

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/317619755>

An adaptive bias – hybrid MD/kMC algorithm for protein folding and aggregation

Article in *Physical Chemistry Chemical Physics* · June 2017

DOI: 10.1039/C7CP03035E

CITATIONS

0

READS

47

2 authors:



Emanuel Peter

Universität Regensburg

25 PUBLICATIONS 140 CITATIONS

[SEE PROFILE](#)



Joan Shea

University of California, Santa Barbara

151 PUBLICATIONS 5,508 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Hydration of molecular ions [View project](#)



Phase diagram of peptides adsorbed to surfaces. [View project](#)

Cite this: DOI: 10.1039/xxxxxx

An adaptive bias - hybrid MD/kMC algorithm for protein folding and aggregation.

Emanuel K. Peter ^{*a}, Joan-Emma Shea ^{b‡}

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxx

www.rsc.org/journalname

In this paper, we present a novel hybrid Molecular Dynamics/kinetic Monte Carlo (MD/kMC) algorithm and apply it to protein folding and aggregation in explicit solvent. The new algorithm uses a dynamical definition of biases throughout the MD component of the simulation, normalized in relation to the unbiased forces. The algorithm guarantees sampling of the underlying ensemble in dependency of one average linear coupling factor $\langle \alpha \rangle_\tau$. We test the validity of the kinetics in simulations of Dialanine and compare dihedral transition kinetics with long-time MD-simulations. We find that for low $\langle \alpha \rangle_\tau$ values, kinetics are in good quantitative agreement. In folding simulations of TrpCage and TrpZip4 in explicit solvent, we also find good quantitative agreement with experimental results and prior MD/kMC simulations. Finally, we apply our algorithm to study growth of the Alzheimer Amyloid A β 16-22 fibril by monomer addition. We observe two possible binding modes, one at the extremity of the fibril (elongation) and one on the surface of the fibril (lateral growth), on timescales ranging from ns to 8 μ s.

1 Introduction

Molecular Dynamics (MD) and Monte Carlo (MC) simulations have emerged as invaluable tools to investigate protein conformations and dynamics. However, energy landscapes for biomolecular systems are riddled with metastable minima¹, such that these systems only rarely change their state during conventional MD or MC simulations. Enhanced sampling methods solve this problem by defining essential reaction pathways, collective variables in the case of Metadynamics² and the s-ensemble weighting³, and an energy bias in the well-tempered ensemble⁴ and transition path sampling⁵. In replica exchange MD (REMD), on the other hand, a reactive coordinate is defined in temperature space^{6,7}, while spatial degrees of freedom are used as reaction pathways in umbrella sampling⁸. In addition, accelerated molecular dynamics, in which a bias potential function is used, has been implemented for the simulation of biomolecules and in ab-initio MD^{9,10}. Another method of note is the paradynamics approach that accelerates sampling using a reference potential. This method has

been applied to free-energy calculations^{11–13} and to investigate protein free energy landscapes^{14,15}. Other techniques, based on renormalization approaches from explicit to implicit solvent models allow the efficient sampling of potentials of mean force of proteins¹⁶. One of the difficulties encountered in the enhanced sampling methodology is extracting the time on which observed processes take place, since the dynamics are affected by applied biases in these approaches. A number of developments have been made to address this issue^{17–21}, however, most algorithms still need the definition of reactive pathways or product states *a priori*. At the present, the only way to extract kinetics from MD simulations without requiring reactive pathways is through long unbiased MD simulations on specific computers such as ANTON²².

An alternate means to obtaining timescales and rates from simulations is to use Kinetic Monte Carlo (kMC). The kMC move contains information about rates and an inherent time-information by which the system propagates during each kMC step^{23,24}. We recently developed hybrid kinetic Monte Carlo-MD (MD/kMC) techniques which we successfully applied to study the folding of α -helical and β -stranded peptides in explicit solvent^{25,26} and to elucidate the signaling-dynamics of photoactivatable proteins^{27,28}. While our MD/kMC algorithms were suitable for the study of monomeric proteins, their application to the process of

^a Department of Pharmacy and Chemistry, Institute of Physical and Theoretical Chemistry, University of Regensburg, Germany

^b Department of Chemistry and Biochemistry, Department of Physics, University of California, Santa Barbara, USA, ^{*Email} : shea@chem.ucsb.edu

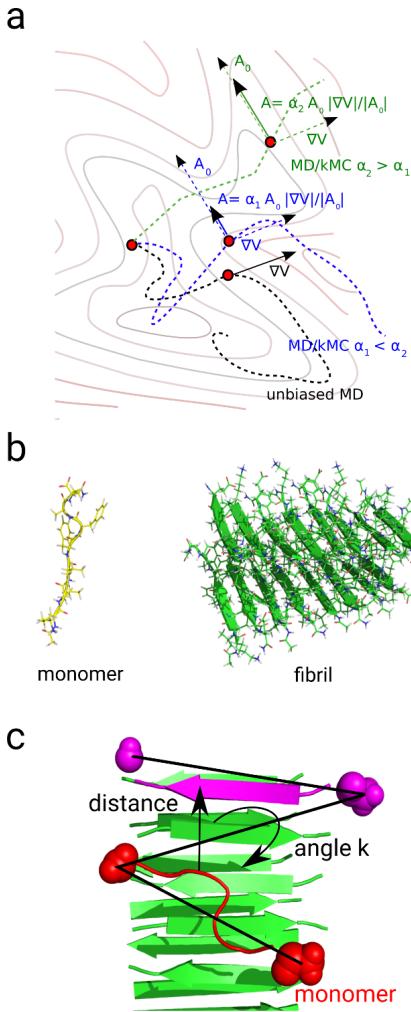


Fig. 1 (a) Schematic description of a time-dependent propagation of a system along a free energy landscape (FEL) using MD and the MD/kMC method as presented in this paper. The novel MD/kMC implementation is based on an adaptive renormalization of the applied bias A_0 in dependency of the un-biased gradient (∇V) and a linear coupling parameter $\langle \alpha \rangle_\tau$. For low $\langle \alpha \rangle_\tau$ values, the distribution converges to the underlying ensemble. (b) Starting configuration for aggregation simulations of one A β 16-22 monomer (sequence : ACE-KLVFFAE-NHE) to a double-stranded fibril of the same peptide. (c) Description of order parameters to describe A β 16-22 monomer aggregation near a fibril of the same peptide : The minimum distance to the tip of the fibril and the dihedral angle k between the termini of the free monomer and the first peptide of the fibril (N (N-NHE (monomer), C-ACE (monomer), N-NHE (fibril-first monomer), C-ACE (fibril-first monomer)).

protein aggregation is hampered by the larger number of MC-moves now required as a result of the increased number of degrees of freedoms associated with multiple proteins.

We present in this paper a novel hybrid MD/kMC technique that can be applied to both protein folding and aggregation problems, based on a dynamical redefinition of biases throughout the simulation. In contrast to earlier MD/kMC implementations, this algorithm is independent of any choice of moves, and collects gradients throughout the MD-simulation using a formalism derived from dissipative particle dynamics (DPD)²⁹⁻³¹. The bias-gradients depend on pair-interactions and on a pseudo-temperature represented by a random-noise. The MD/kMC method selects the bias according to rates and contains an inherent time-information about each process. Through an adaptive renormalization of the bias, the algorithm maintains ergodicity of the dynamics and allows sampling of ensemble averages. We probe the time-behavior and statistical correlation on the dynamics of Dialanine peptide in comparison with long time MD trajectories. Then, we test the parameter dependency in folding simulations of TrpCage and TrpZip4 in explicit solvent. Finally, we apply our new algorithm to simulate fibril growth of A β 16-22.

2 Methods

2.1 General approach

The general approach presented in this paper is based on the idea that kinetic Monte Carlo (kMC) moves only capture the dynamics of a particle-based all-atom protein on a coarse level. In other words, when applying moves such as dihedral rotations and longitudinal moves, i.e. H-bond formation and breaking, there is an inherent *a priori* assumption that these moves are the only rate-limiting events along the folding pathway^{25,26}. For example, it might be the case that the protein system could perform one simultaneous rotation and longitudinal motion (H-bond breaking via rotation of the backbone and longitudinal motion along the H-bond axis), which would be not well represented by 2 separate events : one rotational (backbone rotation) and one longitudinal move (H-bond breaking along the bond-axis). A sequence of 2 separate events would not resemble the transition within 1 event (rotation and longitudinal motion), which involves both in one simultaneous action.

In the case of protein aggregation, the situation is far more complex than for monomeric protein folding, since the number of degrees of freedom increases and it becomes crucial to account for the mutual orientation between the different units. Clearly, each kMC move would have to consider the motion of each monomeric unit itself, which has $3N-6$ spatial degrees of freedom for each monomer.

We introduce a novel approach based on distance-dependent fluctuations derived from dissipative particle dynamics (DPD)^{29,30}. The long-time dynamics depend on a long-

ranged hydrodynamic interaction, which conserves momentum and energy. In other words, this MD/kMC implementation does not impose *a priori* any dynamical pathway and is suitable to describe dynamics of peptide aggregation.

2.2 Methodology

The methodology is based on biased molecular dynamics, where the bias coordinate for each phase is selected during short phases (of length τ), while the system is sampled with MD. During each phase, a bias is applied to one particular segment and the algorithm collects structural data and gradients for each segment of the system, followed by a subsequent rate determination and execution during the next phase over a simulation time τ . In MD/kMC the probability for one segment to be selected for execution of the bias is proportional to its corresponding rate, which itself is a function of the energy content added to each segment. The applied energy fills the metastable basin in free energy and accelerates the escape of the system from local minima. If the applied energy is below a threshold value, the transition state remains unaffected, which is implemented via an adaptive renormalization of the bias-energy^{17,19}. This renormalization depends on one average coupling factor $\langle\alpha\rangle_\tau$ (average of the coupling factor over a sampling period τ), which defines the fraction of applied bias to the unbiased gradient. The underlying kMC algorithm and the related transition time follows the approximation that the applied energy equals the energy of the transition state energy for a particular process. As mentioned before, this particular property of our method is maintained by a bias-force renormalization for each timestep, which depends on a coupling factor $\langle\alpha\rangle_\tau$. As we will show in the Methods and Results sections, the unbiased partition is obtained for $\langle\alpha\rangle_\tau$ -values in a region which allows unperturbed enhanced sampling of partitions. Large $\langle\alpha\rangle_\tau$ -values result in perturbed trajectories, while too small coupling values lead to weak sampling efficiencies.

2.3 Theory

We use a modified Langevin equation of motion of positions \mathbf{x}_t and momenta \mathbf{p}_t to express the propagation of the system with time t in the canonical ensemble. This equation is expressed as follows³² :

$$\begin{aligned} d\mathbf{x}_t &= \mathbf{M}^{-1} \mathbf{p}_t dt \\ d\mathbf{p}_t &= \mathbf{M}^{-1} \nabla V dt + \mathbf{A}(\nabla V, t) dt \delta_{kMC/MC} \\ &\quad - \gamma \mathbf{M}^{-1} \mathbf{p}_t dt + \sqrt{2\gamma\beta^{-1}} dW_t, \end{aligned} \quad (1)$$

where \mathbf{M} stands for the mass tensor, V is the potential energy function, γ is a friction coefficient, W_t stands for a random noise

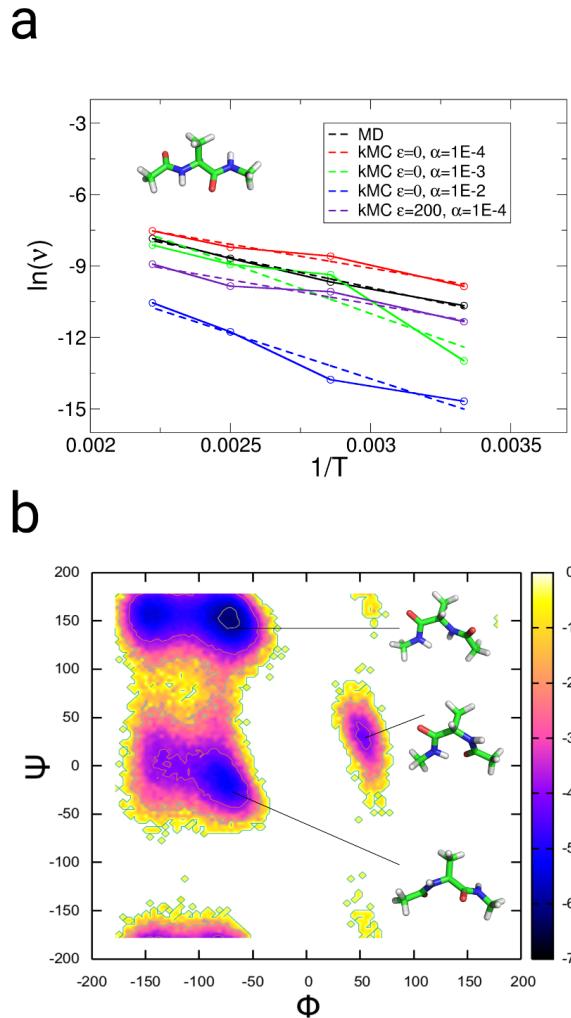


Fig. 2 Results from MD and MD/kMC simulations with different $\langle\alpha\rangle_\tau$ -parameters of the Diananine peptide in explicit solvent ($\varepsilon = 0$ and 200). (a) Arrhenius analysis of the logarithm of the transition frequency of the Φ -angle of the Diananine peptide as a function of the inverse temperature (regression - dashed lines, data - solid lines). (b) Free energy landscape at 300 K as function of dihedral angles Φ and Ψ of the Diananine peptide from a MD/kMC simulation with $\langle\alpha\rangle_\tau = 1E-4$, $\varepsilon=0$.

and β for the inverse temperature $(k_B T)^{-1}$, where k_B stands for Boltzmann's constant. The dynamics described by equation 1 are ergodic, if V is sufficiently smooth (i.e. as long as there are no abrupt changes in the potential energy function such as strong repulsions). The bias \mathbf{A} is expressed as :

$$\mathbf{A}(\nabla V, t) = \alpha \mathbf{A}_0(t) \frac{|\nabla V|}{|\mathbf{A}_0(t)|}, \quad (2)$$

where α is a coupling constant in the range from 1E-2 to 1E-4 (see Figure 1), in analogy to a linear boost factor used in works by McCammon *et al.*^{9,10}. We have found that for α -value below 1E-3 (0.1 % of $|\nabla V|$), the dynamics and kinetics are not affected by the added bias. The bias vector $\mathbf{A}_0(t)$ for each component is calculated during each phase with length τ by :

$$\mathbf{A}_0(t) = \tau^{-1} \sum_{t=0}^{\tau} \sum_{ij, i \neq j}^N \omega((\mathbf{x}_i(t) - \mathbf{x}_j(t))(\mathbf{a} - 2\xi)), \quad (3)$$

where ξ is a vector dependent on a uniform random variable $\in [0, 1]$, \mathbf{a} is a unit vector $\{e_x, e_y, e_z\}$, and N stands for the number of particles (segments). In this work, we used all atoms of each peptide as number of segments. However, if a representation fulfills the equation $\sum_i^{N_k} \mathbf{x}_i - \sum_j^{N_l} \mathbf{X}_j = 0$, or if only parts of the system should be considered, the system consisting of N_k particles can be coarse-grained to a smaller number of N_l segments. ω is expressed as :

$$\omega = (1 - d/R_c), \quad (4)$$

where d is the distance between interacting pairs of atoms and R_c stands for the cutoff. We note that this expression is derived from the conservative force \mathbf{F}^C and the random force \mathbf{F}^R from DPD²⁹⁻³¹, which conserve momentum and energy. The random variable ξ represents a pseudo-temperature, which effectively acts between pairs of particles.

The energy ΔE_j added to each component j corresponds to :

$$\Delta E_j = \alpha \frac{|\nabla V_j|}{|\mathbf{A}_{j0}(t)|} \langle \omega_j d_j | \mathbf{a}_j - 2\xi_j | \rangle_\tau. \quad (5)$$

Simulations are performed in explicit solvent and the solvent relaxation is still on the MD-timescale, which is crucial for the folding and relaxation behavior of the peptide, especially in peptide aggregation. We introduce a fluctuation parameter ε , which guarantees sampling around an average $\langle \alpha \rangle_\tau$ -value, while energy fluctuations are allowed in the range of ε . We tested our methodology at first using a value $\varepsilon = 0$ to evaluate the effect of the $\langle \alpha \rangle_\tau$ value on the dynamics.

For the simulations of peptide aggregation (Aβ 16-22), we modified expression 3, to :

$$\begin{aligned} \mathbf{A}_0(t) = & \tau^{-1} \sum_{t=0}^{\tau} \left[\sum_{ij, i \neq j}^N \omega((\mathbf{x}_i(t) - \mathbf{x}_j(t))(\mathbf{a} - 2\xi)) \right. \\ & \left. - 2 \sum_{ij', i \neq j'}^{N'} \xi \omega(\mathbf{x}_i(t) - \mathbf{x}_{j'}(t)) \right], \end{aligned} \quad (6)$$

where N' segments j' act attractive on segments i depending on a normal distributed random-number ξ' . For Aβ 16-22, we used the first 2 monomers of the fibril as segments N' . We introduce an approach for the coupling strength α , where we allow $\alpha(t)$ to fluctuate depending on a constant ε using :

$$\alpha(t) = \alpha \times (\varepsilon - \xi \varepsilon), \quad (7)$$

where ξ is a normal distributed random number with $\xi \in [0, 1]$ with $\langle \xi \rangle = 0.5$. We set the condition that $\alpha(t) = \alpha$ for a value of $\varepsilon = 0$. In both cases we obtain :

$$\langle \alpha(t) \rangle_\tau = \alpha. \quad (8)$$

This idea is based on the well-tempered ensemble idea, where the potential energy itself serves as bias, while the ensemble averages remain conserved⁴. (For the aggregation simulations, we use $\varepsilon = 200$, while we use $\varepsilon = 100$ for TrpCage and $\varepsilon = 10$ for TrpZip4. The value of ε primarily defines the energy range in which the dynamics take place (A value for $\langle \alpha \rangle_\tau = 1E-4$, $\varepsilon = 10$ would mean that dynamics take place at $\alpha(t) \leq 1E-3$). The magnitude of ε is derived from the importance of energy fluctuations for conformational transitions in the system, which means that the value of ε is in principle system dependent and has to be chosen such as to obtain a high sampling efficiency. In the case of peptide aggregation, we assumed that energy fluctuations are dominant over a long time-range, when the monomeric unit is separated. Thus, we chose a comparatively high value of 200, equal to dynamics in the range $\alpha(t) \leq 5E-2$, with $\langle \alpha \rangle_\tau = 2.5E-4$. For TrpCage, we found that a value of $\alpha(t) \leq 1E-2$ with $\langle \alpha \rangle_\tau = 1E-4$ accelerates sampling with no dramatic shift in the occupation densities along the free energy landscape. The rates and time-dependent dynamics of TrpZip4 are very sensitive to $\langle \alpha \rangle_\tau$, especially due to the higher forces and gradients in the collapsed state (Trp-Trp interaction). In that case, we chose a value of $\varepsilon = 10$. In general, we state that ε -values in the range from 10-100 with $\alpha < 2.5E-4$ allow an accurate sampling, while the efficiency of the kMC/MD simulation is enhanced. For monomeric folding, we use $N' = 0$.)

In terms of the partition function Z of the system which is affected by the bias in the canonical ensemble, this means that³³ :

$$Z = \frac{1}{h^{3N} N!} \int \exp(-\beta[H(\mathbf{r}_N, \mathbf{p}_N) + \Delta E_j]) d\mathbf{r}'_N d\mathbf{p}'_N. \quad (9)$$

Using equation 5, the equivalence of $H = T + V$,

$$\frac{1}{|\mathbf{A}_{j0}(t)|} \langle \omega_j d_j |\mathbf{a}_j - 2\xi_j| \rangle_\tau \leq 1, \quad (10)$$

and $V = \int \nabla V dr$ we obtain :

$$H(\mathbf{r}_N, \mathbf{p}_N) + \sum_j^N \Delta E_j = T(\mathbf{p}_N) + V(\mathbf{r}_N) + \langle \alpha \rangle_\tau V(\mathbf{r}_N) \quad (11)$$

$$= T(\mathbf{p}_N) + V(\mathbf{r}_N)(1 + \langle \alpha \rangle_\tau), \quad (12)$$

where T stands for the kinetic energy in the system.

For the configuration partition function Q , this leads to :

$$Q = \int \exp(-\beta[V(\mathbf{r}_N) + \langle \alpha \rangle_\tau V(\mathbf{r}_N)]) d\mathbf{r}'_N d\mathbf{p}'_N, \quad (13)$$

and :

$$\begin{aligned} Z &= \frac{1}{h^{3N} N!} \int \exp(-\beta(T(\mathbf{p}_N) + V(\mathbf{r}_N)[1 + \langle \alpha \rangle_\tau])) d\mathbf{r}'_N d\mathbf{p}'_N \\ &\approx \frac{1}{h^{3N} N!} \int \exp(-\beta H(\mathbf{r}_N, \mathbf{p}_N)) d\mathbf{r}_N d\mathbf{p}_N, \end{aligned} \quad (14)$$

which is approximately equal to the canonical distribution for if $\langle \alpha \rangle_\tau$ sufficiently small. For $\langle \alpha \rangle_\tau = 0$, we obtain the unbiased canonical partition. The same holds for the transition state partition :

$$Z^\dagger \approx \frac{1}{h^{3N} N!} \int \exp(-\beta H^\dagger(\mathbf{r}_N^\dagger, \mathbf{p}_N^\dagger)) d\mathbf{r}_N^\dagger d\mathbf{p}_N^\dagger,$$

and the biased transition state energy ΔG^\dagger remains approximately unaffected if $\langle \alpha \rangle_\tau$ is sufficiently small.

2.4 Kinetic Monte Carlo Sampling - MD/kMC

In the kinetic Monte Carlo scheme, the rate r_j for each segment j is :

$$r_j = v(\alpha) \exp\left(-\frac{\Delta E_j}{RT}\right), \quad (15)$$

where $R = N_A k_B$ with N_A being Avogadro's constant, and T stands for the instantaneous temperature. We chose a prefactor $v(\alpha)$ of $(1/N) 1E + 10 s^{-1} \times a(\alpha)$ for $\varepsilon=0$ and a prefactor $(1/N) 1E + 12 s^{-1} \times a(\alpha)$ for $\varepsilon>0$ (see Section : Kinetics and convergence behavior), where $a(\alpha)$ is a linear correction factor which depends from the $\langle \alpha \rangle_\tau$ -coupling value. With $R_j = \sum_{i=1}^j r_i$ and $R_N = \sum_{j=1}^N r_j$, we select a segment j for execution of an event (with $\delta_{kMC} = 1$), using:

$$R_{j-1} < \xi R_N \leq R_j, \quad (16)$$

where ξ is a uniform distributed random number $\in [0, 1]$ ^{23,24}. The time is updated according to :

$$\Delta t = -\frac{\ln \xi}{R_N}. \quad (17)$$

We refer to equation 15, and state that Δt corresponds to a realistic transition time for sufficiently small $\langle \alpha \rangle_\tau$ -values.

2.4.1 Algorithm

1. Propagate the system using equation 1, while the bias is reweighted for each timestep with equation 2. Collect bias gradients $\mathbf{A}_{j0}(t)$ for each segment, according to equation 3.
2. At $\Delta t = \tau$, which equals the time of one phase, calculate energies E using equation 5.
3. Calculate corresponding rates r_j for each segment j , using equation 15.
4. Select segment j using equation 16 and update time with equation 17.
5. Apply bias on segment j over period with length τ and continue with 1.

2.5 Simulation parameters and system preparation

In all simulations, we used the AMBER99SB force-field for the description of the protein interactions in each system^{34,35}. For the description of electrostatic interactions, we used the Particle Mesh Ewald algorithm with a real-space cutoff of 1.0 nm. The Lennard-Jones interactions were calculated with the same cutoff. The neighborlist with a cutoff of 1.0 nm was updated every 5th timestep. We used a timestep of 1 fs. We used the Nosé-Hoover thermostat with $t_c = 1.0$ ps and a Parrinello-Rahman barostat with $t_p = 1.0$ ps with 1 atm as reference pressure in the explicit solvent simulations. We used the GROMACS-4.5.5 simulation package for the MD simulations and modified versions of the same package for the enhanced simulations^{36,37}. In all simulations isotropic periodic boundary conditions have been applied. In all explicit solvent simulations, we used the SPC/E water model and the SET-TLE algorithm for a rigid description of the water geometry^{38,39}. In all MD/kMC simulations, we applied $\tau = 10$ ps as length for one kMC-phase.

For the simulations of Diananine peptide, we centered Ace-Ala-NMe into a cubic box with dimensions $2.26703 \times 2.26703 \times 2.26703$ nm³ and filled the box with 371 SPC/E waters. We performed 4 MD simulations over 1 μ s for temperatures 300, 350, 400 and 450 K. For each temperature, we performed MD/kMC simulations with $\langle \alpha \rangle_\tau = 1E-4$, 1E-3 and 1E-2 over a simulation-time of 200 ns

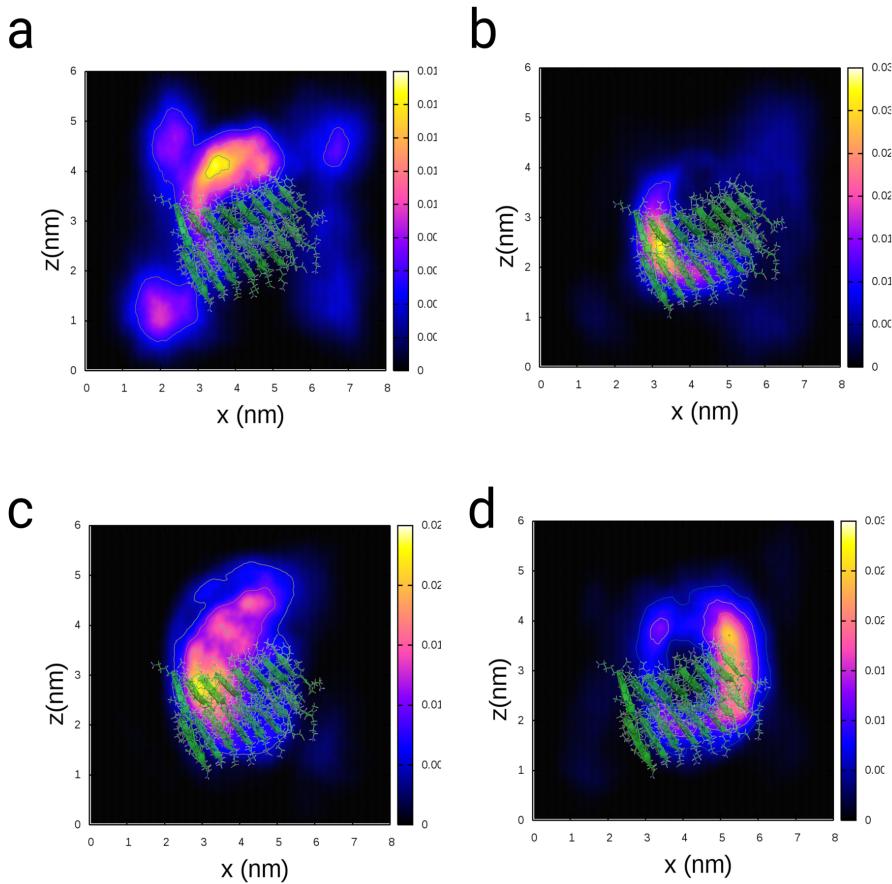


Fig. 3 Probability based projection of the free monomer coordinates in each trajectory to the x-z plane (after fitting to the fibril coordinates as a reference) from simulations 1-4 of an $\text{A}\beta$ 16-22 monomer to a fibril of the same peptide. Simulations 1-4 (a-d). We find a preferential occurrence of the monomer near the termini of the fibril.

MD time. We performed one simulation of TrpZip4 at $\langle\alpha\rangle_\tau=5E-4$. We simulated $\text{A}\beta$ 16-22 using $\langle\alpha\rangle_\tau = 2.5E-4$ and a ε -value of 200.0. For the simulations of TrpCage folding, we centered the extended peptide (NLYIQWLKDGGPSSGRPPPS⁴⁰) in a cubic box with dimensions $7.022 \times 7.022 \times 7.022 \text{ nm}^3$ and filled the box with 11424 SPC/E waters. We added one chloride ion to the system. In the simulations of TrpZip4 (GEWTWDDATKTWTWTEX⁴¹), we centered the extended peptide in a cubic box with dimensions $7.39 \times 7.39 \times 7.39 \text{ nm}^3$ and filled the box with 13242 SPC/E waters. We added 2 sodium ions to the system. We simulated TrpCage over 250 ns MD time for each $\langle\alpha\rangle_\tau$ -value. TrpZip4 was simulated over 250 ns MD time for each $\langle\alpha\rangle_\tau$ -value. In the simulation of aggregation of monomeric $\text{A}\beta$ 16-22, we centered a double stranded fibril of $\text{A}\beta$ 16-22 (8×2 units, taken from ref.⁴²)

in a triclinic box with dimensions $8.356 \times 5.51 \times 6.521 \text{ nm}^3$ and placed an extended monomer (ACE-KLVFFAE-NHE) 1 nm away from the fibril. We filled the box with 9105 SPC/E waters and added 5 sodium / 5 chloride ions. We tested the simulations for convergence in relation to the mobility of the monomer. We performed 4 simulations over approximately 260 ns each of the $\text{A}\beta$ system. For simulations with $\langle\alpha\rangle_\tau \leq 2.5E-4$, we used a value of 1.0 for $a(\alpha)$.

2.6 Analysis

We analyzed the transition time of the Φ -angle of Diananine peptide as the average time τ_a needed to transition between the region $\Phi < 0^\circ$ to the region $\Phi > 0^\circ$. We then determined the fre-

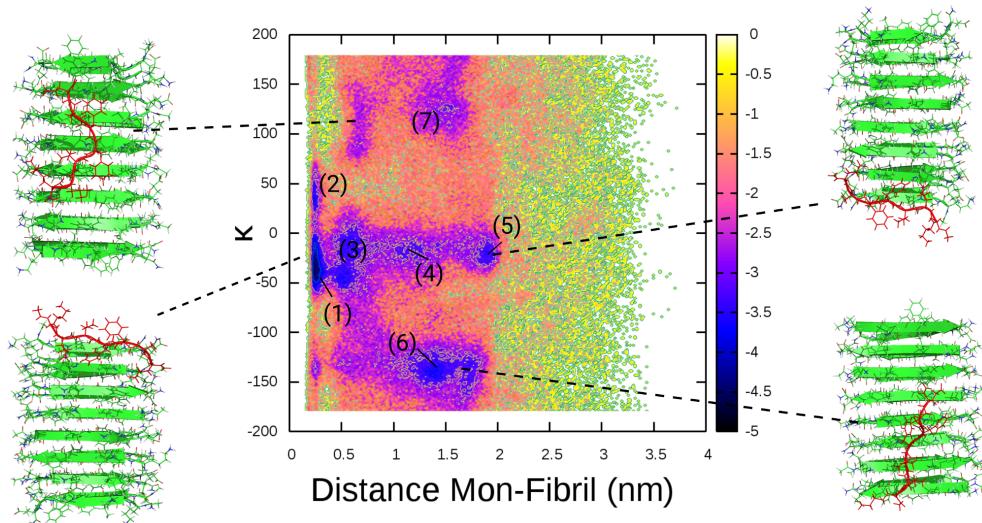


Fig. 4 Free energy landscape as function of angle k and monomer distance to the tip of the fibril averaged over simulations 1-4 of $\text{A}\beta$ 16-22 aggregation to a fibril of the same peptide. We find a preferential orientations of the peptide chain parallel in plane to the strands of the fibril or parallel to the fibril axis. We observe a preferential occurrence of the monomer near the termini of the fibril.

quency using $v_a = 1/\tau_a$ (see Figure 2 and Figure 1S). For the determination of the activation energy E_a , we used the relations :

$$v_a = A_a \exp\left(-\frac{E_a}{RT}\right), \quad (18)$$

and

$$\ln v_a = \ln A_a - \frac{E_a}{RT}. \quad (19)$$

We defined the dihedral Φ by the atoms : C-N-CA-C and Ψ using : N-CA-C-N of the Diananine peptide. For the measurement of the free energies ΔF , we used :

$$\Delta F = -k_B T \ln \frac{P}{P_{min}}, \quad (20)$$

where P_{min} stands for the minimal probability of the projection on 2 quantities.

We determined the folding time using the difference of the total simulation time and the average time in the folded ensemble with $RMSD_{C\alpha-C\alpha} < 0.3$ nm. In the analysis of the $\text{A}\beta$ 16-22 simulation, we defined the dihedral angle k between N (N-NHE (monomer), C-ACE (monomer), N-NHE (fibril-first monomer), C-ACE (fibril-first monomer) (see Figure 1 c). For the determination of the spatial probability distribution of the $\text{A}\beta$ 16-22 monomer, we first fitted the trajectory onto the fibril axis (gmx_trjconv) and second projected the trajectory of the monomer onto the x-z plane and generated a histogram over a mesh in that plane (in house code).

2.7 Implementation

The MD/kMC module has been implemented into the molecular dynamics module of the GROMACS-4.5.5 package into /src/kernel/md.c. A global call to collect forces and coordinates is performed and bias forces are re-distributed onto each parallel core.

2.8 Performance of MD/kMC

The computational performance of the MD/kMC simulation method is equal to MD/MC, while both techniques need to perform global calls to collect and re-distribute forces, which restricts the scalability of the algorithm. As an estimate, the computational performance regarding the MD-time is about 10-20 % lower than the conventional GROMACS-4.5.5 implementation. For $\langle \alpha \rangle_\tau$ equal to 1E-4, the acceleration is 8-fold for Diananine in comparison with conventional MD. We note that this factor increases with the system size. We estimate an acceleration equal to a factor of 29 in the case of TrpCage for $\langle \alpha \rangle_\tau$ equal 1E-4, $\epsilon = 0$. A factor of ϵ on the order of 100 increases the sampling efficiency by a factor of approximately 2-5.

3 Results and Discussion

3.1 Kinetics and convergence behavior : Diananine

We tested the reliability of the rate equation of the MD/kMC algorithm in simulations of Diananine with different $\langle \alpha \rangle_\tau$ -parameters for $\epsilon = 0$ and compared with the kinetics obtained from 1 μs trajectories over a temperature range from 300 to 450 K (see Figure 2). We measured the transition frequency of the Φ -angle of this peptide, and found that the transition frequencies and activation

energies for $\langle \alpha \rangle_\tau$ -values in the range from 1E-2 to 1E-3 are qualitatively in agreement with the unbiased MD trajectory. A coupling parameter of 1E-4 quantitatively reproduces the transition frequency of the Φ -angle (see Figures 2 a, 1S). While the average transition time τ for MD ranges from 42696.7 to 2550.8 ps in a temperature range from 300 to 450 K, the average transition time of MD/kMC with $\langle \alpha \rangle_\tau$ of 1E-4 ranges from 18984.6 to 1853.0 ps for the same temperatures ($\epsilon=0$) and from 84268.42 to 7457.9 ps ($\epsilon=200$). The shift in the transition time for $\epsilon=200$ is caused by the small system size of Dinalanine and a comparatively low impact of energy fluctuations on that system. In other words, the system does not fully perform dihedral transitions at the level of allowed energy fluctuations mainly due to its small system size and its low solvent accessible surface, i.e. a low level of interaction with surrounding water. That means that a certain fraction of the added bias as function of ϵ dissipates rather into longitudinal diffusion of Dinalanine than into an internal dihedral transition. A larger solvent accessible surface would lead to a redistribution of the applied bias into internal degrees of freedom. Taking into account a statistical error of approx. 8 ns at 300 K and approx. 30 ns at 450 K at an elevated thermal fluctuation, we note that this is a good quantitative agreement with unbiased MD. In a graphical comparison of the transition frequencies at the different temperatures, we determined a match of MD/kMC transition times in the case of an $\langle \alpha \rangle_\tau$ -value of approximately 2.5E-4 at a temperature of 300 K at $\epsilon=0$ (see Figure 4S). We determined the dependency of the transition time of the Φ -angle from $\langle \alpha \rangle_\tau$ in a linear fit and compared the value with conventional MD (see Figure 3S). We obtained a linear factor $a(\alpha)$ of 2.296E+8 ps $\langle \alpha \rangle_\tau^{-1}$ by which the transition time depends from $\langle \alpha \rangle_\tau$ (see equation 15). The activation energy in the unbiased MD-trajectory is 21.1 kJ/mol, while for the MD/kMC trajectory with a coupling parameter of 1E-4 we obtain a value of 16.7 kJ/mol ($\epsilon=0$) and 16.9 kJ/mol ($\epsilon=200$). For coupling parameters with $\langle \alpha \rangle_\tau$ -values 1E-3 and 1E-2, we obtain values of 35.11 and 31.8 kJ/mol (see Figure 3S c). We note that for $\langle \alpha \rangle_\tau$ -values at $\langle \alpha \rangle_\tau = 1E-4$ kinetics are in good agreement for values of $\epsilon=0$ and 200. Therefore, we chose simulations at $\langle \alpha \rangle_\tau = 1E-4$ to $\langle \alpha \rangle_\tau = 5E-4$ in folding and aggregation simulations as reference for $a(\alpha)$.

We compared the free energy landscapes of Dinalanine as function of Φ and Ψ from MD/kMC at 300 K. The free energy landscapes of Dinalanine at 300 K is in good quantitative agreement with the MD result (see Figure 2S and Figure 2 b).

Dinalanine has been investigated in a number of computational studies^{2,43,44}. Our value of the effective activation barrier of 21 kJ/mol (5.02 kcal/mol) is in agreement with results from Metadynamics simulations (8 kcal/mol)^{17,18}. We mention that Tiwary *et al.* stiffened the dihedral potential along the Φ -angle, while Bolhuis *et al.* loosened the potential along the same dihedral^{17,18}. That might influence the kinetics and lead to different kinetic

observables. We have used the standard AMBER99SB forcefield parametrization and used conventional MD simulations with the same forcefield as reference to validate our approach.

3.2 Kinetics and convergence behavior : TrpCage and TrpZip4

For both systems, TrpCage and TrpZip4, we find that the dynamical evolution with $\langle \alpha \rangle_\tau$ -values larger than 1E-3 is perturbed. For these high coupling parameters, the native state is not populated sufficiently, with primarily near native states at $RMSD_{Ca-Ca}$ 0.5 nm seen. The larger the $\langle \alpha \rangle_\tau$ -value, the higher the fluctuations of the system and the deviation from the unbiased partition (see section Methods), which is underlined by the simulations at $\langle \alpha \rangle_\tau$ values of 1E-3 and 1E-2 ($\epsilon=0$). For $\langle \alpha \rangle_\tau$ -values lower than 1E-3, folding of both systems can be sampled accurately with kinetics in good agreement with experimental and computational results from other studies^{21,25,26,45-62}. Energy fluctuations play a larger role in β -stranded folding, with both systems sampled with a higher efficiency using $\epsilon > 0$ (See Supplementary Material SI: Section 2, Figures 5S-9S). Folding times and conformational dynamics of both systems are consistent with previous kMC/MD studies^{25,26}.

3.3 Fibril growth: Monomer addition to an $\text{A}\beta$ 16-22 fibril

We simulated the addition of one extended monomeric fragment of $\text{A}\beta$ 16-22 to a double stranded fibril of $\text{A}\beta$ 16-22 in explicit solvent using $\langle \alpha \rangle_\tau = 2.5E-4$, $\epsilon = 200$. The results are shown in Figures 3, 4 and Figures 10S and 11S.

By monitoring the distances of the free monomer to the fibril as function of time, we observe that the free monomer remains in the bulk solvent for 7.7 μs (simulation 1), 3.89 μs (simulation 2), 3.9 μs (simulation 3) and 300 ns (simulation 4) until an aggregation event occurs. In that diffusive period, the monomer forms hydrophilic contacts of N-ACE and C-NHE with the termini of the fibril at 1.2, 2.0, 4.7, 5.8 μs (simulation 1), 1.7 and 2.0 μs (simulation 2), 161 ns (simulation 4) (see Figures 10S, 11S). The aggregation event is correlated with a preferential orientation of the monomer in relation to the fibril, while hydrophobic contact formation occurs between residues Phe19, Phe20 with the same residues of the peptides located at the fibril. In simulations 1 and 2, we observe a preferential dihedral angle k (see Section Methods) of 0° at the time of the aggregation event at 7.7 and 3.89 μs , while the monomer forms contacts with the fibril at an angle of -180° and -40° at 3.9 μs in simulation 3 and at 300 ns in simulation 4 (see Figure 10S). The angles in the range of 0, 180 and -180° correlate with a parallel alignment of the monomer, while angles in between these values are related to an alignment of the monomer parallel to the fibril axis. At an angle of 0 degrees the N-terminus of the monomer is oriented towards the C-terminus of the first monomer of the fibril. After formation of a hydropho-

bic contact, the monomer remains in the orientation at 0° in the middle of the fibril for a simulation time ranging from $7.7 \mu\text{s}$ to $13.8 \mu\text{s}$ in simulation 1. In simulations 2-4, the monomer remains at the termini of the fibril with a preferential orientation parallel to the strands of the fibril, while the diffusion of the peptide is decreased in that aggregated state (see Figures 10S, 11S). That behavior is also reflected by the free energy landscapes as function of the angle k and the monomer distance, where the minima of approximately $-4 k_B T$ correlate with a stabilized aggregated state after a diffusive period (see Figure 10S).

In the spatial profiles of the monomer trajectories, which have been projected on the x - z plane, we observe the behavior described above (see Figure 3). In simulation 1, we observe occurrences of the monomer in the range $0 < x < 3 \text{ nm}$ and $0 < z < 2 \text{ nm}$, and at $0 < x < 2.5 \text{ nm}$, $4 < z < 6 \text{ nm}$, which correspond to contact formation process prior to the aggregation event of the monomer. The aggregated species in the middle of the fibril persists in the range $3 < x < 5 \text{ nm}$, $3 < z < 5 \text{ nm}$ with an occurrence of more than 38% (see Figure 3a). In simulations 2-4, bulk conformations as we observed in simulation 1 are less populated. Here, we observe highly populated areas at the termini of the fibril, which is correlated with a fast initial aggregation process in these simulations at these regions along the fibril (see Figure 3b-d). For simulation 2, we find a population at $2.5 < x < 4 \text{ nm}$, $1.5 < z < 3 \text{ nm}$ with a probability larger than 70% at the terminus of the fibril (see Figure 3b). Another area in the range from $5 < x < 7 \text{ nm}$ and $0 < z < 6 \text{ nm}$ is related to free diffusion of the monomer. For simulation 3, we find that the monomer preferentially resides at $2.5 < x < 3.5 \text{ nm}$ and $2 < z < 3.3 \text{ nm}$ with a probability larger than 70%, while we find a preference for $4.8 < x < 6 \text{ nm}$ and $1.7 < z < 4.3 \text{ nm}$ in the case of simulation 4 (see Figures 3c, d).

In the free energy landscape as function of angle k and the monomer distance to the tip of the fibril averaged over 4 simulations, we find 7 minima at $d=0.2 \text{ nm}$, $k=-45^\circ$ (1), $d=0.2 \text{ nm}$, $k=45^\circ$ (2), $d=0.5 \text{ nm}$, $k=-45^\circ$ (3), $d=1 \text{ nm}$, $k=-45^\circ$ (4), $d=1.8 \text{ nm}$, $k=-45^\circ$ (5), $d=1.5 \text{ nm}$, $k=-145^\circ$ (6) and $d=1.5 \text{ nm}$, $k=125^\circ$ (7) (see Figure 4). The minima (1-5) correspond to an antiparallel alignment (N-ACE-C-NHE - N-ACE-C-NHE) of the free monomer to the first monomer of the fibril, while the monomer is aligned parallel to the strands of the fibril. These minima correspond to a free energy of approx. -4 to $-5 k_B T$. That means that a preference exists for the free monomer to reside at the termini of the fibril in a parallel alignment. Minima (6) and (7) represent configurations in which the monomer is aligned parallel to the fibril axis, with these configurations less energetically favorable than the states 1-5. The states away from the fibril are energetically homogeneous and higher in energy than aggregated states.

We observed initial nonspecific hydrophilic contact formation between the N- and C-termini of peptide in the fibril and the monomer, with no specified hydrophobic contacts which are sepa-

rated by water. That initial sequence, where the monomer resides in the bulk, is followed by the formation of a strong hydrophobic contact by residues Phe19 and Phe20 with the fibril. This process occurs on a wide range of timescales from several hundreds of ns to approx. $8 \mu\text{s}$. The peptide remains in that aggregated state, while it changes its orientation and location on the fibril in the range of few nm. Parallel orientations of the free monomer to the strand-axis of the fibril and a position at the termini of the fibril are energetically favorable. In general, we observed in all 4 simulations, that a surface-aggregated state of the monomer to the fibril is the energetically favored state. That might be a result from the double stranded fibril model we used in that study or might indicate the importance of a dependency from the free-monomer concentration.

Aggregation of $\text{A}\beta$ fragments has been studied in a number of simulations⁶³⁻⁷⁶. It has been reported that the dock-lock mechanism can involve initial nonspecific or specific contacts between the free monomer and the fibril, followed by a final and slow conversion to the fibril via a hydrophobic contact formation^{63,65,66,71,72}. Our observation of a surface-aggregated state as part of the aggregation pathway is in agreement with another recent work by Schor *et al.*⁷⁴ and a coarse-grained simulation study by Frenkel and coworkers⁷⁷.

4 Conclusions

In this paper, we introduced a novel MD/kMC technique, which is based on a dynamical redefinition of biases and a renormalization of its magnitude in relation to the unbiased gradient. The renormalization guarantees sampling of the underlying partition function. To account for the energy fluctuations in peptide aggregation simulations, we introduced a factor ϵ which applies fluctuations on the bias, while the ensemble average is maintained. We tested our novel algorithm on Dialanine and folding of 2 peptides : TrpCage and TrpZip4. Finally, we applied our algorithm in aggregation simulations of one monomeric fragment of $\text{A}\beta$ 16-22 to a double stranded fibril of the same peptide. We observe quantitative agreement of folding times of Dialanine and both peptides with unbiased MD simulations and prior kMC/MD results. In the aggregation simulations, we find a heterogeneous free energy landscape of the free monomer. The initial aggregation events are in agreement with the dock-lock mechanism presented in prior simulation studies. We observe prior formation of hydrophilic contacts between the N- and C-termini of the monomer and the fibril, which is followed by the formation of hydrophobic contacts (Phe19, Phe20). In that strongly aggregated state, the monomer diffuses along the surface of the fibril with a preference of the monomer to reside near the termini of the fibril in a parallel orientation to the fibril strand axis. The bias introduced into the kMC/MD method is adaptive and does not directly introduce *a priori* a reaction coordinate for the attachment

of the monomer to the terminus of the fibril, which is a significant advantage in contrast to widely used umbrella sampling and Metadynamics methods, while the computational cost is much lower compared to conventional REMD. In REMD-simulations of TrpZip2 and TrpZip4^{53,78}, which reached a comparable timescale in the range of tens of microseconds, the total computational time (number of CPUs (Intel Xeon E5-2695) times total wall-time (25 days)) was larger than in our approach by a factor of approximately 36. Additionally, the kMC/MD method contains an *on the fly* information about rates and the time in which aggregation events occur in the system.

5 Acknowledgments

The authors thank Dr. Chun Wu for providing the structural data for the A β 16-22 fibril. EKP acknowledges support by a grant from the Swiss Platform for Advanced Scientific Computing. Simulations were carried out at the Texas Advanced Supercomputing Center, XSEDE grant no. (TG-MCA05S027). JES acknowledges support by the grant no. MCB-1158577 from the National Science Foundation.

References

- 1 A. F. Voter, *J. Chem. Phys.*, 1996, **106**, 4665–4677.
- 2 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 12562–12566.
- 3 A. S. Keys, D. Chandler and J. P. Garrahan, *Phys. Rev. E*, 2014, **92**, 022304.
- 4 M. Bonomi and M. Parrinello, *Phys. Rev. Lett.*, 2010, **104**, 190601.
- 5 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
- 6 T. Okabe, M. Kawata, Y. Okamoto and M. Mikami, *Chem. Phys. Lett.*, 2001, **335**, 435–439.
- 7 K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.*, 1996, **65**, 1604–1608.
- 8 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
- 9 D. Hamelberg, J. Mongan and J. A. McCammon, *J. Chem. Phys.*, 2004, **120**, 11919–11929.
- 10 D. Bucher, L. C. T. Pierce, J. A. McCammon and P. R. L. Markwick, *J. Chem. Theor. Comput.*, 2011, **7**, 890–897.
- 11 N. V. Plotnikov, S. C. L. Kamerlin and A. Warshel, *J. Phys. Chem. B*, 2011, **115**, 7950–7962.
- 12 N. V. Plotnikov and A. Warshel, *J. Phys. Chem. B*, 2012, **116**, 10342–10356.
- 13 J. Lameira, I. Kupchenko and A. Warshel, *J. Phys. Chem. B*, 2016, **120**, 2155–2164.
- 14 Z. Z. Fan, J.-K. Hwang and A. Warshel, *Theor. Chem. Acc.*, 1999, **103**, 77–80.
- 15 M. Roca, H. Liu, B. Messer and A. Warshel, *Biochemistry*, 2007, **46**, 15076–15088.
- 16 A. Dryga and A. Warshel, *J. Phys. Chem. B*, 2010, **114**, 12720–12728.
- 17 P. Tiwary and M. Parrinello, *Phys. Rev. Lett.*, 2013, **111**, 230602.
- 18 P. G. Bolhuis, C. Dellago and D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 5877.
- 19 H. Grubmüller, *Phys. Rev. E*, 1995, **52**, 2893.
- 20 J. Juraszek, G. Saladino, T. S. van Erp and F. L. Gervasio, *Phys. Rev. Lett.*, 2013, **110**, 108106.
- 21 F. Marinelli, F. Pietrucci, A. Laio and S. Piana, *PLOS Comput. Biol.*, 2009, **5**, e1000452.
- 22 K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, **334**, 517–520.
- 23 A. B. Bortz, M. H. Kalos and J. L. Lebowitz, *J. Comput. Phys.*, 1975, **17**, 10–18.
- 24 D. T. Gillespie, *J. Comput. Phys.*, 1976, **2**, 403–434.
- 25 E. K. Peter and J.-E. Shea, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6330–6340.
- 26 E. K. Peter, I. V. Pivkin and J.-E. Shea, *J. Chem. Phys.*, 2015, **142**, 144903.
- 27 E. K. Peter, B. Dick, I. Stambolic and S. A. Baeurle, *Proteins*, 2014, **81**, 394.
- 28 E. Peter, B. Dick and S. A. Baeurle, *J. Chem. Phys.*, 2012, **136**, 124112.
- 29 R. D. Groot and P. B. Warren, *J. Chem. Phys.*, 1997, **107**, 4423.
- 30 P. Español and P. B. Warren, *Europhys. Lett.*, 1995, **30**, 191–196.
- 31 P. J. Hoogerbrugge and J. M. V. A. Koelman, *Europhys. Lett.*, 1992, **19**, 155.
- 32 J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille and C. Chipot, *J. Phys. Chem. B*, 2015, **119**, 1129–1151.
- 33 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, New York, NY, USA, 1989.
- 34 P. A. Kollman, *Acc. Chem. Res.*, 1996, **29**, 461–469.
- 35 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.
- 36 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
- 37 B. Hess, D. van der Spoel and E. Lindahl, *Gromacs User Manual* 4.6, 2012, 1.
- 38 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
- 39 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.
- 40 J. W. Neidigh and R. M. Fesinmeyer, *Nat. Struct. Biol.*, 2002, **9**, 425–430.

- 41 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 5578–5583.
- 42 C. Wu, Z. Wang, H. Lei, Y. Duan, M. T. Bowers and J.-E. Shea, *J. Mol. Biol.*, 2008, **384**, 718–729.
- 43 D. J. Tobias and C. L. Brooks III, *J. Phys. Chem.*, 1992, **96**, 3864–3870.
- 44 W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, Y. Zhestkov and R. Zhou, *J. Chem. Phys. B*, 2004, **108**, 6582–6594.
- 45 D. Du, Y. Zhu, C.-Y. Huang and F. Gai, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 15915–15920.
- 46 Y. Xu, R. Oyola and F. Gai, *J. Am. Chem. Soc.*, 2003, **125**, 15388–15394.
- 47 J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 76–88.
- 48 C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen and V. S. Pande, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 4077–4082.
- 49 J. Juraszek and P. G. Bolhuis, *J. Phys. Chem. B*, 2009, **113**, 16184–16196.
- 50 B. N. Markiewicz, L. Yang, R. M. Culik, Y. Q. Gao and F. Gai, *J. Phys. Chem. B*, 2014, **118**, 3317–3325.
- 51 A. M. Razavi and V. A. Voelz, *J. Chem. Theor. Comput.*, 2015, **11**, 2801–2812.
- 52 G. H. Zerze, B. Uz and J. Mittal, *Proteins*, 2015, **83**, 1307–1315.
- 53 E. K. Peter, J.-E. Shea and I. V. Pivkin, *Phys. Chem. Chem. Phys.*, 2016, **18**, 13052–13065.
- 54 J. Juraszek and P. G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 15859–15864.
- 55 J. Juraszek and P. G. Bolhuis, *Biophys. J.*, 2008, **95**, 4246–4257.
- 56 C. D. Snow, B. Zagrovic and V. S. Pande, *J. Am. Chem. Soc.*, 2002, **124**, 14548–14549.
- 57 H. Meuzelaar, K. A. Marino, A. Huerta-Viga, M. R. Panman, L. E. J. Smeenk, A. J. Kettelarij, P. T. J. H. van Maarseveen, P. G. Bolhuis and S. Woutersen, *J. Phys. Chem. B*, 2013, **117**, 11490–11501.
- 58 H. Ren, Z. Lai, J. D. Biggs, J. Wang and S. Mukamel, *Phys. Chem. Chem. Phys.*, 2013, **15**, 19457–19464.
- 59 H. Neuweiler, S. Doose and M. Sauer, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 16650–16655.
- 60 L. Qiu, S. A. Pabit, A. E. Roitberg and S. J. Hagen, *J. Am. Chem. Soc.*, 2002, **124**, 12952–12953.
- 61 R. Zhou, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 13280–13285.
- 62 J. Zhang, M. Qin and W. Wang, *Proteins*, 2006, **62**, 672–685.
- 63 N. Schwierz, C. V. Frost, P. L. Geissler and M. Zacharias, *J. Am. Chem. Soc.*, 2016, **138**, 527–539.
- 64 T. Takeda and D. K. Klimov, *Biophys. J.*, 2009, **96**, 442–452.
- 65 N.-V. Buchete, *Biophys. J.*, 2012, **103**, 1411–1413.
- 66 M. Schor, J. Vreede and P. G. Bolhuis, *Biophys. J.*, 2012, **103**, 1296–1304.
- 67 P. H. Nguyen, M. S. Li, G. Stock, J.-E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 111–116.
- 68 G. Reddy, J.-E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 11948–11953.
- 69 M. Cheon, I. Chang and C. K. Hall, *Biophys. J.*, 2011, **101**, 2493–2501.
- 70 J. Nasica-Labouze and N. Mousseau, *PLoS Comput. Biol.*, 2012, **8**, e1002782.
- 71 W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghiraldi, J. P. Lee, P. W. Mantyh and J. E. Maggio, *Biochemistry*, 2000, **39**, 6288–6295.
- 72 F. Massi, J. W. Peng, J. P. Lee and J. E. Straub, *Biophys. J.*, 2001, **80**, 31–44.
- 73 J. A. Luiken and P. G. Bolhuis, *J. Phys. Chem. B*, 2015, **119**, 12568–12579.
- 74 M. Schor, A. S. J. S. Mey, F. Noé and C. E. MacPhee, *Chem. Phys. Lett.*, 2015, **6**, 1076–1081.
- 75 M. G. Krone, P. S. L. Hua, R. Zhou, B. J. Berne and J.-E. Shea, *J. Am. Chem. Soc.*, 2008, **130**, 11066–11072.
- 76 P. Soto, M. A. Griffin and J.-E. Shea, *Biophys. J.*, 2007, **93**, 3015–3025.
- 77 A. Saric, Y. C. Chebaro, T. P. J. Knowles and D. Frenkel, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 17869–17874.
- 78 E. K. Peter, M. Agarwal, B. K. Kim, I. V. Pivkin and J.-E. Shea, *J. Chem. Phys.*, 2014, **141**, 22D511.