Resolution exchange with tunneling for enhanced sampling of protein landscapes

Fatih Yasar ,* Alan J. Ray,† and Ulrich H. E. Hansmann ‡
Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, USA

(Received 24 October 2021; revised 22 May 2022; accepted 15 June 2022; published 5 July 2022)

Simulations of protein folding and protein association happen on timescales that are orders of magnitude larger than what can typically be covered in all-atom molecular dynamics simulations. Use of low-resolution models alleviates this problem but may reduce the accuracy of the simulations. We introduce a replica-exchange-based multiscale sampling technique that combines the faster sampling in coarse-grained simulations with the potentially higher accuracy of all-atom simulations. After testing the efficiency of our Resolution Exchange with Tunneling (ResET) in simulations of the Trp-cage protein, an often used model to evaluate sampling techniques in protein simulations, we use our approach to compare the landscape of wild-type and A2T mutant Aβ_{1-42} peptides. Our results suggest a mechanism by that the mutation of a small hydrophobic alanine (A) into a bulky polar threonine (T) may interfere with the self-assembly of Aβ fibrils.

DOI: 10.1103/PhysRevE.106.015302

I. INTRODUCTION

While molecular dynamics is now commonly used to study folding, association, and aggregation of proteins and other biological macromolecules [1–9], biochemical processes such as the formation of amyloid fibers from monomers [5,9,10] often occur on timescales [10,11] that exceeds what can be covered in all-atom simulations. Coarse graining, i.e., lowering the resolution of a system [4,12–16], allows one to reduce the computational difficulties and to access timescales not obtainable to the fine-grained all-atom models [4,12], but it often results in lower accuracy. This is because the smaller number of degrees of freedom lowers the entropy of the system, and it is difficult to compensate for this reduction by modifying the enthalpic contributions accordingly [12]. Multiscale techniques try to combine the advantages of fine-grained models (that are more accurate but costly to evaluate) with that of coarse-grained models (which are less detailed but enable larger time steps).

One example is Resolution Exchange [17] where the replica-exchange protocol [18] is used to induce a walk in resolution space. In the same way that for Replica Exchange Molecular Dynamics (REMD) [18,19] the walk in temperature space leads to faster sampling at low temperatures, does exploration of resolution space enable a faster convergence of simulations at an all-atom level [17,20]. However, the replica-exchange step requires reconstruction of the fine-grained degrees of freedom of a previously coarse-grained configuration, for instance, by adding side chains to a conformation that was previously described only by the backbone.

Various approaches [17,20–23] have been developed to address this problem, but often they result in high energies of the proposal configuration (and therefore low acceptance rates) [20,21], or introduce biases [22,23]. This dilemma can be alleviated by introducing a potential energy made of three terms:

$$E_{\text{pot}} = E_{FG} + E_{CG} + \lambda E_{\lambda}. \tag{1}$$

The first term is the energy E_{FG} of the protein system and the surrounding environment as described by an *all-atom* (*fine-grained*) model. The second term E_{CG} describes the same system by a suitable *coarse-grained* model. Both models are coupled by a system-specific penalty term E_{λ} [24,25] that measures the similarity between the configurations at both levels of resolution, with the strength of coupling controlled by a replica-specific parameter λ. Hence, Hamilton Replica Exchange [26,27] of the above defined multiscale system leads to an exchange of information between fine-grained and coarse-grained models, with measurements taken at the replica where $\lambda = 0$. However, while avoiding the problem of steric clashes in resolution exchange, the exchange probability is often still small [28], and the resulting need for multiple replicas to bridge the two levels of resolution makes this approach not appealing.

As an alternative, we propose here a Resolution Exchange with Tunneling (ResET) approach as a way to "tunnel" through the unfavorable proposal configuration that cause the low acceptance rates in the above described Hamilton Replica Exchange variant. Working and efficiency of our approach, which requires only two replicas and replaces exchanges by replacement moves, is tested in simulations of the Trp-cage [29,30] miniprotein [Protein Data Bank (PDB) Identifier 1L2Y], an often used model for testing new sampling techniques. As a first application, we use in the second part ResET to compare the landscape of Aβ_{1-42} wild-type peptides, implicated in Alzheimer's disease, with that of A2T mutants which seems to protect against the disease [31–33]. Our results suggest a mechanism by that the mutation of a small hydrophobic alanine (A) into a bulky polar threonine (T) may interfere

*Present Address: Department of Physics Engineering, Hacettepe University, Ankara 06800, Turkey; fyasar@ou.edu
†alanjray@ou.edu
‡uhansmann@ou.edu

 ©2022 American Physical Society

with the self-assembly of Aβ fibrils, decreasing the chance for formation of the disease-associated Aβ amyloids [34–36].

II. RESOLUTION EXCHANGE WITH TUNNELING

Resolution Exchange with Tunneling (ResET) utilizes two replicas, each containing both a coarse-grained and a fine-grained representation of the system. On each replica, both representations evolve separately by molecular dynamics. On the first replica, A, is the *coarse-grained model* in a configuration A_{CG} and has a potential energy $E_{CG}^{\text{pot}}(A_{CG})$ and a kinetic energy $E_{CG}^{\text{kin}}(A_{CG})$. On the other hand, the fine-grained model is in a configuration A_{FG} that has a kinetic energy $E_{FG}^{\text{kin}}(A_{FG})$ and a total potential energy $E_{FG}^{\text{biased}}(A_{FG})$ which adds to the intraprotein and protein-water interaction energy $E_{FG}^{\text{pot}}(A_{FG})$ a biasing term that depends on the configuration A_{CG} of the coarse-grained model by $E_{FG}^{\text{biased}}(A_{FG}) = E_{FG}^{\text{pot}}(A_{FG}) + \lambda_1 E_\lambda(A_{FG}, A_{CG})$. Hence, the two models on this replica interact only by the term $\lambda_1 E_\lambda(A_{FG}, A_{CG})$ that biases the fine-grained model, but are otherwise invisible to each other. The effect of this biasing term is that configurations of the fine-grained model are favored which resemble the coarse-grained model configuration, with the strength of the bias controlled by parameter λ_1.

The opposite situation is found on the replica B. Here lives an independent fine-grained model with configuration B_{FG} that has a potential energy $E_{FG}^{\text{pot}}(B_{FG})$ and kinetic energy $E_{FG}^{\text{kin}}(B_{FG})$, while, on the other hand, the configuration B_{CG} of the coarse-grained model has a kinetic energy $E_{CG}^{\text{kin}}(B_{CG})$ and a total (biased) potential energy $E_{CG}^{\text{biased}}(B_{CG}) = E_{CG}^{\text{pot}}(B_{CG}) + \lambda_2 E_\lambda(B_{CG}, B_{FG})$ that depends on the fine-grained model by a term $\lambda_2 E_\lambda(B_{CG}, B_{FG})$. This biasing term now ensures that on replica B the coarse-grained configuration resembles the one of the fine-grained model.

Statistical properties are calculated only from data that are sampled from the unbiased models. Hence, for calculating averages of fine-grained properties we use solely data recorded for the (unbiased) fine-grained model on replica B, while averages of coarse-grained quantities are calculated from the data recorded for the unbiased coarse-grained model on replica A.

While the time step for integrating fine-grained and coarse-grained models may differ, they have to be the same for the corresponding models on both replicas. This is because after a certain number of molecular dynamics steps a decision is made on whether to replace on replica B the configuration B_{FG} in the unbiased fine-grained model by the configuration A_{FG} of the auxiliary (biased) fine-grained model living on replica A. This replacement goes together with a rescaling of the velocities $v_{FG}(A_{FG})$ such that after rescaling $\hat{E}_{FG}^{\text{kin}}(A_{FG}) = E_{FG}^{\text{kin}}(B_{FG})$. Note, however, that ours is *not* an exchange move and configuration A_{FG} is *not* replaced by B_{FG} with rescaled velocities $v_{FG}(B_{FG})$, and neither the kinetic energy nor the weight of the biased fine-grained model of the replica A changes. On the other hand, while the old configuration of the unbiased fine-grained model on replica B has a weight

$$p_{\text{old}} = \exp\left[-\beta\left(E_{FG}^{\text{pot}}(B_{FG}) + E_{FG}^{\text{kin}}(B_{FG})\right)\right], \tag{2}$$

the new configuration will have, if the replacement move is accepted, a weight

$$p_{\text{new}} = \exp\left[-\beta\left(E_{FG}^{\text{pot}}(A_{FG}) + E_{FG}^{\text{kin}}(B_{FG})\right)\right]. \tag{3}$$

However, the proposal configuration A_{FG} is derived from a distribution skewed by the interaction with the coarse-grained model on replica A, and the probability to get A_{FG} as proposal configuration is given by

$$q(A_{FG}) = \exp\left[-\beta\left(\lambda_1 E_\lambda(A_{FG}, A_{CG}) + \Delta E_{FG}^{\text{kin}}\right)\right], \tag{4}$$

where $\Delta E_{FG}^{\text{kin}} = E_{FG}^{\text{kin}}(A_{FG}) - E_{FG}^{\text{kin}}(B_{FG})$ accounts for the change in kinetic energy resulting from the rescaling of velocities in the proposal configuration. The replacement of configuration B_{FG} by A_{FG} is accepted with a probability $w(B_{FG} \to A_{FG})$ that can be calculated from

$$p_{\text{old}} \times q(A_{FG}) \times w(B_{FG} \to A_{FG}) = p_{\text{new}}, \tag{5}$$

leading to the Metropolis-Hastings acceptance criterion:

$$w(B_{FG} \to A_{FG}) = \min\left(1, \exp\left[-\beta\left(E_{FG}^{\text{pot}}(A_{FG}) - E_{FG}^{\text{pot}}(B_{FG}) - \lambda_1 E_\lambda(A_{FG}, A_{CG}) - \Delta E_{FG}^{\text{kin}}\right]\right). \tag{6}$$

Note that detailed balance

$$p_{\text{old}} \times q(A_{FG}) \times w(B_{FG} \to A_{FG}) = p_{\text{new}} \times q(B_{FG}) \times w(A_{FG} \to B_{FG}) \tag{7}$$

is preserved as

$$q(B_{FG}) = \exp\left[-\beta\left(\lambda_1 E_\lambda(B_{FG}, A_{CG}) - \Delta E_{FG}^{\text{kin}}\right)\right] \tag{8}$$

and

$$w(A_{FG} \to B_{FG}) = \min\left(1, \exp\left[-\beta\left(E_{FG}^{\text{pot}}(B_{FG}) - E_{FG}^{\text{pot}}(A_{FG}) - \lambda_1 E_\lambda(B_{FG}, A_{CG}) + \Delta E_{FG}^{\text{kin}}\right)\right]\right). \tag{9}$$

Similarly, at other times, the the coarse-grained configuration A_{CG} on replica A is replaced by the configuration B_{CG} of the biased coarse-grained model on replica B with probability:

$$w(A_{CG} \to B_{CG}) = \min\left(1, \exp\left[-\beta\left(E_{CG}^{\text{pot}}(B_{CG}) - E_{CG}^{\text{pot}}(A_{CG}) - \lambda_2 E_\lambda(B_{FG}, B_{CG}) - \Delta E_{CG}^{\text{kin}}\right)\right]\right) \tag{10}$$

with $\Delta E_{CG}^{\text{kin}} = E_{CG}^{\text{kin}}(B_{CG}) - E_{CG}^{\text{kin}}(A_{CG})$. Reweighting the velocities of configuration B_{CG} such that $\hat{E}_{CG}^{\text{kin}}(B_{CG}) = E_{CG}^{\text{kin}}(A_{CG})$ and the Metropolis-Hastings acceptance criterion are again used to correct for the skewed probability by which the configuration B_{CG} is proposed.

Note that the replacement moves for the (unbiased) fine-grained model on replica B and the (unbiased) coarse-grained model on replica A are not coupled. An update of the unbiased coarse-grained configuration on replica A also changes the E_λ biasing term in the ancillary fine-grained configuration on replica A, as does the update of the unbiased fine-grained configuration on replica B changes the corresponding biasing term in the steered coarse-grained configuration on replica B. Hence, in order to minimize this disturbance, we also rescale the velocities in the biased models such that the change in kinetic energy compensates for the change in E_λ.

The overall effect of the replacement moves is a more exhaustive sampling of the (unbiased) fine-grained configurations on replica B. This is because Eq. (6) allows for the

TABLE I. Alternative implementation of ResET using four replicas.

Model	Replica	Potential energy	Kinetic energy	Lambda	Lambda energy
Unbiased fine-grained model	0	P_0	K_0		
Biased fine-grained model	1	P_1	K_1	λ_1	$E_\lambda(1, 3)$
Biased coarse-grained model	2	P_2	K_2	λ_2	$E_\lambda(0, 2)$
Unbiased coarse-grained model	3	P_3	K_3		

occasional acceptance of configurations taken from replica A that, being biased by the coarse-grained model on this replica, evolve according to the timescale of the coarse-grained model, i.e., faster than usual in a fine-grained model. As the fine-grained configuration on the unbiased model is replaced by another fine-grained model, the problem of reconstructing fine-grained degrees of freedom (the main limitation of resolution exchange) does not appear. Similarly, the corresponding replacement move for the unbiased coarse-grained model on replica A effectively increases the accuracy of the coarse-grained simulation as it represents an influx of information from the fine-grained degrees of freedom.

We remark that in software packages such as GROMACS [37] it is sometimes simpler to separate the biased and unbiased models onto different replicas. In this case one would have four replicas, with a possible distribution of the models sketched in Table I.

In this implementation, the replica 0 and 2, and replica 1 and 3, communicated during the molecular dynamics evolution of the configurations; and the ResET move replaces the configuration of replica 0 by that of replica 1, and/or the configuration on replica 3 by that of replica 2.

III. MATERIALS AND METhODS

Setup of the ResET simulations

Our simulations utilize a modified version of the GROMACS [37] molecular package available from the authors. Initial tests of the working and efficiency are for the Trp-cage protein [29,30], an often used system for evaluating new algorithms. In order to compare our simulations with previous studies, we follow closely the setup of Han *et al.* [38] for the coarse-grained model, and that of Kouza *et al.* [39] for the fine-grained model. Hence, our coarse-grained Trp-cage protein model is described by PACE force field [4], with the uncapped protein solvated by 1118 MARTINI [40] coarse-grained water molecules, and buffered 0.15M Na^+ and Cl^- ions, in a cubic box of length 5.18 nm, leading to a total of 1313 coarse-grained particles. On the other hand, in our fine grained model is the N-terminus capped by an acetyl group and at C-terminus by methylamine, leading to a total number of 313 atoms for the protein that are solvated with 2645 extended simple point charge (SPC/E [41]) water molecules in a cubic box with an edge length of 4.4 nm. One chlorine ion (Cl^-) is added to neutralize the system. In total, the system contains 8249 fine-grained particles, with the interactions between them described by the AMBER94 force field [42].

As a first application, we compare in the second part of this study the ensemble of configurations sampled by ResET simulation of $A\beta_{1-42}$ wild-type and A2T mutant peptides. While aggregates of the wild-type $A\beta$-peptides are implicated in Alzheimer's disease, the A2T mutant appears to be protective, i.e., reducing the probability for acquiring the disease. We use in our simulations for both wild-type and mutant the MARTINI force field [40]. This is because this coarse-grained model is computationally efficient and has been already used earlier in $A\beta$ simulations [43,44]. The main chain of each amino acid is represented by one bead, and the side chains by up to four beads depending on the size of the amino acid. Our wild-type protein thus contains 91 beads, and the mutant 92 beads. Each peptide is placed in a cubic box and solvated with the MARTINI-CG water molecules represented by single beads. Together with 3 Na^+ MARTINI-ion beads and a box size of 7.16 nm (wild-type) and 7.24 nm A2T mutant) we arrive at 2925 and 3189 particles, respectively. On the other hand, the fine-grained representations of wild-type and mutant peptides are modeled by the CHARMM36 force field [45] which we found in previous work to be efficient for simulations of intrinsically disordered and amyloid-forming proteins. The N- and C-termini are capped with acetyl and methyl groups, respectively. The protein is placed in the center of a cubic box using a 1 nm distance between the atoms of the protein and box. The each system is solvated with TIP3 water molecules [46] and neutralized with 3 Na^+ ions. This leads to a box size of 7.5 nm and a total number of 41412 particles for the wild-type. Correspondingly, we get a box size of 7.6 nm and a total number of 44509 particles for the mutant.

In simulations of both the Trp-cage protein and the $A\beta$ peptides we use for both fine-grained and coarse-grained models shift functions with a cutoff of 1.2 nm in the calculations of Coulomb and van der Waals interactions. Because of periodic boundary conditions we employ Particle Mesh Ewald (PME) [47] summation to account for long-range electrostatic interactions. Hydrogen atoms and bond distances are constraint in the fine-grained model by the LINCS algorithm [48]. Equations of motion are integrated using a leap-frog algorithm, with a time step of 2 fs for both the fine-grained model and coarse-grained model. The v-rescale thermostat [49] with a coupling time of 0.01 ps is used to maintain the temperature in the coarse-grained models, while a Nose-Hoover [50,51] thermostat with the coupling time of 0.5 ps controls the temperature in the fine-grained models.

A key element of the ResET sampling technique is the restraining potential E_λ which quantifies the similarity between fine-grained and coarse-grained configurations, described by the coordinates q_{FG} of the atoms in the fine-grained model and coordinates q_{CG} for the ones in the coarse-grained model. Choice of this function depends on the system under consideration, and by specifying how the configurations are biased one may affect the sampling efficiency of the approach. Therefore, this function has to be selected carefully. In our case,

TABLE II. Simulation details.

	Trp cage			$A\beta_{1-42}$		
Method	Force field	Sampling no	time (ns)	Force field	Sampling No	time (ns)
Canonical FG	AMBER94	3	5000	CHARMM36	—	—
REMD FG	AMBER94	1	200	—	—	—
ResET FG+CG	AMBER94+PACE	6	200(1000)	CHARMM36+MARTINI v2.2	2	100(500)

we choose a function, introduced in Ref. [25] that is often used to compare configurations in simulations of proteins. For this purpose, a protein configuration is described by the set of distances $\delta(ij)$ between all possible pairs of C_α atoms i and j in a protein. Hence, the differences $\Delta(ij) = \delta_{FG}(ij) - \delta_{CG}(ij)$ can be used to quantify the dissimilarity between fine-grained and coarse-grained configurations. However, in order to ensure well-behaved forces in the molecular dynamic simulations are the $\Delta(ij)$ scaled for the calculation of the restraining potential:

$$E_\lambda(q_{FG}, q_{CG}) = \sum_{ij} K_{ij}(q_{FG}, q_{CG}), \tag{11}$$

where the sum goes over all possible pairs of C_α atoms i and j, and

$$K_{ij}(q_{FG}, q_{CG}) = \begin{cases} \frac{1}{2}[\Delta^2(i,j)] & -ds < \Delta(i,j) < ds \\ A + \frac{B}{\Delta^S(i,j)} + f_{\max}\Delta(i,j) & \Delta(i,j) > ds \\ A + \frac{B}{\Delta^S(i,j)}(-1)^S - f_{\max}\Delta(i,j) & \Delta(i,j) < -ds \end{cases} \tag{12}$$

In this expression, ds is a distance threshold at which $K_{ij}(q_{FG}, q_{CG})$ switches to a soft asymptote approaching a maximum force set by the control parameter $f_{\max}$ as $\Delta(i,j) \to \infty$, and S determines how fast this maximum value is approached. In our case, $S = 1$, $f_{\max} = 0$, and $ds = 3$ Å. The parameters A and B are included to ensure continuity of $K_{ij}(q_{FG}, q_{CG})$ and its first derivative at values where $\Delta(i,j) = \pm ds$, i.e., where the functional form of Eq. (12) changes. These parameters are given by

$$A = \left(\frac{1}{2} + \frac{1}{S}\right) ds^2 - \left(\frac{1}{S} + 1\right) f_{\max} ds$$

$$\text{and} \quad B = \left(\frac{f_{\max} - ds}{S}\right) ds^{S+1}. \tag{13}$$

In the ResET simulations is the biased fine-grained model on replica A coupled to the unbiased coarse-grained model by a parameter $\lambda_1 = 0.5$, while on replica B the biased coarse-grained models is coupled to the free fine-grained models by a parameter $\lambda_2 = 2.5$. These values were chosen to ensure that the replacement moves [Eq. (6) and Eq. (10)] are not dominated by the λE_λ term. The ResET replacement move is tried every 250 ps, with the bias-correction factor $\lambda_1 E_\lambda(A_{FG}, A_{CG}) - \Delta E^{\text{kin}}_{FG}$ limited to the interval (0,100), and $\lambda_2 E_\lambda(B_{FG}, B_{CG}) - \Delta E^{\text{kin}}_{CG}$ to the interval (0,20), choices that in preliminary test runs led to increased numerical stability.

Start structures for both fine-grained and coarse-grained models are generated by heating up the experimental structures of PDB-ID: 1L2Y (Trp-cage) and PDB-ID: 1Z0Q ($A\beta_{1-42}$) [34,52] to 500 or 1000 K in short molecular dynamics simulations under NVT conditions (0.5 ns and 1 ns), and cooling them down to the respective temperatures (with the exception of the REMD simulations is this 310 K). Simulations of the various systems start from the so-generated configurations and are performed in the NVT ensemble, with the simulation details listed in Table II.

For most of our analysis we use GROMACS tools [37] such as gmx-rms, which calculates the root-mean-square deviation (RMSD) and the root-mean-square fluctuations (RMSFs) of residues with respect to an initial configuration. For visualization we use the VMD software [53], which we also use to calculate the solvent accessible surface area (SASA) using a probe radius of 1.4 Å. Other quantities are calculated with in-house programs and defined in the manuscript. An example are dynamic cross-correlation maps which are calculated using the definition of [54,55]:

$$C(i,j) = \frac{\langle \Delta\mathbf{r}_i . \Delta\mathbf{r}_j \rangle}{\langle \Delta\mathbf{r}_i^2 \rangle \langle \Delta\mathbf{r}_j^2 \rangle}, \tag{14}$$

where $\Delta\mathbf{r}_i$ and $\Delta\mathbf{r}_j$ are the displacement vectors of ith and jth residues of the system and angle brackets represent ensemble averages. Positive values mark correlated motions of the respective residues while negative values indicate anticorrelated motion.

IV. RESULTS AND DISCUSSION

A. Efficiency of ResET

In order to test the working and efficiency of our multiscale approach ResET, we perform first simulations of the Trp-cage [29,30] miniprotein, an often used model for testing sampling techniques. Choice of this system, with which we are familiar from previous work, therefore allows a direct comparison with past simulations. An example are the Replica Exchange Molecular Dynamics (REMD) simulations of Ref. [39,56], where 40 replicas of equal volume are simulated at 40 temperatures spanning a range from $T = 280$ K to $T = 540$ K. Configurations are exchanged between neighboring tempera-

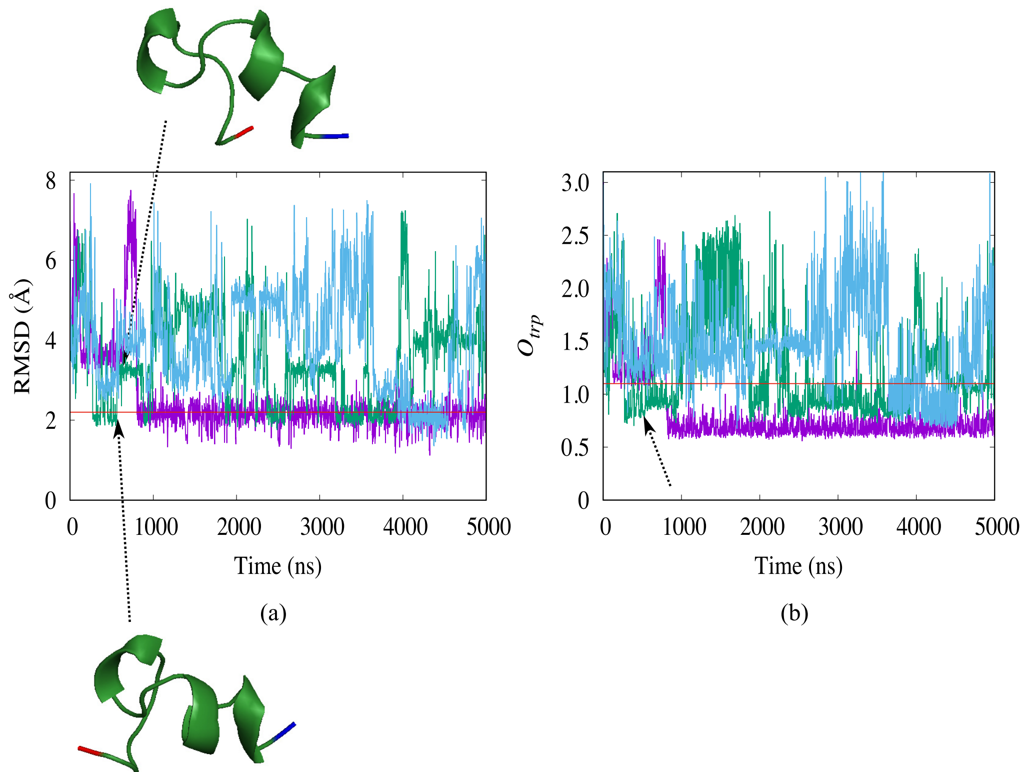

FIG. 1. The time evolution of RMSD (a) and folding parameter O_{trp} (b) as measured in regular molecular dynamics simulations at $T = 310$ K. Trajectory 1 is drawn in purple, trajectory 2 in green and trajectory 3 in blue. A red line marks the values of the two quantities below which we consider the configurations as native-like according to the criteria discussed in the text. The two snapshots are taken from trajectory 2 at 601.0 ns (snapshot at the bottom) and 602.3 ns (snapshot on top). Both snapshots show similar configurations while the RMSD changes from 2.0 Å to 3.6 Å. On the other hand, no such change is seen for folding parameter O_{trp}; see the corresponding region in (b), marked by the arrow. N- and C-terminal residues in the snapshots are marked in blue and red, respectively.

tures according to a generalized Metropolis criterion, leading to a random walk in temperature that allows replicas to find local minima (when at low temperatures) and escape out of them (when at high temperatures). The net effect is an enhanced sampling at the target temperature. Defining a configuration as native-like if the RMSD to the PDB structure (PDB-ID: 1L2Y) is less than 2.5 Å, we find at $T = 310$ K native-like configuration with a frequency of 87%, using the more restrictive criterion of a RMSD smaller than 2.2 Å, the frequency reduces to 55%. Note that these frequencies do not change beyond statistical fluctuations once the REMD simulation has reached 50 ns, and we therefore neglect the first 50 ns of our 200 ns long trajectories when calculating the frequencies. While these frequencies are similar to the ones observed in earlier work [39,56], we suspect that our values overestimate the frequency of folded configurations that reside at a certain time at $T = 310$ K. This is because the systems are simulated at each temperature with the same volume. This volume, while sufficiently large at the target temperature may at the higher temperature suppress extended configurations, therefore artificially stabilizing folded configurations. For this reason, we prefer instead to compare our ResET simulations directly with regular constant temperature molecular dynamics, simulating the Trp-cage protein in three independent trajectories at $T = 310$ K over 5000 ns, a value that is comparable to the experimental measured folding times of around 4 μs [57]. The RMSD as function of time is shown for all three trajectories in Fig. 1(a). Visual inspection of the three trajectories points to another problem. For a small protein such as Trp-cage is the RMSD not good measure for similarity as configurations that appear as similar by visual inspection may differ by relatively large RMSD values. This can be seen, for instance, in the second trajectory where at around 600 ns the RMSD increases from 2.0 Å to 3.6 Å, i.e., from native-like to configurations to one considered no longer native-like according to the above definition of a native configuration (i.e., having a RMSD of less than 2.5 Å). However, visual inspection shows that the molecule keeps its native-like fold; see the corresponding configurations in the Fig. 1(a). This contradiction between our RMSD-based definition and visual inspection made us consider another quantity as a measure for similarity. The two main characteristics of the Trp-cage native structure are its two helices (residues 2–9 and 11–14), and the contact between residue 6W (a tryptophan) and residue 18P (a proline). Hence we define as marker for Trp-cage folding a new quantity:

$$O_{\text{trp}} = d_{6-18} + 1/(n_H + 1). \tag{15}$$

Here d_{6-18} is the difference between residues 6W and 18P, and n_H the number of residues that have dihedral angles as seen in a helix. The time evolution of this quantity in Fig. 1(b) shows that the new coordinate allows indeed a better description of the folding transitions, as its behavior differs less from the visual inspection. Especially, we do not see for the second trajectory at 600 ns the false signal for nonnative configurations

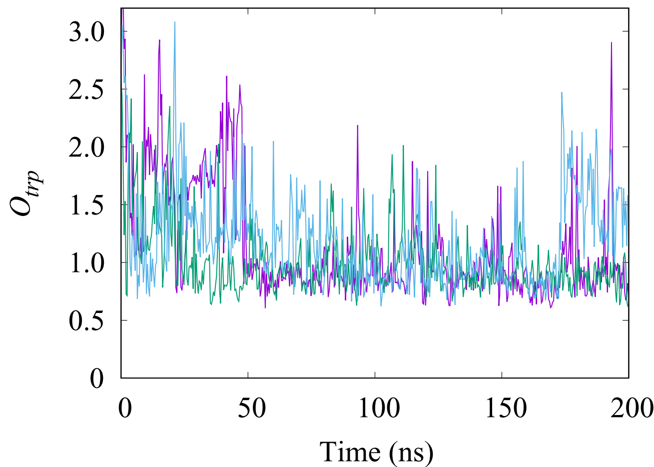

FIG. 2. The time evolution of the order parameter in our ResET simulations of the Trp-cage protein. Trajectory 1 is drawn in purple, trajectory 2 in green, and trajectory 3 in blue.

that we see in the RMSD plot. Comparing O_{trp} as function of time with visual inspection of configurations along the trajectories suggests that folded configurations are characterized by values of $O_{\text{trp}} < 1$, and we use in the following this definition to quantify frequencies of folded configurations.

With this definition, we observe the first folding event at $t = 11.6$ ns (in trajectory 2), and the systems stays folded for about 600 ns before unfolding again. For trajectory 1 folding is observed at $t = 800$ ns, and no folding is observed within 3500 ns in the third trajectory where the protein unfolds afterwards again at about 4500 ns. As a consequence, we find between 250 ns and 500 ns folded configurations with a frequency of about 26% and between 750 ns and 1000 ns, with about 49%. The frequencies increase only slowly as the simulations proceed, and between 3000 ns and 5000 ns we find native-like configurations with about 58%. The above numbers are consistent with the experimentally measured folding times of about 4 μs [57].

How does our multiscale method fit in this discussion? The time evolution of our marker function $O_{\text{trp}}(t)$ is shown in Fig. 2. Native-like configurations according to our criterion are observed after around 30 ns, and between 50 ns and 100 ns seen with a frequency of about 59%. The frequencies do not change much as the simulation progresses, and between 150 ns and 200 ns are native-like configurations observed with 65%. We remark that these frequencies do not depend on the choice of parameters with which we scale the λ energy contribution in the ResET update.

These frequencies for folded configurations are similar to what is seen in long-time canonical runs, but require shorter simulation times. Hence, our simulations of the Trp-cage protein indicate that our multiscale simulation method leads indeed to an increase in sampling efficiency. If we take as a criterion for the comparison the time it takes to have (on average) about 50% of configuration folded (about 800 ns for the canonical runs and 50 ns for the ResET run) we find that ResET is about 16 times faster than the canonical simulations. While the gain in efficiency will depend on the specifics of the coarse-grained model (i.e., how much faster it samples the configuration space) and its coupling to the physical force field, our data demonstrate the faster sampling properties of our multiscale approach.

B. Comparing Aβ wild-type and A2T mutant

Our evaluation of the sampling efficiency of ResET relies on a rather simple test case. As a more interesting first application, we use in the second part our sampling technique to compare the ensembles of wild-type and A2T mutant Aβ_{1-42} peptides. Fibrils containing Aβ_{1-40} or the more toxic Aβ_{1-42} are a hallmark of Alzheimer's disease and the focus of intense research [31]. A large number of familial mutations are known that worsen the symptoms of Alzheimer's disease or hasten its outbreak [32,33], but there have been also mutations identified that are protective, i.e., lower the risk to fall ill with Alzheimer's disease. One example is the mutant A2T where the second residue (counted from the N-terminus) is changed from a small hydrophobic alanine (A) into a bulky polar threonine (T) [34]. It has been not yet established why this mutation is protective [35,36], but one possibility is that this mutation alters the pathway for amyloid formation, for instance, by making it more difficult to form aggregates. In order to test this hypothesis we simulate here Aβ_{1-42} wild-type and A2T mutant monomers, and compare the ensembles of sampled configurations for their aggregation propensities.

Under physiological conditions are Aβ-peptides intrinsically disordered, and we do not expect the appearance of folded structures. Instead, we assume that the ensemble of configurations contains such with transiently formed β-strands that would encourage aggregation. We conjecture that such transient ordering appears more often for wild-type Aβ_{1-42} than for the A2T mutant peptides. In order to identify these differences in local ordering, we have measured the RMSFs of residues for both cases, taking as reference structure the corresponding start configuration, but discarding for the calculation of the RMSF the first 50 ns of the simulation. The RMSF is chosen because this quantity describes the flexibility of residues or segments of the protein, and the more flexible a segment is, the less likely will it form stable structures. Our data are shown in Fig. 3. While for the first 20 residues there are only small differences between wild-type and mutant, the situation is different for the C-terminal half of the chain. For residues 21–37, the RMSF is considerably lower for the mutant then for the wild-type. We remark that this picture does not change if we recalculate the RMSF, including now all heavy atoms (i.e, not only backbone but also side-chain atoms).

The lower flexibility of the segment 21–37 in the mutant is not correlated with increased secondary structure. Residues take dihedral angle values as seen in a helix or a β strand with about 10%, independent on whether wild-type and mutant. However, there is a change in the average radius of gyration (RGY, a measure for the volume), which with 10.6(1) Å is larger for the mutant than for the wild-type where it is 10.5(1) Å. Similarly, the average solvent accessible surface area (s) of the peptide is in the mutant with 38.0(1) nm^2 less fluctuating than in the wild-type (38.0(3) nm^2), reflecting the gain in surface area resulting from the more bulky

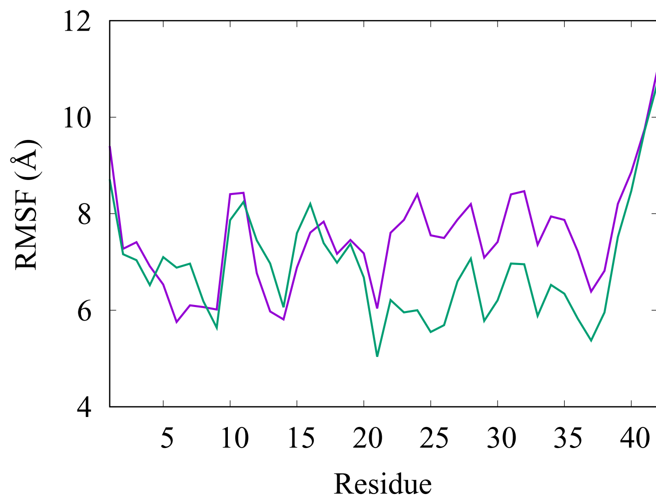

FIG. 3. Root-mean-square fluctuations of residues in either wild-type (purple) or A2T mutant (green) $A\beta_{1-42}$ peptides. Only heavy atoms are considered in the calculation, and the first 50 ns of the 100 ns trajectories discarded to allow for convergence of the simulations.

threonine. However, the relation is different for the segment of residues 21–37, where the wild-type has a SASA value of 18.2(3) nm^2 and the mutant a SASA of 18.1(2) nm^2. The differences for the segment result from polar residues as the solvent accessible surface area of hydrophobic residues is with 4.1(1) nm^2 the same for both mutant and wild-type. Hence, the differences in SASA values for this segment indicate that in the mutant polar residues, which are exposed to solvent in the wild-type, form contacts with other residues. In order to understand the differences between mutant and wild-type in more detail, we have also analyzed the contacts and cross-correlations between residues, focusing again on the final 50 ns of the trajectories for both systems. The resulting maps for both systems are shown in Figs. 4(a) and 4(b), with the coloring describing the degree of correlation between residues.

Unlike in the wild-type are in the A2T mutant the disordered N-terminus (residues 1–9) and residues 27–33 correlated. This correlation results from electrostatic interactions, for instance, between the NH3+ group of residue K28 (a lysin) with negatively charged COO- group of residue E7 (a glutamic acid) seen in the snapshot shown in Fig. 4(d). Hence, the replacement of the small hydrophobic alanine by a bulky polar threonine allows for the above electrostatic interactionsin the mutant that do not exist in the wild-type, and whose importance for inhibiting amyloid formation in the A2T mutant has been already noticed earlier in Ref. [58]. These interactions likely stabilize not only the segment 27–33, but are responsible for the lower RMSF seen for residues 21–37. The interactions between N-terminus and residues 27–33 compete now in the A2T with hydrophobic interactions between the segment formed by residues 13–21, which include the central hydrophobic core ($L_{17}VFFA_{21}$), and the mostly

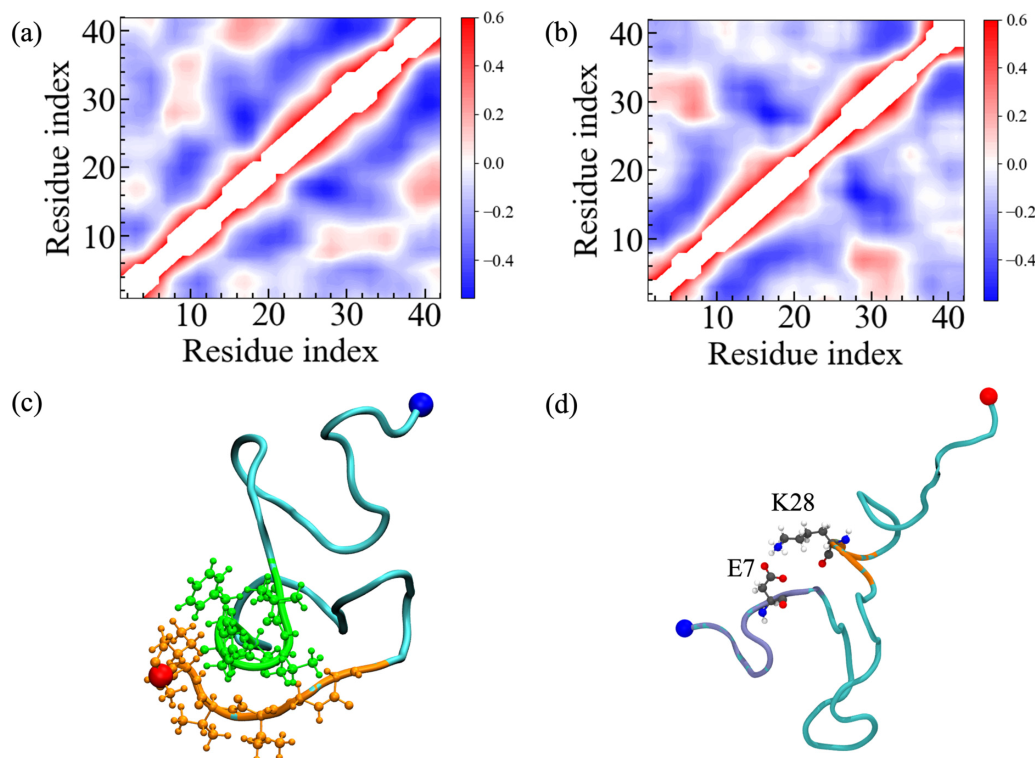

FIG. 4. Two-dimensional dynamic cross-correlation map extracted from (a) wild-type and (b) mutant $A\beta_{1-42}$ ResET simulations. A representative snapshot obtained from the wild-type simulations is shown in (c), where the central hydrophobic core $L_{17}VFFA_{21}$ and the C-terminal hydrophobic residues $G_{37}GVVIA_{42}$ are drawn in green and orange, respectively. A corresponding snapshot from the mutant simulation is shown in (d), where the disordered N-terminus (residues 1–10) and residues 27–31 are colored in ice-blue and orange, respectively. N- and C-terminal residues are represented by blue and red spheres.

hydrophobic C-terminus (residues 37–42), see the corresponding snapshot in Fig. 4(c). As a result the two segments are correlated in the wild-type but not in the mutants. These interactions between the peptide's two main hydrophobic domains are thought to be crucial for the self-assembly of Aβ fibrils [59,60], but are now missing in the A2T mutant, reducing the risk for aggregation.

V. CONCLUSIONS

We have described a replica-exchange-based multiscale simulation method, Resolution Exchange with Tunneling (ResET), that is designed for simulations of protein folding and aggregation. Our approach combines the faster sampling of coarse-grained simulations with the potentially higher accuracy of all-atom simulations. It avoids the problem of low acceptance rates plaguing similar approaches, and it requires only a few replicas. We first test for the small Trp-cage protein the accuracy and efficiency of our approach in comparison to long-scale (5 μs) regular molecular dynamic simulations. In a second step, we use our method to compare the ensemble of Aβ_{1-42} wild-type peptides, implicated in Alzheimer's disease, with that of A2T mutants which seems to protect against the disease. Our ResET simulations indicate that the replacement of a small alanine (A) by a bulky threonine (T) as residue 2 alters the pathway for amyloid formation. Introducing steric constraints on the mostly polar N-terminal residues it encourages electrostatic interactions with residues 27–33. In turn, these interactions reduce the flexibility of the extended segment 21–37, therefore contributing to the overall larger volume, more exposed surface and resulting higher solubility of the mutant. At the same time do this interactions also interfere with the hydrophobic interactions between the central hydrophobic core ($L_{17}VFFA_{21}$), and the mostly hydrophobic C-terminus (residues 37–42), known to be crucial for the self-assembly of Aβ fibrils, decreasing therefore the chance of formation of Aβ amyloids. The enlarged exposed hydrophobic surface area that in connection with a higher solubility may trigger faster degradation of the mutant, could further contributing to this mechanism, which may explain why the A2T mutant protects against Alzheimer's disease. We plan to test this hypothesis by comparing the A2T mutant with suitable double mutants that would interfere with this mechanism.

ACKNOWLEDGMENTS

The simulations in this work were done using the SCHOONER cluster of the University of Oklahoma, XSEDE resources allocated under Grant No. MCB160005 (National Science Foundation), and TACC resources allocated under grant under Grant No. MCB20016 (National Science Foundation). We acknowledge financial support from the National Institutes of Health under Grants No. GM120634 and No. GM120578. F.Y. acknowledges support by the Scientific and Technological Research Council of Turkey (TUBITAK) under the BIDEB programs. We are grateful to Dr. N. Bernhardt for contributions at an early stage of this project. F.Y. also thanks the Department of Chemistry and Biochemistry for kind hospitality during his sabbatical stay at the University of Oklahoma.

[1] K. A. Dill and J. L. MacCallum, Science **338**, 1042 (2012).
[2] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl, Annu. Rev. Biophys. **37**, 289 (2008).
[3] C. M. Dobson, Nature (London) **426**, 884 (2003).
[4] W. Han, C. K. Wan, and Y. D. Wu, J. Chem. Theory Comput. **6**, 3390 (2010).
[5] C. M. Chiti and F. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).
[6] J. C. Stroud, C. Liu, P. K. Teng, and D. Eisenberg, Proc. Natl. Acad. Sci. USA **109**, 7717 (2012).
[7] W. M. Berhanu, F. Yasar, and U. H. Hansmann, ACS Chem. Neurosci. **4**, 1488 (2013).
[8] T. M. Doran, A. J. Kamens, N. K. Byrnes, and B. L. Nilsson, Proteins **80**, 1053 (2012).
[9] M. Eisenberg and D. Jucker, Cell **148**, 1188 (2012).
[10] Y. S. Eisele, Brain Pathol. **23**, 333 (2013).
[11] K. Weise, D. Radovan, A. Gohlke, N. Opitz, and R. Winter, ChemBioChem **11**, 1280 (2010).
[12] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski, Chem. Rev. **116**, 7898 (2016).
[13] W. Han and Y.-D. Wu, J. Chem. Theory Comput. **3**, 2146 (2007).
[14] C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker, in *Numerical Computer Methods, Part D*, edited by L. Brand and M. L. Johnson, Methods in Enzymology, Vol. 383 (Academic Press, 2004),pp. 66–93.
[15] A. Kolinski, Acta. Biochim. Pol. **51**, 349 (2004).
[16] A. Liwo, M. Baranowski, C. Czaplewski, E. Golas, Y. He, D. Jagiela, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska *et al.*, J. Mol. Model. **20**, 2306 (2014).
[17] E. Lyman, F. M. Ytreberg, and D. M. Zuckerman, Phys. Rev. Lett. **96**, 028105 (2006).
[18] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[19] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[20] E. Lyman and D. M. Zuckerman, J. Chem. Theory Comput. **2**, 656 (2006).
[21] P. Liu, Q. Shi, E. Lyman, and G. A. Voth, J. Chem. Phys. **129**, 114103 (2008).
[22] P. Liu and G. A. Voth, J. Chem. Phys. **126**, 045106 (2007).
[23] R. Lwin and T. Z. Luo, J. Chem. Phys. **123**, 194904 (2005).
[24] K. Moritsugu, T. Terada, and A. Kidera, J. Chem. Phys. **133**, 224105 (2010).
[25] W. Zhang and J. Chen, J. Chem. Theory Comput. **10**, 918 (2014).
[26] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).
[27] W. Kwak and U. H. E. Hansmann, Phys. Rev. Lett. **95**, 138102 (2005).
[28] W. Nadler and U. H. E. Hansmann, Phys. Rev. E **76**, 065701(R) (2007).
[29] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, Nat. Struct. Biol. **9**, 425 (2002).

[30] C. Simmerling, B. Strockbine, and A. E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002).
[31] J. Hardy and D. Selkoe, Science **297**, 353 (2002).
[32] M. Morishima-Kawashima and Y. Ihara, J. Neurosci. Res. **70**, 392 (2002).
[33] D. J. Selkoe and M. B. Podlisny, Annu. Rev. Genomics Hum. Genet. **3**, 67 (2002).
[34] T. Jonsson, J. K. Atwal, S. Steinberg, J. Snaedal, P. V. Jonsson, S. Bjornsson, H. Stefansson, P. Sulem, D. Gudbjartsson, J. Malone *et al.*, Nature (London) **488**, 96 (2012).
[35] J. A. Maloney, T. Bainbridge, A. Gustafson, S. Zhang, R. Kyauk, P. Steiner, M. van der Brug, Y. Liu, J. A. Ernst, R. J. Watts, and J. K. Atwal, J. Biol. Chem. **289**, 30990 (2014).
[36] I. Benilova, R. Gallardo, A. A. Ungureanu, V. Castillo Cano, A. Snellinx, M. Ramakers, C. Bartic, F. Rousseau, J. Schymkowitz, and B. De Strooper, J. Biol. Chem. **289**, 30977 (2014).
[37] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[38] W. Han and K. Schulten, J. Phys. Chem. B **117**, 13367 (2013).
[39] M. Kouza and U. H. Hansmann, J. Chem. Phys. **134**, 044124 (2011).
[40] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).
[41] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[42] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[43] M. Zhang, R. Hu, H. Chen, X. Gong, F. Zhou, L. Zhang, and J. Zheng, J. Chem. Inf. Model. **55**, 1628 (2015).
[44] A. B. Poma, H. V. Guzman, M. S. Li, and P. E. Theodorakis, Beilstein J. Nanotechnol. **10**, 500 (2019).
[45] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. J. Mackerell, J. Chem. Theory Comput. **8**, 3257 (2012).
[46] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[47] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[48] B. Hess, J. Chem. Theory Comput. **4**, 116 (2008).
[49] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[50] S. Nosé, Mol. Phys. **52**, 255 (1984).
[51] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[52] S. Tomaselli, V. Esposito, P. Vangone, N. A. J. van Nuland, A. M. J. J. Bonvin, R. Guerrini, T. Tancredi, P. A. Temussi, and D. Picone, ChemBioChem **7**, 257 (2006).
[53] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[54] S. Swaminathan Jr., W. E. Harte, and D. L. Beveridge, J. Am. Chem. Soc. **113**, 2717 (1991).
[55] W. Wang, P. Khatua, and U. H. E. Hansmann, J. Phys. Chem. B **124**, 1009 (2020).
[56] D. Paschek, H. Nymeyer, and A. E. Garcia, J. Struct. Biol. **157**, 524 (2007).
[57] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen, J. Am. Chem. Soc. **124**, 12952 (2002).
[58] P. Das, B. Murray, and G. Belfort, Biophys. J. **108**, 738 (2015).
[59] S. Zhang, K. Iwata, M. Lachenmann, J. Peng, S. Li, E. Stimson, Y.-A. Lu, A. Felix, J. Maggio, and J. Lee, J. Struct. Biol. **130**, 130 (2000).
[60] S. L. Bernstein, T. Wyttenbach, A. Baumketner, J.-E. Shea, G. Bitan, D. B. Teplow, and M. T. Bowers, J. Am. Chem. Soc. **127**, 2075 (2005).