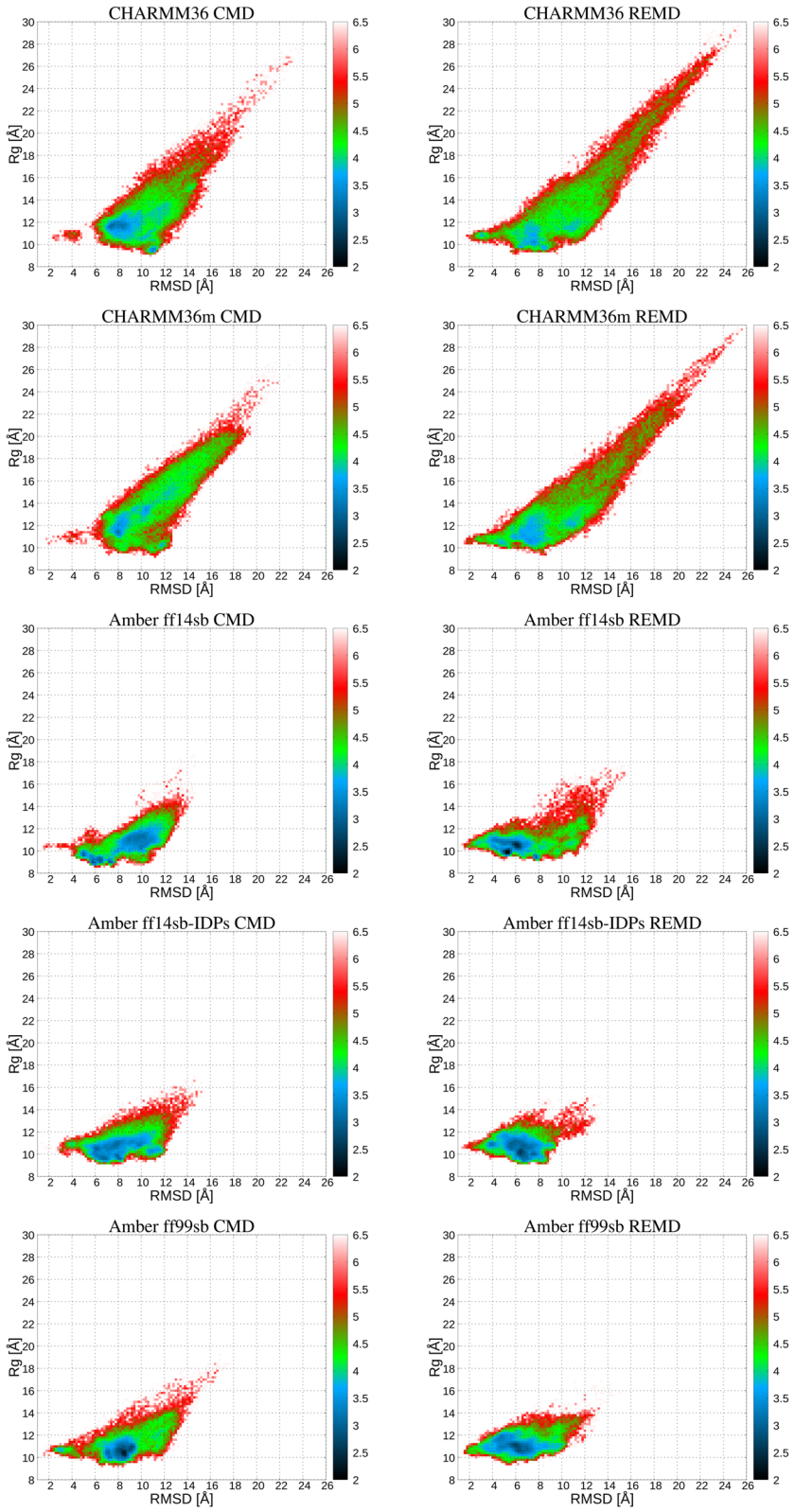

FIG. 1. Free energy maps (kcal/mol) for CMD (left panels) and REMD (right panels) simulations.

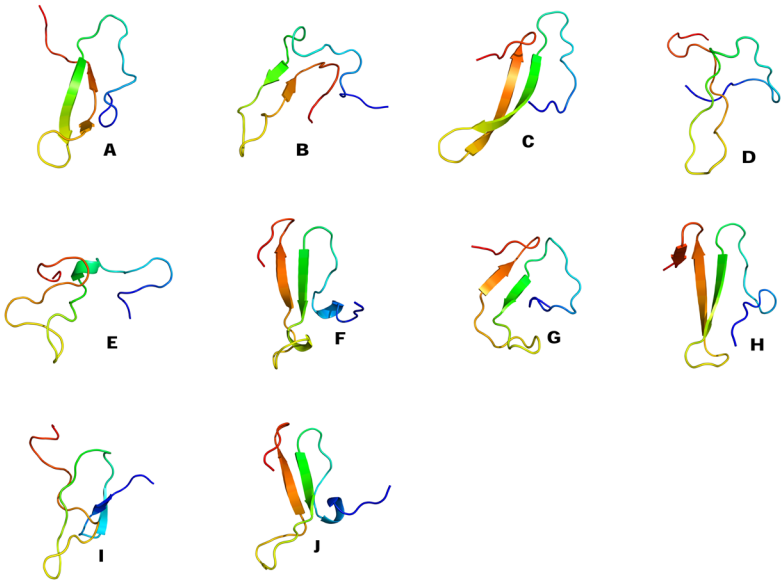

FIG. 2. Cartoon representations of the cluster representative structures for CHARMM36 [(a) and (b)], CHARMM36m [(c) and (d)], Amber ff14sb [(e) and (f)], ff14sb_IDPs [(g) and (h)], and ff99sb [(i) and (j)] obtained from the clustering second halves of the CMD and REMD simulations, respectively.

F. Solvent accessible surface area (SASA) and CCS

SASA values are significantly higher for CHARMM than for Amber force fields with the difference of about 1000 Å^2 [Table I and Figs. 4(a) and 4(b)]. Distribution of the SASA values from REMD simulations is broader and shifted to higher values than from CMD runs. The presence of a second SASA peak at around 5000–5500 Å^2 [Fig. 4(b)] suggests an occurrence of significant fraction of largely unfolded structures in REMD runs, caused by the migration to and from higher temperature replicas. It may indicate that REMD provided better sampling than CMD which failed to recover the particular state related to the second peak. SASA values around 3200 Å^2 were reported previously from Amber ff99SB for monomeric Aβ42,[48] while values around 4200 Å^2 were found for the single chains of Aβ40 using CHARMM36 force field[61] and for monomeric Aβ42 in Amber03.[62] The analysis of the contributions of each residue into the total SASA shows that, in general, the behavior of residues is understandable, hydrophobic residues and glycine have the smallest SASA contributions (Ala2, Gly9, Ala21, Gly25, Gly29, Gly33, Gly37, and Gly38) [Figs. 5(a)–5(c)]. However, interestingly, the biggest difference between Amber and CHARMM force fields is in Lys16, which especially in Amber ff14sb_IDPs is completely buried inside the protein (as a part of a β-strand), while in CHARMM force fields, it is one of the most solvent-exposed residues. Lys16 (along with Leu17) is in the α-secretase cleavage site, and its substitution to other amino acid residue can lead to significant decrease in toxicity of Aβ42;[63] therefore, its position and availability for solvent and enzymes may play a very important role in Alzheimer disease.

CCS values were calculated using three methods in Mobcal tool: EHSS, which should provide too high values, and PA and TM, which should be closer to real values. Tendency of the changes between CCS values is consistent with the SASA. Observed values (Table II)

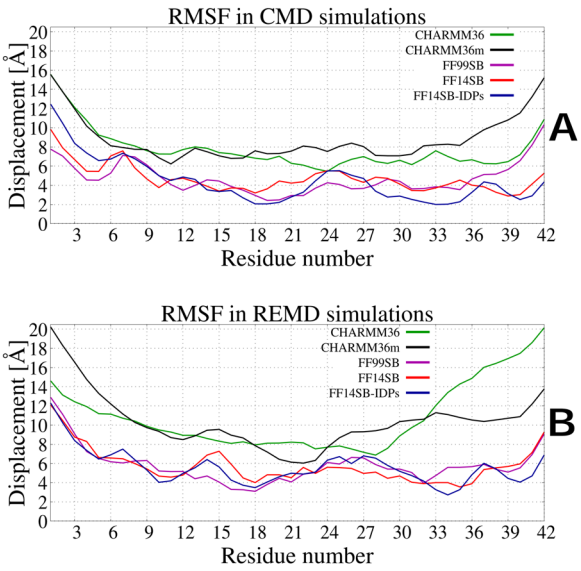

FIG. 3. RMSF values calculated from the second halves of CMD [panel (a)] and REMD [panel (b)] simulations.

TABLE I. Average SASA values (nm^2).

	CMD	REMD
CHARMM36	41.62 ± 3.07	46.01 ± 4.39
CHARMM36m	42.36 ± 3.77	45.89 ± 5.37
Amber ff14SB	33.43 ± 2.13	33.52 ± 2.61
Amber ff14SB_idps	32.85 ± 2.47	34.42 ± 2.40
Amber ff99SB	32.46 ± 2.06	35.19 ± 2.10

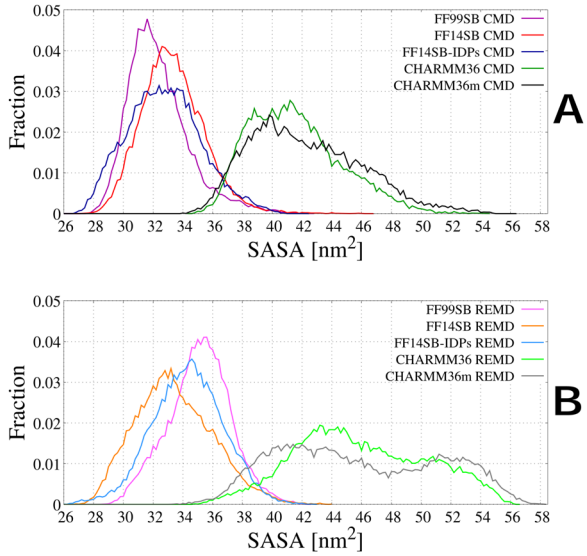

FIG. 4. Diagrams of the populations of structures with given solvent accessible surface areas for CMD [panel (a)] and REMD [panel (b)] simulations.

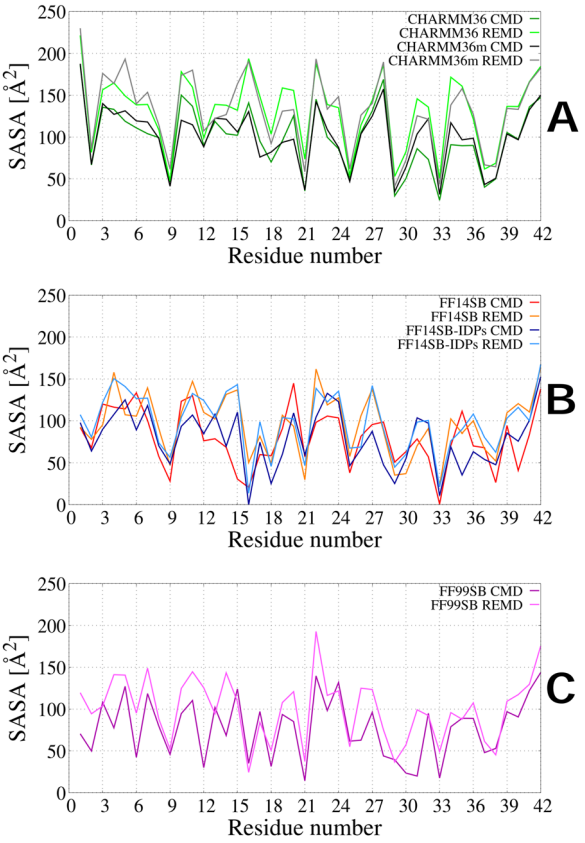

FIG. 5. Diagrams of SASA per residues values for CHARMM (a) and Amber [(b) and (c)] force fields.

are well corresponding to the experimentally measured values of the CCS for monomeric Aβ42, equal to 702 and 774 Å^2 for −3 and −4 total charges, respectively.[64] Only CHARMM36 with REMD and CHARMM36m provided slightly too high CCS values, which are a result of the high content of unfolded structures. Because of that, structures from REMD possess higher CCS values than these from CMD runs.

G. Contact maps and correlations

Contact maps were calculated using snapshots collected in the second halves of the simulations; therefore, contacts are averaged over a vast number of different conformations, making them difficult to be interpreted. Some residues in the middle of the chain (region 14–20, depending on the force field) can interact during simulation with almost all other residues (Fig. S8). Additionally, some conformations, especially in CHARMM force fields, form very small number of intrachain hydrogen bonds, indicating that they are very unfolded (Fig. S9); however, one of the residues, Val18, seems to play a special role in most of the force fields. It strongly interacts with Ile31, Gly33, and Leu34, respectively, in ff14sb_IDPs, ff99sb, and CHARMM36, forming a hydrophobic core of monomeric Aβ42. The special role of Val18 was experimentally disclosed by Soto *et al.*, who found that the substitution of this residue to alanine causes an increase in α-helical content in Aβ42 and a drastic reduction in fibrilogenesis.[65]

Because the analysis of raw averaged contact maps is difficult due to the presence of many residues being accidentally close to others, correlations between movements between every pair of residues were calculated for every simulation (Fig. S10). In all force fields, in CMD simulations, the correlation between residues 17–21 and 31–36 can be observed, except of CHARMM force fields in REMD simulations, for which due to the very large flexibility of the system no significant correlations were found. In Amber force fields, in REMD scheme, the range of highly correlated residues is shifted to 11–18 and 35–42. In all cases, the important role of Val18 in interactions and movement correlation is preserved.

H. Secondary structures

Except of the ff14sb, for all force-fields, the β-content in the CMD simulations is higher than in the REMD variant (Table III). It is especially visible for both CHARMM force fields, for which the average β-content is more than two-fold higher than in the CMD run due to the high flexibility of the amyloid chain in these force fields. This may be a typical behavior due to the intrinsically disordered nature of monomeric Aβ42 caused by the low stability of semistable conformations, but it may also be caused by the tendency of the CHARMM force fields to overpopulate highly unfolded conformations for some proteins.[33]

The average β-content for CMD trajectories is on the level 15%–34%, while α-content is in the 0.1%–9.0% range, which is roughly in agreement with experimental circular dichroism spectroscopy findings, which estimates α-helical content between 3% and 9%, and β-content of 12%–25%.[66,67] Amber ff14sb overestimated

TABLE II. CCS values (nm^2).

		PA	EHSS	TM	Average
CHARMM36	CMD	7.68 ± 0.59	9.24 ± 0.65	6.54 ± 0.61	7.818
	REMD	8.45 ± 1.09	10.10 ± 1.20	7.03 ± 0.55	8.528
CHARMM36m	CMD	7.88 ± 0.81	9.46 ± 0.88	6.69 ± 0.61	8.009
	REMD	8.44 ± 1.15	10.08 ± 1.24	7.06 ± 0.37	8.527
ff14sb	CMD	6.84 ± 0.28	8.27 ± 0.34	6.48 ± 0.75	7.198
	REMD	6.73 ± 0.35	8.15 ± 0.42	6.28 ± 0.63	7.052
ff14sb_idps	CMD	6.65 ± 0.38	8.03 ± 0.42	6.93 ± 0.18	7.203
	REMD	6.89 ± 0.34	8.35 ± 0.41	6.54 ± 0.72	7.259
ff99sb	CMD	6.73 ± 0.38	8.11 ± 0.43	7.00 ± 0.22	7.279
	REMD	6.96 ± 0.39	8.42 ± 0.39	6.66 ± 0.57	7.348

the α-helical content in REMD run, while ff99sb overestimated β-content in both simulation types. Such overfavor of the secondary structure remains a significant problem for force field optimization even for regular, stable proteins.[30]

In most of the stable structures, there are at least two β-strands, which are formed by residues 20–23 and 25–28, while residue 24 is usually the one which is bending two strands to each other [Figs. 6(a)–6(J)]. α-helical content is low in all force fields, and only very short (up to 4 residues) α-helices are formed temporarily during simulations. The most likely regions are 4–8, 13–16, and 21–25.

NMR studies suggested a presence of an antiparallel β-sheet between residues 16–21 and 29–36 for monomeric Aβ42,[68] which is in very good agreement with CHARMM36m CMD, FF14SB-IDPs REMD, and partially FF14SB REMD and FF99SB REMD. Also formation of a short β-strand by the C-terminal residues is consistent with the experimental data[68] and suggested to be connected to the oligomerization of the Aβ.[69]

Average secondary content in the CMD simulations is fluctuating (Fig. S11) so much that there is no clear convergence of the simulation and values are periodically increasing and decreasing—it is reasonable because of the disordered character of monomeric Aβ42. In REMD, the results seem to be converged in the second half of the simulation due to taking a much better averaging over multiple replicas and independence from the initial structure. Both behaviors suggest that the simulations are long enough to reach equilibrated states and reasonable averages. However, especially in the case of Amber force fields, it may be beneficial for sampling to perform REMD runs using different initial structures for different replicas. Previous studies provided a wide variety of secondary structure content, e.g., 10% of β- and 4.2% of α-content were reported by Pham *et al.* using REMD simulations in FF99SB for monomeric Aβ42,[48] and similar values were observed by Ball *et al.* with the use of the same methods but different solvents (TIP4P-Ew),[68] while Rosenman *et al.* reported 20% β- and below 5% of α-contents in OPLS force field with REMD simulations.[70] The same values were

TABLE III. Average secondary structure content (%) during simulations.

		Beta	Alpha	Turn	Coil
CHARMM36	CMD	20.1 ± 7.9	0.7 ± 2.4	8.9 ± 5.5	70.3 ± 10.6
	REMD	6.2 ± 7.6	3.8 ± 5.2	10.9 ± 6.5	0.791 ± 12.0
CHARMM36m	CMD	23.8 ± 14.6	2.1 ± 3.9	12.1 ± 5.6	62.0 ± 15.8
	REMD	10.6 ± 9.1	1.2 ± 3.1	8.1 ± 5.7	80.1 ± 12.0
ff14sb	CMD	14.8 ± 5.7	8.7 ± 6.3	17.8 ± 7.1	58.6 ± 8.5
	REMD	20.1 ± 9.3	11.7 ± 10.4	17.7 ± 6.3	50.5 ± 9.0
ff14sb_idps	CMD	31.5 ± 5.5	2.7 ± 4.9	16.4 ± 4.7	49.4 ± 6.8
	REMD	26.9 ± 8.6	2.0 ± 3.7	14.9 ± 5.6	56.2 ± 11.0
ff99sb	CMD	34.0 ± 7.0	1.3 ± 3.0	14.8 ± 5.6	50.0 ± 9.4
	REMD	27.9 ± 8.8	3.8 ± 6.1	15.6 ± 7.0	52.6 ± 11.6

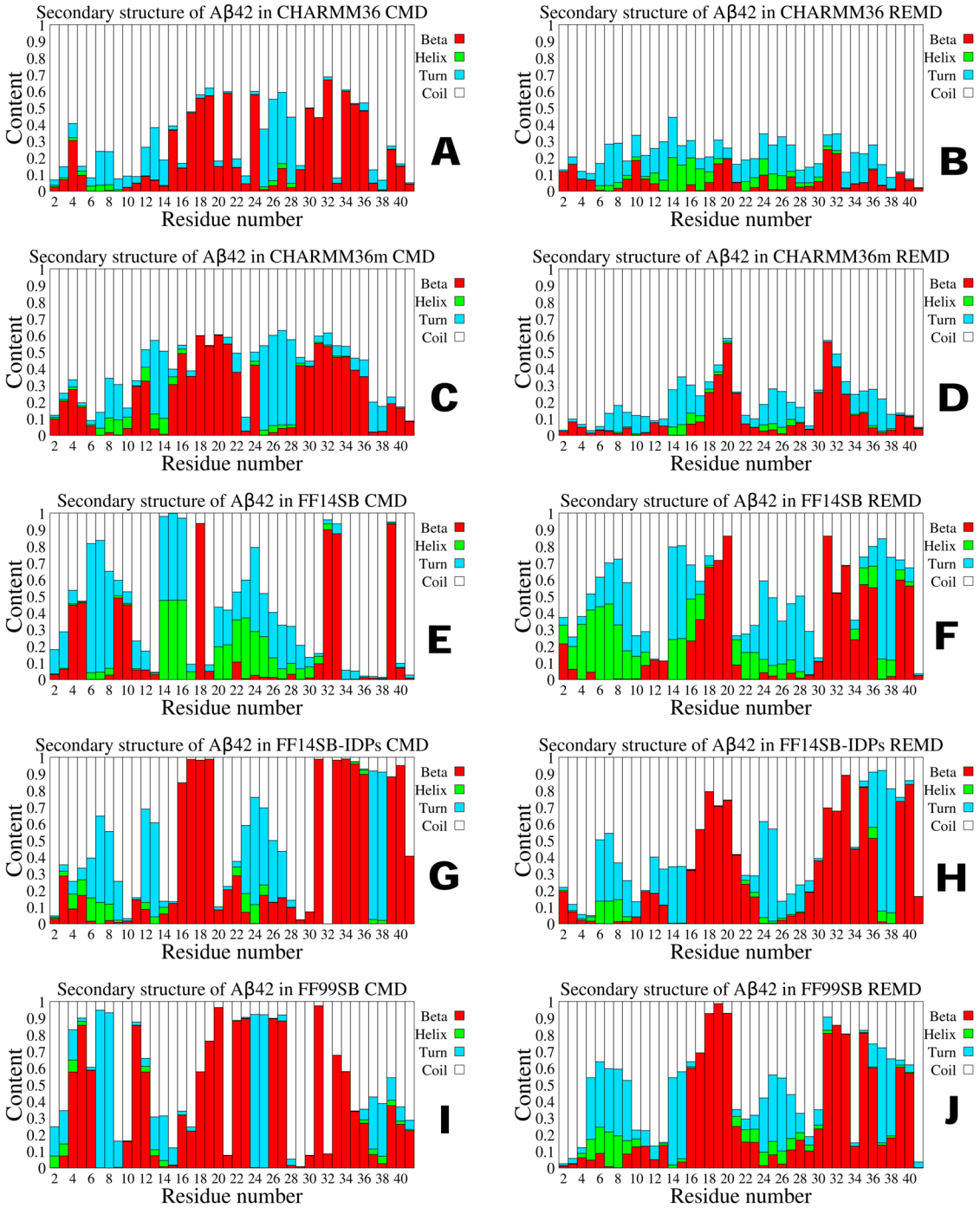

FIG. 6. Diagrams of secondary structure per residue for CMD [(a), (c), (e), (g), and (i)] and REMD [(b), (d), (f), (h), and (j)] simulations, left and right panels, respectively.

obtained by Carballo-Pacheco and Strodel for CHARMM22; however, they obtained around 9% of both α- and β-contents in OPLS.[71] Weber and Uversky simulated monomeric Aβ42 with different water concentrations in Amber FF99SB and CHARMM22/CMAP, obtaining very diverse α-helical content of 6%–14% and 6%–43% at 310 K, respectively.[72] All these results prove that the obtained secondary structure of Aβ42 monomer strongly depends on simulation and analysis details.

TABLE IV. Correlation coefficients (R) for chemical shifts.

		Chemical shifts				
		Hα	Cα	Cβ	N	Average
CHARMM36	CMD	0.717	0.991	0.995	0.890	0.898
	REMD	0.845	0.994	0.998	0.924	0.940
CHARMM36m	CMD	0.729	0.993	0.996	0.916	0.908
	REMD	0.893	0.994	0.999	0.912	0.949
ff14sb	CMD	0.591	0.979	0.995	0.817	0.845
	REMD	0.878	0.990	0.997	0.913	0.945
ff14sb_idps	CMD	0.827	0.988	0.996	0.870	0.920
	REMD	0.841	0.994	0.998	0.930	0.941
ff99sb	CMD	0.717	0.986	0.994	0.746	0.861
	REMD	0.774	0.988	0.994	0.856	0.903

I. Chemical shifts

Analysis of the chemical shifts predicted for the structures from the simulations shows that in all cases, REMD simulations provided better agreement with the experimental data (Table IV and Figs. S12–S15), with the smallest difference for FF14SB_IDPs. Improvement of the sampling is especially significant for conventional Amber force fields: FF99SB and FF14SB. The best overall agreement was observed for CHARMM36m, but FF14SB, CHARMM36, and FF14SB_IDPs were not much worse. Only the oldest force field, FF99SB, gave significantly worse results. Obtained correlations (R) for calculated Cα chemical shifts with the experimentally obtained values are higher than values obtained by Weber and Uversky using Amber FF99SB and CHARMM22/CMAP at 310 K, being 0.960 and 0.965 for 20 Å and 30 Å of water layers, respectively,[72] indicating that simulation length of the simulation is as important as choice of the sampling method and force field. It is consistent with the observation of Lin and Pande, who stated that β-propensity of monomeric Aβ require very long simulation time, longer than obtained by them 200 μs of aggregated simulation time.[73] Obtained in this work, R values are significantly higher than those obtained for CHARMM22 by Carballo-Pacheco and Strodel,[71] which are 0.680, 0.730, and 0.380 for Cα, Cβ, and Hα chemical shifts, respectively. The difference may be caused by different tools used for the estimation of computed chemical shifts and sampling through equilibrated part of longer trajectory.

J. Role of energy components

Decomposition of potential energy was performed for 10 000 snapshots from equilibrated part of trajectory at each temperature and from each simulation type to obtain electrostatic and van der Waals (vdW) components for the solute (Aβ42) and solute-solvent (interactions between Aβ42 and water and counterions) [Figs. 7(a)–7(c), Figs. S16 and S17]. Due to the empirical nature of the force fields, relative energies were compared instead of

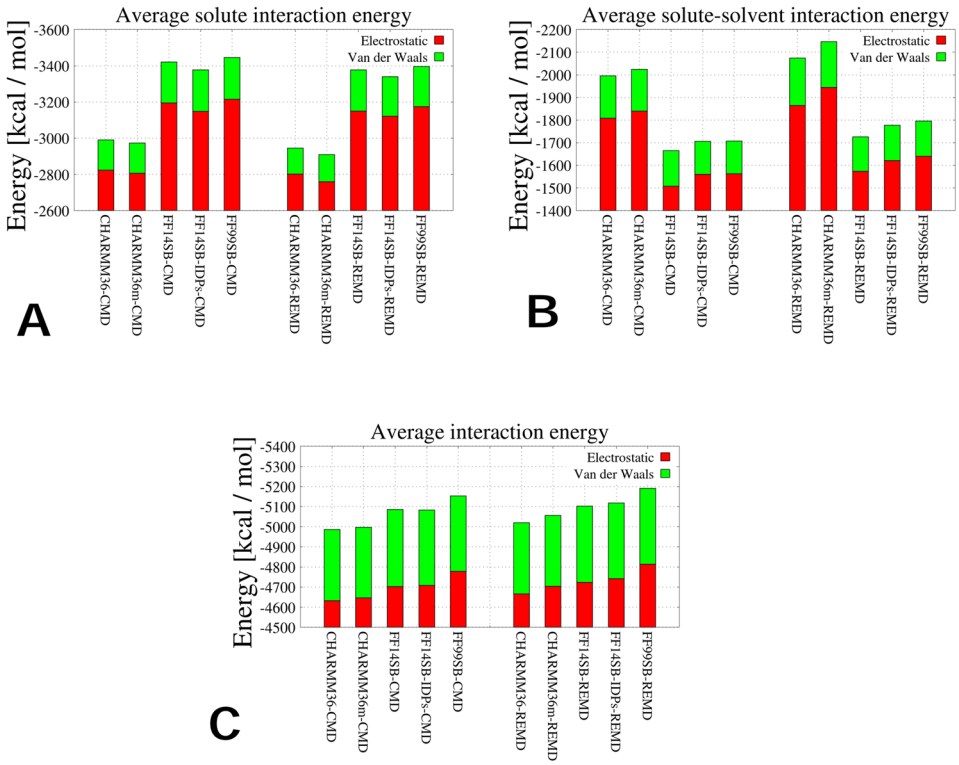

FIG. 7. Average values of the long-range interactions in Aβ42 (a), between Aβ42 and water and ions (b), and combined interaction energies (c).

absolute ones. The nonbonded potential energy of Aβ42 is the lowest in both CHARMM force fields [Fig. 7(a)], while solute-solvent energy is the highest [Fig. 7(b)], indicating that interactions with water and ions play a major role in the structure and dynamics of monomeric Aβ42 in CHARMM force fields. Such behavior is responsible for the high flexibility of Aβ42 in CHARMM force fields [Figs. 3(a) and 3(b)] and lower β-content (Table III) comparing to the Amber force fields. It is especially noticeable for REMD sampling in CHARMM force fields for which solute-solvent interaction is very strong with high fluctuations (Figs. S16 and S17). The strong interaction of Aβ42 with a solvent in the CHARMM force fields is also the reason why in these force fields, the SASA values and hydrophilicity are much higher than in Amber [Figs. 4(a) and 4(b)]. The SASA values seems to correlate well with the solute and solute-solvent vdW interactions; e.g., in CHARMM36m with REMD sampling, the vdW term is the only one with two visible interaction energy minima peaks (Fig. S16) and it is also the only one with two peaks in SASA distribution plot [Fig. 4(b)]. It is an example of a case in which a minor energy contribution can have a significant impact on the structure.

For monomeric Aβ42, the electrostatic component prevails in all conditions, force fields and sampling methods, contributing over 90% to nonbonded energy. In CHARMM force fields, the role of vdW interactions is the smallest, constituting approximately 5.5% and 5% in CMD and REMD simulations, respectively, while it is the greatest for the solute-solvent interactions, constituting approximately 9.2% and 9.8% [Figs. 7(a)–7(c), and Table S4]. Nonbonded potential energy is the highest for Amber FF99SB, for which a particularly electrostatic term is high and has the lowest fluctuations (Figs. S16 and S17 and Table S4), which, along with other high energy components, is the reason why the structure of Aβ42 is overstabilized, staying for most of the time in one deep energy minimum (Fig. 1) and why it possesses the highest β-content (Table III).

The calculated entropy is higher with REMD sampling than CMD (Table S5) and the difference caused by the sampling methods is more pronounced in the CHARMM force fields. The entropy values for the Amber force fields are in the range of 480–490 kcal/mol, while for CHARMM, the observed values are lower, on average, around 400 kcal/mol. The low entropic difference between the unfolded and folded states is favorable to the thermodynamic stability of a protein;[74] however, in the case of a disordered peptide, like Aβ42, it is difficult to assess. The obtained total enthalpy values of Aβ42 (Table S5) are more negative for the Amber force fields, confirming its higher stability in these force fields.

The strange behavior of both variants of the Amber ff14sb force fields was observed—the total potential energy of Aβ42 is often very close to 0 in both CMD and REMD sampling methods (Figs. S16 and S17), which may be the result of their parametrization. The distribution of the electrostatic components for solute-solvent and both vdW terms is significantly wider in CHARMM force fields, especially in the REMD variants [Figs. 7(a)–7(c)]. For Aβ42, electrostatic components slightly decrease (becoming more negative) with increasing temperature for all force fields, while vdW terms decrease in Amber force fields (Fig. S17) with the increase in temperature. For the CHARMM force fields, the vdW term first decrease with increasing temperature and then begins to increase. For both nonbonded terms in all force fields, a rise in temperature causes a significant increase in the solute-solvent interaction energy; therefore, a change of the temperature has major effect on the solute-solvent interactions, rather than on the solute itself.

IV. CONCLUSIONS

In this work, two different sampling methods, long conventional single-trajectory MD simulation and shorter-run replica-exchange MD, were used with five popular all-atom force fields to simulate the structure, dynamics, and flexibility of Aβ42 peptide. Our study shows that the proper choice of the force field and the sampling method is not easy and heavily depends on the studied system, the available computational resources and investigated phenomena. In general, REMD is expected to provide better conformational search due to the possibility to overcome energy barriers with higher temperature replicas. However, while it is essential for conventional systems (e.g., proteins with stable structures), its usage for the intrinsically disordered proteins (IDPs), such as Aβ42 monomer, may not be necessary; therefore, for IDPs, one may safely use the conventional sampling method to reduce the computational cost of the studies.

To show clearly differences between sampling methods, two very contrast approaches were compared: single-trajectory 10 μs CMD runs and 48-trajectory REMD runs with a total time of 28.8 μs. The CMD runs were performed by using GPGPU calculations, which is not possible for REMD simulations, decreasing the computational time by an order of magnitude, resulting in about 30 times lower computational time for CMD sampling. It should be noted that in regular simulations, CMD sampling can be improved by running multiple shorter trajectories starting from different initial conformations; however, for the purpose of this comparison, the most different sampling approaches were used.

Behavior of Aβ42 in standard Amber force fields FF99SB and FF14SB is significantly better with REMD than CMD runs; nevertheless, older variant performed the worst among all tested force fields, significantly overstabilizing the structure and providing the least correct values of the chemical shifts. The use of REMD sampling in CHARMM36 and 36m and Amber FF14SB_IDPs force fields also improved the results, but to a lesser extent, providing satisfactory results with both sampling approaches.

In general, the newer force fields Amber FF14SB and CHARMM36m yielded better results than their older versions Amber FF99SB and CHARMM36, which gives hope that with the development of new force fields, even greater accuracy can be achieved. This supports the claims of the developers that FF14SB and CHARMM36m can be successfully used to study a variety of proteins and peptides, including Aβ42. However, the usability of every force field should be assessed for each system and the studied phenomena separately, before starting production runs.

On the other hand, FF14SB_IDPs is an interesting example of a force field modification that does not improve the results provided by the original version, but it allows us to use CMD simulations instead of the REMD variant and maintains good conformational search and high flexibility of the Aβ42 (Fig. S18). The use of CMD sampling with GPGPU can significantly reduce computational costs

and speed up calculations without compromising much of the accuracy. One should keep in mind that the use of a dedicated force field, such as Amber FF14SB_IDPs, is probably limited to one group of proteins,[24] while other force fields must provide satisfactory results for majority of the systems.

We showed that, especially in CHARMM force fields, the interaction of Aβ42 with a solvent is more important than the interaction within Aβ42 and its role increases with increasing temperature. The total potential energy of Aβ42 only slightly levels up with temperature, while for solute-solvent, the increase is much bigger. The main energy contribution is electrostatic interactions, which is at least an order of magnitude larger than the vdW one; nevertheless, both play an important role in the structure and dynamics of monomeric Aβ42. Our study has revealed the important role of water in determining the structure and dynamics of Aβ42. A stronger solute-solvent interaction in CHARMM compared to Amber resulted in the fact that this peptide is less stable and more hydrophilic in CHARMM. It would be interesting to extend our research to other IDPs.

SUPPLEMENTARY MATERIAL

Supplementary material includes the information on representative structures from MD and REMD simulations; similarity to the S- and U-shape fibril structures; total entropy and enthalpy of Aβ42 and its interaction with solvent; initial structure used in all simulations; time dependence of RMSD and gyration radius for all cases; free energy landscapes obtained by Cartesian and dihedral PCA; time dependence of RMSD to the S- and U-shape structure; contact maps; populations of the structures with given number of intrachain hydrogen bonds; correlations between motions of the residues; correlation plots of predicted and experimentally derived chemical shifts; populations of electrostatic and van der Waals interaction energies; and the most representative structures from CMD simulation in ff14sb_IDPs force field.

ACKNOWLEDGMENTS

This work was supported by the National Science Center (Poland) Grant Nos. 2015/19/B/ST4/02721 and 2017/27/N/NZ1/02871, and Department of Science and Technology, Ho Chi Minh city, Vietnam. Computational resources were also provided by the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk.

REFERENCES

[1] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer, 2010), pp. 425–518.
[2] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell, J. Chem. Theory Comput. **9**(12), 5430–5449 (2013).
[3] P. E. M. Lopes, O. Guvench, and A. D. MacKerell, *Molecular Modeling of Proteins* (Humana Press, New York, NY, 2015), Vol. 1215, pp. 47–71.
[4] P. D. Dans, I. Ivani, A. Hospital, G. Portella, C. González, and M. Orozco, Nucleic Acids Res. **45**(7), 4217–4230 (2017).
[5] K. H. DuBay, M. L. Hall, T. F. Hughes, C. Wu, D. R. Reichman, and R. A. Friesner, J. Chem. Theory Comput. **8**(11), 4556–4569 (2012).
[6] O. F. Lange, D. van der Spoel, and B. L. de Groot, Biophys. J. **99**(2), 647–655 (2010).
[7] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Chem. Rev. **116**(14), 7898–7936 (2016).
[8] S. Marrink and H. Risselada, J. Phys. Chem. B **111**(27), 7812–7824 (2007).
[9] A. Liwo, M. Baranowski, C. Czaplewski, E. Gołaś, Y. He, D. Jagieła, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska, A. Niadzvedtski, S. Ołdziej, H. A. Scheraga, A. K. Sieradzan, R. Ślusarz, T. Wirecki, Y. Yin, and B. Zaborowski, J. Mol. Model. **20**(8), 2306 (2014).
[10] J. Maupetit, P. Tuffery, and P. Derreumaux, Proteins **69**(2), 394–408 (2007).
[11] S. Kmiecik and A. Kolinski, Proc. Natl. Acad. Sci. U. S. A. **104**(30), 12330–12335 (2007).
[12] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**(11), 1333–1338 (1993).
[13] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**(20), 9058–9067 (2002).
[14] A. S. Karczyńska, C. Czaplewski, P. Krupa, M. A. Mozolewska, K. Joo, J. Lee, and A. Liwo, J. Comput. Chem. **38**(31), 2730–2746 (2017).
[15] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**(1-2), 141–151 (1999).
[16] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, and K. Schulten, J. Comput. Chem. **28**(16), 2618–2640 (2007).
[17] H. Nymeyer, J. Chem. Theory Comput. **4**(4), 626–636 (2008).
[18] K. A. Beauchamp, Y.-S. Lin, R. Das, and V. S. Pande, J. Chem. Theory Comput. **8**(4), 1409–1414 (2012).
[19] P. Florová, P. Sklenovský, P. Banáš, and M. Otyepka, J. Chem. Theory Comput. **6**(11), 3569–3579 (2010).
[20] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, J. Phys. Chem. B **119**(16), 5113–5123 (2015).
[21] R. B. Best, W. Zheng, and J. Mittal, J. Chem. Theory Comput. **10**(11), 5113–5124 (2014).
[22] R. B. Best and G. Hummer, J. Phys. Chem. B **113**(26), 9004–9015 (2009).
[23] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, Proteins **78**(8), 1950–1958 (2010).
[24] D. Song, R. Luo, and H.-F. Chen, J. Chem. Inf. Model. **57**(5), 1166–1178 (2017).
[25] P. Robustelli, S. Piana, and D. E. Shaw, Proc. Natl. Acad. Sci. U. S. A. **115**(21), E4758–E4766 (2018).
[26] A. Sandoval-Perez, K. Pluhackova, and R. A. Böckmann, J. Chem. Theory Comput. **13**(5), 2310–2321 (2017).
[27] H. B. Nam, M. Kouza, H. Zung, and M. S. Li, J. Chem. Phys. **132**(16), 165104 (2010).
[28] P. H. Nguyen, M. S. Li, and P. Derreumaux, Phys. Chem. Chem. Phys. **13**(20), 9778 (2011).
[29] C. R. Watts, A. Gregory, C. Frisbie, and S. Lovas, Proteins Struct. Funct. Bioinf. **86**(3), 279–300 (2018).
[30] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins Struct. Funct. Bioinf. **65**(3), 712–725 (2006).
[31] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**(8), 3696–3713 (2015).
[32] J. Huang and A. D. MacKerell, J. Comput. Chem. **34**(25), 2135–2145 (2013).
[33] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, Nat. Methods **14**(1), 71–73 (2017).
[34] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, Comput. Phys. Commun. **91**(1-3), 1–41 (1995).
[35] M. Yang and D. B. Teplow, J. Mol. Biol. **384**(2), 450–464 (2008).
[36] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**(2), 926–935 (1983).
[37] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**(5), 1902 (1998).
[38] S. Jo, T. Kim, V. G. Iyer, and W. Im, J. Comput. Chem. **29**(11), 1859–1865 (2008).
[39] S. Jo, X. Cheng, J. Lee, S. Kim, S.-J. Park, D. S. Patel, A. H. Beaven, K. I. Lee, H. Rui, S. Park, H. S. Lee, B. Roux, A. D. MacKerell, J. B. Klauda, Y. Qi, and W. Im, J. Comput. Chem. **38**(15), 1114–1124 (2017).

[40]A. Patriksson and D. van der Spoel, Phys. Chem. Chem. Phys. **10**(15), 2073 (2008).
[41]W. Kabsch and C. Sander, Biopolymers **22**(12), 2577–2637 (1983).
[42]M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold, J. Phys. Chem. **100**(40), 16082–16086 (1996).
[43]J. Weiser, P. S. Shenkin, and W. C. Still, J. Comput. Chem. **20**(2), 217–230 (1999).
[44]K. Osapay and D. A. Case, J. Am. Chem. Soc. **113**(25), 9436–9444 (1991).
[45]L. Hou, H. Shao, Y. Zhang, H. Li, N. K. Menon, E. B. Neuhaus, J. M. Brewer, I. J. L. Byeon, D. G. Ray, M. P. Vitek, T. Iwashita, R. A. Makula, A. B. Przybyla, and M. G. Zagorski, J. Am. Chem. Soc. **126**(7), 1992–2005 (2004).
[46]R. Zhou, Methods Mol. Biol. **350**, 205–223 (2007).
[47]S. Nag, B. Sarkar, A. Bandyopadhyay, B. Sahoo, V. K. A. Sreenivasan, M. Kombrabail, C. Muralidharan, and S. Maiti, J. Biol. Chem. **286**(16), 13827–13833 (2011).
[48]P. D. Q. Huy, Q. V. Vuong, G. La Penna, P. Faller, and M. S. Li, ACS Chem. Neurosci. **7**(10), 1348–1363 (2016).
[49]A. Kitao, F. Hirata, and N. Gō, Chem. Phys. **158**(2-3), 447–472 (1991).
[50]G. G. Maisuradze, P. Senet, C. Czaplewski, A. Liwo, and H. A. Scheraga, J. Phys. Chem. A **114**(13), 4471–4485 (2010).
[51]C. C. David and D. J. Jacobs, Methods Mol. Biol. **1084**, 193–226 (2014).
[52]A. Altis, P. H. Nguyen, R. Hegger, and G. Stock, J. Chem. Phys. **126**(24), 244111 (2007).
[53]W. M. Berhanu and U. H. E. Hansmann, PLoS One **7**(7), e41479 (2012).
[54]N. G. Sgourakis, Y. Yan, S. A. McCallum, C. Wang, and A. E. Garcia, J. Mol. Biol. **368**(5), 1448–1457 (2007).
[55]Y. Yan and C. Wang, J. Mol. Biol. **364**(5), 853–862 (2006).
[56]H. L. Nguyen, T. Thi Minh Thu, P. M. Truong, P. D. Lan, V. H. Man, P. H. Nguyen, L. A. Tu, Y.-C. Chen, and M. S. Li, J. Phys. Chem. B **120**(30), 7371–7379 (2016).
[57]Y. Xiao, B. Ma, D. McElheny, S. Parthasarathy, F. Long, M. Hoshi, R. Nussinov, and Y. Ishii, Nat. Struct. Mol. Biol. **22**(6), 499–505 (2015).
[58]L. C. Serpell, Biochim. Biophys. Acta **1502**(1), 16–30 (2000).
[59]S. Kumar and J. B. Udgaonkar, *Current Science* (Current Science Association, 2010), pp. 639–656.
[60]D. Q. H. Pham, M. S. Li, and G. La Penna, J. Phys. Chem. B **122**(29), 7243–7252 (2018).
[61]C. R. Watts, A. J. Gregory, C. P. Frisbie, and S. Lovas, Proteins Struct. Funct. Bioinf. **85**(6), 1024–1045 (2017).
[62]S. Duan, X. Guan, R. Lin, X. Liu, Y. Yan, R. Lin, T. Zhang, X. Chen, J. Huang, X. Sun, Q. Li, S. Fang, J. Xu, Z. Yao, and H. Gu, Neurobiol. Aging **36**(5), 1792–1807 (2015).
[63]D. Kaden, A. Harmeier, C. Weise, L. M. Munter, V. Althoff, B. R. Rost, P. W. Hildebrand, D. Schmitz, M. Schaefer, R. Lurz, S. Skodda, R. Yamamoto, S. Arlt, U. Finckh, and G. Multhaup, EMBO Mol. Med. **4**(7), 647–659 (2012).
[64]S. L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condron, G. Bitan, D. B. Teplow, J.-E. Shea, B. T. Ruotolo, C. V. Robinson, and M. T. Bowers, Nat. Chem. **1**(4), 326–331 (2009).
[65]C. Soto, E. M. Castaño, B. Frangione, and N. C. Inestrosa, J. Biol. Chem. **270**(7), 3063–3067 (1995).
[66]M. D. Kirkitadze, M. M. Condron, and D. B. Teplow, J. Mol. Biol. **312**(5), 1103–1119 (2001).
[67]K. Ono, M. M. Condron, and D. B. Teplow, Proc. Natl. Acad. Sci. U. S. A. **106**(35), 14745–14750 (2009).
[68]K. A. Ball, A. H. Phillips, D. E. Wemmer, and T. Head-Gordon, Biophys. J. **104**(12), 2714–2724 (2013).
[69]B. Urbanc, L. Cruz, S. Yun, S. V. Buldyrev, G. Bitan, D. B. Teplow, and H. E. Stanley, Proc. Natl. Acad. Sci. U. S. A. **101**(50), 17345–17350 (2004).
[70]D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, and A. E. García, J. Mol. Biol. **425**(18), 3338–3359 (2013).
[71]M. Carballo-Pacheco and B. Strodel, Protein Sci. **26**(2), 174–185 (2017).
[72]O. C. Weber and V. N. Uversky, Intrinsically Disord. Proteins **5**(1), e1377813 (2017).
[73]Y.-S. Lin and V. S. Pande, Biophys. J. **103**(12), L47–L49 (2012).
[74]S. Dagan, T. Hagai, Y. Gavrilov, R. Kapon, Y. Levy, and Z. Reich, Proc. Natl. Acad. Sci. U. S. A. **110**(26), 10628–10633 (2013).