RESEARCH ARTICLE

PROTEINS WILEY

Efficient and accurate binding free energy calculation of Aβ$_{9-40}$ protofilament propagation

Christina V. Frost[1] | Nadine Schwierz[2] | Martin Zacharias[1]

[1]Center of Functional Protein Assemblies, Technical University of Munich, Garching, Germany

[2]Institute of Physics, University of Augsburg, Augsburg, Germany

Correspondence
Martin Zacharias, Center of Functional Protein Assemblies, Technical University of Munich, Ernst-Otto-Fischer-Str. 8, Garching 85748, Germany.
Email: martin.zacharias@ph.tum.de

Abstract

Self-assembled aggregation of peptides and proteins into regular amyloid fibrils is associated with several neurodegenerative diseases. In case of Alzheimer's disease proteolytic cleavage products of the amyloid precursor protein form pathological amyloid-beta fibrils in a nucleation and propagation phase. The molecular details and thermodynamic driving forces of amyloid formation are not well understood, but are of high relevance for potential pharmacological interference. We used atomistic binding free energy simulations to calculate the free energy of protofilament propagation by an additional Aβ$_{9-40}$ peptide binding to the protofilament tip. It requires sampling of relevant conformational transitions which is challenging since the monomeric Aβ$_{9-40}$ peptide is intrinsically disordered. However, the convergence of umbrella simulations can be enhanced by applying additional restraining potentials on the axial, orientational and conformational degrees of freedom. The improved convergence leads to a much closer agreement with experimental binding free energy data compared to unrestrained umbrella sampling. Moreover, the restraining approach results in a separation of contributions to the total binding free energy. The calculated contributions indicate that the free energy change associated with the restriction of conformational freedom upon propagation makes a large opposing contribution of higher magnitude than the total binding free energy. Finally, optimization of the approach leads to further significant reduction of the computational demand which is crucial for systematic studies on mutations, denaturants and inhibitors in the fibril propagation step.

KEYWORDS
Alzheimer's disease, amyloid propagation, amyloid-beta formation, binding free energy calculation, disordered peptide binding, molecular dynamics simulation

1 | INTRODUCTION

Most proteins adopt well-defined globular structures necessary for metabolic, regulatory, or other functions. However, many proteins and peptides can also self-assemble into oligomeric and filamentous conformational states.[1–4] The latter include ordered amyloid fibrils with a characteristic cross-β-structure.[4] Such amyloid fibrils are associated with many neurodegenerative diseases including Parkinson's, Huntington's, and Alzheimer's disease.[5,6] In case of Alzheimer's disease pathological amyloid aggregates are formed by Aβ-peptides, which are products of the proteolytic cleavage of the amyloid precursor protein (APP).[7–10]

During formation of Aβ-fibrils one can distinguish a nucleation phase and a subsequent propagation step in which the preformed fibril grows by monomers binding to the ends (tips) of the fibril.[6,11] Even at high concentrations of Aβ-monomers the nucleation phase can take hours or days (in vitro), followed by the more rapid propagation (elongation) phase.

The structures of many amyloid fibrils (including Aβ) have been determined in recent years by x-ray crystallography, NMR spectroscopy, or CryoEM.[3,4,12] Interestingly, for many peptides and proteins not only one amyloid fibril structure but several different structures have been found depending on experimental setup or whether the fibrils were obtained in vitro or ex vivo.[3,11,13,14] Despite these extensive structural studies as well as kinetic and other biochemical experiments on amyloid fibril formation, little is known about the molecular mechanism and process of how amyloid fibrils nucleate and propagate.

In principle, molecular dynamics (MD) simulations are well suited to investigate the mechanism of the aggregation processes in atomic detail and at high time resolution of femtoseconds. Indeed, the process of aggregation and amyloid formation has already been subject of several simulation studies.[15-20] Unfortunately, both nucleation and propagation steps of amyloid formation involve time scales beyond the capabilities of standard MD simulations and thus require enhanced sampling schemes to obtain insights at the molecular level. In addition, most of the current protein force fields are well suited for studying globular folded proteins but might be less suited or less tested in case of alternative disordered or fibril like structures. It is therefore crucial to benchmark the simulations with experimental data such as the binding affinity of monomers at the fibril tips.[17,21]

The nucleation step of fibril formation is a complex process that may involve multiple long-lived metastable intermediates that are difficult to sample during MD simulations.[2] Protofibril propagation, on the other hand, can be investigated by extensive free energy simulations for instance using the Aβ$_{9-40}$ peptide as a model system.[17,19] In our previous study, umbrella sampling simulations yielded binding free energy estimates of Aβ$_{9-40}$ monomers associating at the fibril tips during the propagation step.[17] In addition, it was possible to characterize the free energy landscape and to distinguish a docking and locking phase for monomer association/dissociation. Furthermore, calculation of free energy profiles and diffusivity provided insight into the kinetics of the propagation step in reasonable agreement with available experimental data. However, the free energy calculations overestimated the binding affinity and it remained unclear whether the deviations resulted from shortcomings of the atomistic force fields or from insufficient sampling. Moreover, it was not possible to distinguish different contributions to the binding free energy resulting from translational, orientational and conformational changes during the propagation step.

For example, the binding of an Aβ monomer to the fibril tip results in a significant reduction of conformational phase space, as the peptide transits from an ensemble of intrinsically disordered conformations in bulk to a well-defined β-hairpin structure in the bound state. Such conformational restriction together with a restriction of the translational and orientational degrees of freedom of the binding partners make an opposing contribution to the binding process. For a comprehensive understanding of the amyloid propagation step, it is necessary to calculate the different contributions underlying the binding process. At the same time, however, sampling all possible conformational transitions from the unbound to the bound state can constitute a significant challenge for the simulations in particular if conformational states are separated by high energetic barriers.

In the current work, we present a computationally efficient method to calculate the absolute binding free energy of Aβ$_{9-40}$ monomer binding at the fibril tip during the propagation step. Restraining the translational, orientational and conformational degrees of freedom as suggested by Woo and Roux for protein-ligand binding[22] significantly improves the convergence and allows us to reproduce a binding free energy in excellent agreement with available experimental data. Moreover, calculation of the free energy contributions indicates that the restriction of conformational freedom upon binding adds a strong opposing component. Finally, reducing the computational cost is of general importance for studying amyloid fibril systems and could be helpful for the design of pharmacologically relevant compounds that modulate or inhibit amyloid propagation.

2 | METHODS

2.1 | Protofilament model system

The protofilament model system was composed of 11 wild type Aβ$_{9-40}$ monomers in all-atom resolution (see amino acid sequence in Figure 1A). The disordered residues 1–8 were omitted from the simulation model, as these had been suggested as less relevant for the growth of a stable protofilament core.[23-25] Based on solid-state NMR models of Petkova and coworkers (PDB code 2LMN),[23,24,26] the protofilament was constructed by stacking 11 monomers in native, U-shaped β-hairpin conformation along the protofilament axis Z with an inter-strand spacing of 0.48 nm (see Figure 1B). The resulting parallel β-sheets were referred to as NT- and CT-sheet, with each monomer contributing its N-terminal strand (amino acids 9–21) or C-terminal strand (amino acids 30–40) to one sheet, respectively (see Figure 1B, D). The connecting loop region consisted of amino acids 22–29.[27] As illustrated in Figure 1C, the NMR structure indicated that the two β-sheets were displaced against each other, resulting in asymmetric protofilament tips. This so called staggering enabled compact internal side-chain packing and charge compensation within the protofilament core (see Figure 1E).[25] In accordance with the model of Petkova and coworkers,[23,24,26] a stagger of +2 (in units of the inter-strand displacement of 0.48 nm) was used, which resulted in two unpaired, protruding strands per sheet, referred to as NT- and CT- tip (see Figure 1C). Note, that the NT-tip is also frequently referred to as even and the CT-tip as odd fibril tip. During simulations of protofilament growth, position restraints were applied to the protofilament tip to prevent the otherwise evolving strong twisting motions and to hence mimic larger protofilaments with a structurally stable binding site.
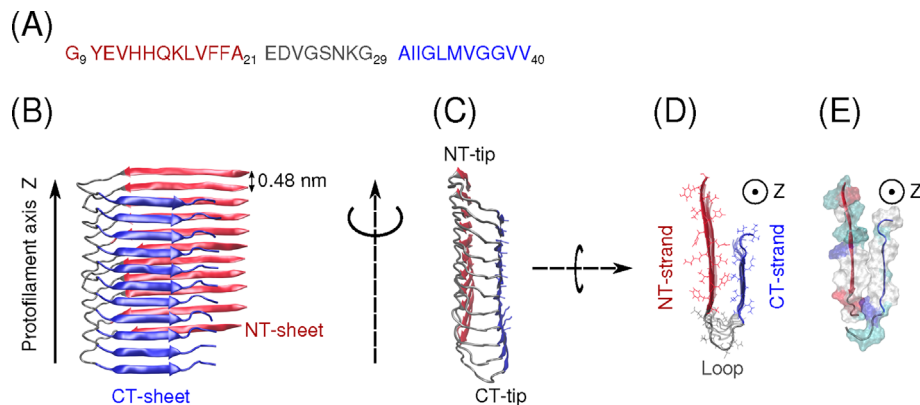
FIGURE 1 Protofilament model system. (A) Amino acid sequence of an Aβ$_{9-40}$ peptide colored according to its secondary structure in the bound form: NT-strand (red), CT-strand (blue) and connecting loop region (gray). (B) The protofilament model system was composed of 11 monomers. According to the secondary structure of the monomers, the β-sheets were denoted as NT- and CT-sheet. (C) The sheets were displaced against each other, resulting in asymmetric protofilament tips. These were denoted as NT- or CT-tip according to the protruding monomeric strand. (D) Top view onto the β-hairpin structure of the monomer at the NT-tip. (E) Van-der-Waals representation of amino acids within the cross section of two opposing, paired strands, illustrating the tight packing of predominantly hydrophobic (white) amino acids inside the protofilament core. Charged amino acids were colored in red (negative) and blue (positive), polar amino acids in green.

2.2 | Simulation setup and parameters

2.2.1 | Interaction parameters

Interactions were parametrized using the all-atom force field Charmm27 in combination with the explicit water model TIP3P.[28] Charmm27 has proven appropriate in describing both thermodynamic stability and conformational dynamics of Aβ filaments[25,27] and was used already in our previous work on the same system.[17] 150 mM NaCl were added to the solvent for neutralizing the negative net charge per Aβ$_{9-40}$ peptide and for imposing physiological boundary conditions. Short-range non-bonded interactions were calculated using grid-based Verlet neighborlists. The van-der-Waals and real-space Coulomb cutoff were set to 1.4 nm, which was larger than the recommended value of 1.0 nm.[29] Periodic boundary conditions were applied and the long-range Coulomb interactions were calculated using the fast smooth particle-mesh Ewald (PME) method with a grid spacing of 0.12 nm and default Gromacs PME parameters.[30]

2.2.2 | General simulation protocol

Simulations were performed with the Gromacs simulation software, version 4.6.5[29] as in our previous work.[17] The equations of motion were integrated using a standard leapfrog algorithm. All bonds were constrained with the LINCS algorithm,[31] which enabled an integration time step of 2 fs. Unless stated otherwise, the following simulation protocol was applied: The total internal energy of each system was first minimized in 50 000 steps of steepest descent algorithm. Then, the system was gradually heated up to the target temperature of 300 K in steps of 50 K and 50 ps each. A subsequent two-step equilibration was performed using the Berendsen weak-coupling scheme,[32] consisting of 500 ps in NVT ensemble at 300 K and 500 ps in NPT

ensemble at 300 K and at 1 bar with coupling time constants of $\tau_T = 0.1$ ps and $\tau_p = 1.0$ ps. During equilibration, harmonic position restraints were applied to backbone C_α atoms with a force constant of 1000 kJ/(mol nm^2). Umbrella simulations were performed in NPT ensemble at standard conditions of 300 K and at 1 bar using the Nose-Hoover thermostat[33,34] with a coupling time constant of $\tau_T = 0.5$ ps and the Parrinello-Rahman barostat[35] with $\tau_p = 5.0$ ps.

2.3 | Binding free energy calculation of Aβ$_{9-40}$ protofilament propagation

Previously, we used standard umbrella sampling for the simulation of Aβ$_{9-40}$ protofilament propagation.[17] This approach provided insights into the intermediate states during the propagation step and yielded additional information on the kinetics of fibril growth. Briefly, a one-dimensional potential of mean force (PMF) $W(\zeta_z)$ was calculated along the separation coordinate ζ_z. The latter was defined as the center-of-mass separation along the protofilament axis Z between the eleventh monomer (M11) at the protofilament tip and an additional twelfth monomer (M12) (see Figure 2). The phase space orthogonal to ζ_z remained unrestrained, including axial and orientational degrees of freedom of the monomer M12 relative to the protofilament, as well as internal conformational degrees of freedom of M12. Further details of the approach are given in Schwierz et al.[17] and in the Supporting Information (SI). However, obtaining converged results with the standard umbrella sampling setup was challenging since the peptide could adopt a large variety of conformations. To improve the accuracy of the calculations, we here used the absolute binding free energy method developed by Wovo and Roux[22] (WR-method). The method employs a set of orientational and conformational restraints during the induced dissociation of binding partners to improve convergence of free energy calculations.
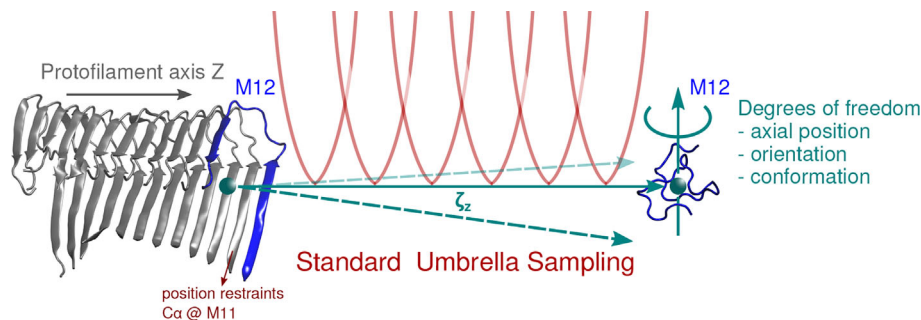
FIGURE 2 Illustration of the unrestrained standard umbrella sampling approach for $A\beta_{9-40}$ protofilament propagation at the NT-tip. The monomer at the NT-tip was denoted as M11 and the additional, twelfth monomer as M12. The separation coordinate ζ_z was defined as the center-of-mass separation (green spheres) between M11 and M12 along the protofilament axis Z. Orthogonal degrees of freedom remained unrestrained, including the axial position and orientation of M12 relative to the protofilament as well as internal conformational degrees of freedom of M12. Position restraints on M11 prevented the otherwise strong twisting motions of the protofilament (see also SI).

2.3.1 | Absolute binding free energy calculation using restraints

In the WR-approach,[22] the orthogonal phase space during umbrella sampling is reduced by introducing several harmonic restraints, resulting in a faster convergence of the distance PMF. Applied to $A\beta_{9-40}$ protofilament propagation, the distance coordinate was defined as the radial center-of-mass distance ζ_r between monomer M11 at the protofilament tip and the incoming monomer M12 (see Figure 3). Geometrical restraints were applied to orientational and axial angles of M12 in order to fix the axial position of the distance coordinate along the protofilament axis Z and to keep the monomer in its native orientation relative to the protofilament. The axial position angles were defined as θ and ϕ in a spherical coordinate system, with the corresponding harmonic restraint potential denoted as $u_a(\theta, \phi)$. The orientational restraining potential $u_o(\Theta, \Phi, \Psi)$ was defined as a harmonic function of the three Euler angles Θ, Φ, Ψ (for details see the Methods paragraph "Definition of axial and orientational angle restraints"). Furthermore, the conformational degrees of freedom of monomer M12 were restrained to fluctuations around its native β-hairpin conformation using a distance-root-mean-square (dRMSD) coordinate ζ_c (see Methods paragraph "Conformational restraints based on distance-RMSD"). With $A\beta_{9-40}$ being intrinsically disordered, the conformational restraints played a major role in enhancing the convergence of the distance PMF $W(\zeta_r)$.

To obtain a physically correct standard binding free energy despite the presence of the restraints, the PMF difference ΔW between binding site and bulk was corrected for the free energy/entropy loss associated with the restraints in a set of separate calculations (see Figure 3). Further corrections included the transfer of M12 from standard state volume to the sampled bulk volume.

Based on the correction steps, a physically meaningful standard binding free energy could be calculated from restrained umbrella sampling using the following expression[22]::

$$\Delta G_{\text{bind}}^0 = -k_{\text{B}}T \ \ln\left(I^* S^* C^0\right) + \left(\Delta G_o^{\text{bulk}} + \Delta G_c^{\text{bulk}}\right) - \left(\Delta G_a^{\text{site}} + \Delta G_o^{\text{site}} + \Delta G_c^{\text{site}}\right). \tag{1}$$

The individual terms are discussed in the following. The main contribution I^* corresponded to a one-dimensional integral over the distance ζ_r in the binding site region, with the integrand being the Boltzmann-weighted difference of the restrained distance PMF $W(\zeta_r)$ and the PMF at a reference bulk value ζ_r^*. Note, we use the term PMF here directly for the free energy profile obtained from the WHAM analysis without a Jacobian correction which for a radial coordinate corresponds to a term proportional to ln(r) (but is very complicated to calculate for a dRMSD free energy profile).

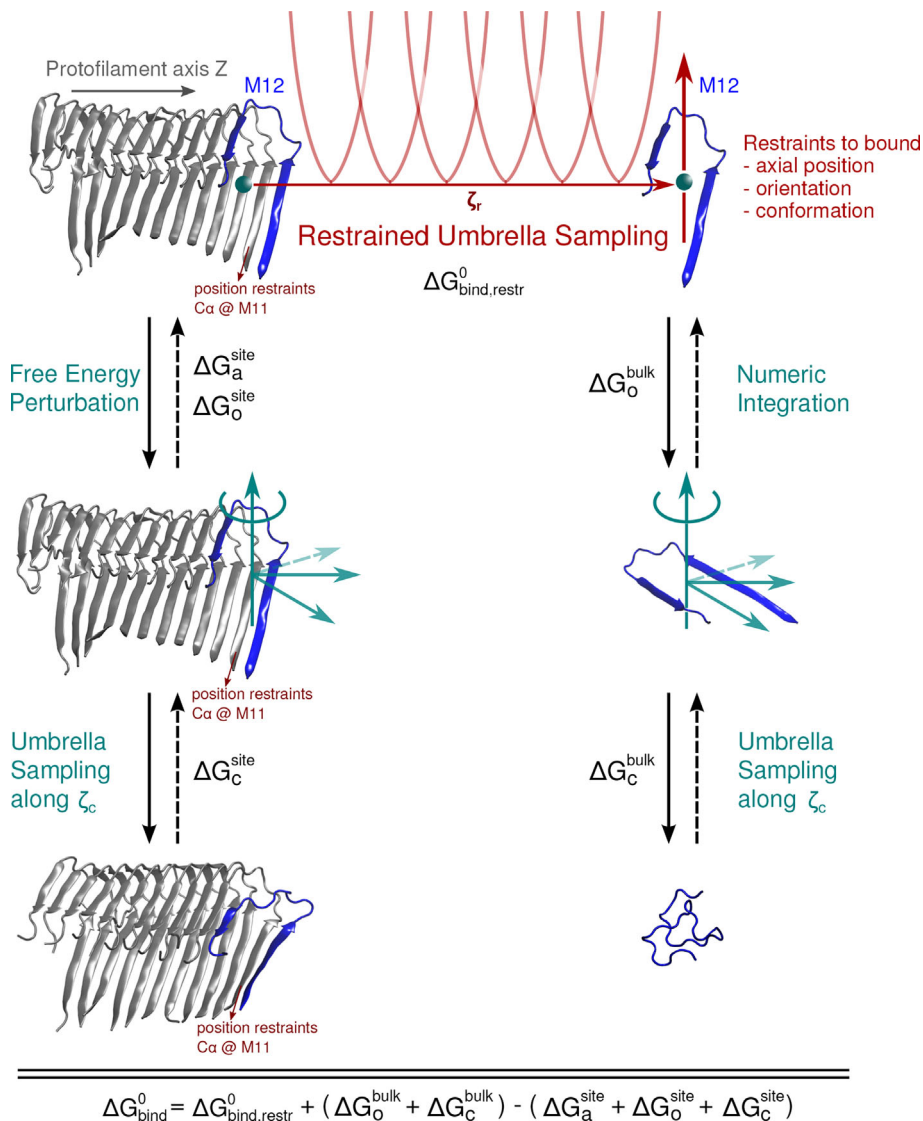$$I^* = \int_{\text{site}} d\zeta_r \ e^{-\beta\left[W(\zeta_r) - W(\zeta_r^*)\right]}, \tag{2}$$

with $\beta = 1/(k_{\text{B}}T)$. As in the case of standard umbrella sampling, the binding site region was defined using a threshold of $5k_{\text{B}}T = 3$ kcal/mol around the absolute minimum of the PMF. However, the resulting standard binding free energy was largely independent of the threshold, as the PMF $W(\zeta_r)$ steeply increased around its absolute minimum and entered I^* through a Boltzmann factor. Even a maximum threshold of $35k_{\text{B}}T = 21$ kcal/mol, which corresponded to the well depth ΔW, lead to variations of only around 0.1 kcal/mol in the standard binding free energy. The standard binding free energy was furthermore independent of the exact choice of the reference bulk radius ζ_r^* with variations in the order of 0.1 kcal/mol.

In equation 1, S^* corresponded to the spherical surface area at the reference bulk radius ζ_r^* which was accessible in the presence of the positional restraint potential $u_a(\theta, \phi)$. Due to the bulk isotropy, S^* could be calculated via numeric integration:

$$S^* = \left(\zeta_r^*\right)^2 \int_0^\pi \sin(\theta) \ d\theta \int_0^{2\pi} d\phi \ e^{-\beta u_a(\theta, \phi)}. \tag{3}$$

The expression $S^* C^0 = S^*/V_0$ in equation (1) accounted for the transfer of M12 from the standard state volume V_0 to S^*, which was comparable to the correction term ΔG_{bulk} in case of standard umbrella sampling (see equation 3 and 5 in SI section "Unrestrained binding free energy calculation using standard umbrella sampling"). Note that

FIGURE 3 Illustration of the WR advanced sampling method applied to Aβ$_{9-40}$ protofilament propagation. The application of axial, orientational and conformational restraints served to enhance the convergence of umbrella sampling, yielding a restrained distance PMF and a restrained binding free energy $\Delta G^0_{bind,restr}$ (see top row of the figure). The free energy loss associated with the restraints was accounted for in a set of correction steps both in binding site (left column) and bulk (right column). The steps could be performed bidirectional, in the form of a restraint release (down arrows) or a restraint introduction (dashed-up arrows). In the binding site, axial and orientational restraints were corrected for via free energy perturbation, whereas in bulk, their contributions were calculated numerically and the axial correction was implicitly accounted for in $\Delta G^0_{bind,restr}$. In order to correct for the conformational restraints, PMFs were calculated via umbrella sampling along a distance-RMSD coordinate ζ_c.

$$\Delta G^0_{bind} = \Delta G^0_{bind,restr} + \left(\Delta G^{bulk}_o + \Delta G^{bulk}_c\right) - \left(\Delta G^{site}_a + \Delta G^{site}_o + \Delta G^{site}_c\right)$$

the product $I^*S^*C^0$ is dimensionless and the term $-k_BT \ln\left(I^*S^*C^0\right)$ in equation 1 corresponds to a restrained standard binding free energy

$$\Delta G^0_{bind,restr} = -k_BT \ln\left(I^*S^*C^0\right). \tag{4}$$

The remaining terms in Equation (1) include free energy corrections for the axial, orientational and conformational restraints in both binding site and bulk. These corrections could be calculated in a bidirectional manner, following either a stepwise introduction or release of the restraints (see illustration in Figure 3). In the present work, we used both directions to provide insights into the convergence of the simulations. To avoid confusion about the directions and signs associated with the free energy contributions, all terms are regarded as absolute values and signs are attributed according to Equation (1).

The conformational restraints in binding site and bulk are corrected by calculating conformational PMFs in the binding site ($W^{site}_c(\zeta_c)$) and in bulk ($W^{bulk}_c(\zeta_c)$) along the dRMSD coordinate ζ_c,

using the potential $u_c(\zeta_c)$ to restrain the conformation of the monomer. The associated free energy loss was calculated from:

$$\Delta G^{site/bulk}_c = \frac{\int d\zeta_c \ e^{-\beta\left[W^{site/bulk}_c(\zeta_c)+u_c(\zeta_c)\right]}}{\int d\zeta_c \ e^{-\beta\left[W^{site/bulk}_c(\zeta_c)\right]}}. \tag{5}$$

The free energy loss associated with the stepwise introduction of axial and orientational restraints in the binding site, ΔG^{site}_a and ΔG^{site}_o, was calculated using free energy perturbation (FEP) in the presence of the conformational restraint potential $u_c(\zeta_c)$ in addition to the system's internal energy U:

$$\Delta G^{site}_o = -\frac{1}{\beta} \ln\left\langle e^{-\beta u_o(\Theta,\Phi,\Psi)}\right\rangle_{(site, \ U+u_c)} \tag{6}$$

$$\Delta G^{site}_a = -\frac{1}{\beta} \ln\left\langle e^{-\beta u_a(\theta,\phi)}\right\rangle_{(site, \ U+u_c+u_o)}. \tag{7}$$

Due to bulk isotropy, the corresponding orientational contribution in bulk, ΔG_o^{bulk}, was calculated via numeric integration analogously to S^*:

$$\Delta G_o^{bulk} = \frac{1}{8\pi^2} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \sin(\Theta) \, d\Theta \, d\Phi \, d\Psi \, e^{-\beta u_o(\Theta,\Phi,\Psi)} \quad (8)$$

The axial correction in bulk was implicitly accounted for in the transfer from standard state volume to bulk in Equation (4), as the strength of the axial restraint potential $u_a(\theta,\phi)$ determined the size of the accessible area S^*.

2.3.2 | Definition of axial and orientational angle restraints

In order to restrain the axial angles (which determine the monomer's position relative to the receptor together with the separation distance ζ_r) and orientational angles during the PMF calculations, six reference points are required: The reference centers F1, F2, and F3 are located within monomer M11 at the protofilament tip and the centers P1, P2, and P3 within the additional monomer M12 (see Figure 4A), all of which are defined as the center-of-mass of several backbone atoms and are technically implemented as virtual sites (see SI Figure S1). In case of monomer M12, the virtual sites are located within the conformationally restrained regions to ensure well-defined reference angles throughout the simulations.

The axial position of the radial center-of-mass distance coordinate ζ_r(F1-P1) was defined by the angle θ (F2-F1-P1) and the dihedral angle ϕ (F3-F2-F1-P1) (see Figure 4A). The three Euler angles, which restrained the orientation of M12 relative to the protofilament, are defined as the angle Θ (F1-P1-P2) and the dihedral angles Φ (F2-F1-P1-P2) and Ψ (F1-P1-P2-P3) (see Figure 4A). Based on the angles, the harmonic axial and orientational restraint potentials are then defined as

$$u_a(\theta,\phi) = \frac{1}{2} k_a \left[(\theta - \theta_0)^2 + (\phi - \phi_0)^2 \right] \quad (9)$$

and

$$u_o(\Theta,\Phi,\Psi) = \frac{1}{2} k_{o,1} \left[(\Theta - \Theta_0)^2 + (\Phi - \Phi_0)^2 \right] + \frac{1}{2} k_{o,2} (\Psi - \Psi_0)^2, \quad (10)$$

with force constants $k_a = k_{o,1} = 1000 \, \text{kJ}/\left(\text{mol rad}^2\right)$. A weaker force constant of $k_{o,2} = 500 \, \text{kJ}/\left(\text{mol rad}^2\right)$ was chosen for Ψ, as this angle involved all three reference points in M12 and hence resulted in higher local forces. The choice of force constants follows the magnitudes used in the original WR method[22] or was obtained from short test simulations.

Axial and orientational reference positions were chosen from a 10 ns simulation in which monomer M12 was left unrestrained in the protofilament binding site. The reference angles $\theta_0, \phi_0, \Theta_0, \Phi_0$, and Ψ_0 were set to the maximum value of the respective angle distributions (see Set 1 in Figure 4B). To evaluate the reproducibility of the resulting binding free energy, the calculation was repeated with a slightly different set of angles. The latter were chosen as the most probable angles with M12 being conformationally restrained in the binding site (see Set 2 in Figure 4B).

2.3.3 | Conformational restraints based on distance-RMSD

In order to restrain monomer M12 to its native β-hairpin conformation, a distance-based RMSD (dRMSD) coordinate ζ_c was introduced, while the original WR-method uses position-based root-mean-square deviation (RMSD).[22] Both RMSD and dRMSD are suitable parameters for measuring the average deviation of typically backbone-C_α atoms with respect to a reference conformation. In contrast to a position-based RMSD, however, the dRMSD is translationally and rotationally invariant (in house implementation). For restraining monomer M12, ζ_c was calculated with respect to the native, energy-minimized β-hairpin conformation predicted by NMR studies[23,24,26] and was defined as

$$\zeta_c = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (d_n - d_{n,0})^2}, \quad (11)$$

(B)

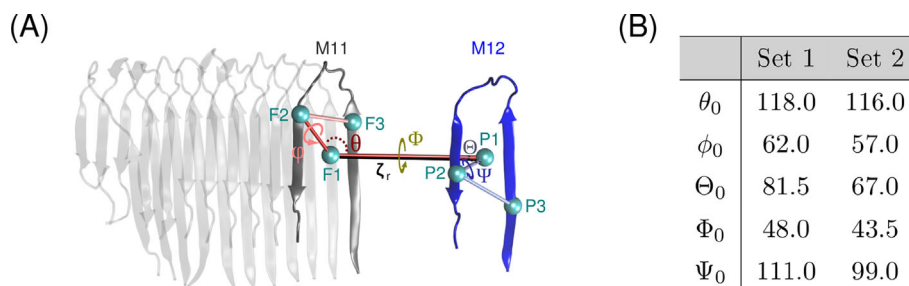|	Set 1	Set 2
θ_0	118.0	116.0
ϕ_0	62.0	57.0
Θ_0	81.5	67.0
Φ_0	48.0	43.5
Ψ_0	111.0	99.0

FIGURE 4 Definition of angles in order to restrain the axial position and orientation of monomer M12 relative to M11 at the protofilament tip. (A) Virtual interaction sites F1, 2, 3 and P1, 2, 3 were used to define the axial angles θ, ϕ and the orientational Euler angles Θ, Φ, Ψ. (B) Two different sets of reference angles were compared. Set 1 corresponded to the most probable angles of an unrestrained monomer bound to the protofilament tip. Set 2 corresponded to the most probable angles in the presence of conformational restraints.

with N denoting the number of restraint atom pairs, d_n the current distance of the nth atom pair and $d_{n,0}$ the corresponding distance in the reference structure. The corresponding harmonic restraint potential was defined as

$$u_c(\zeta_c) = \frac{1}{2}k_c(\zeta_c - \zeta_{c,0})^2. \tag{12}$$

During calculation of the distance PMF $W(\zeta_r)$, the reference dRMSD $\zeta_{c,0}$ was set to zero in order to restrain M12 to the native β-hairpin conformation. In the case of the conformational PMFs in binding site and bulk, $W^{\text{site/bulk}}(\zeta_c)$, $\zeta_{c,0}$ was varied in order to shift M12 to different regions in conformational space. In total 37 intra- and inter-strand restraint atom pairs were used, with a subset of them illustrated in Figure 5A (for a full list of atompairs and further details see SI Table S1). With a force constant of $k_c = 800$ kJ/(mol·nm²), the restraints were able to maintain the native β-hairpin conformation in solution (see Figure 5B). At the same time, the restraint potential was able to induce sufficient conformational transitions during calculation of the conformational PMFs.

2.3.4 | Potential-of-mean force along separation/distance coordinate

The protofilament's long axis Z and hence the distance coordinates $\zeta_{z/r}$ were aligned along the z-axis of the external coordinate system and fixed via position restraints on the C_α-atoms of M11. This enabled the construction of an asymmetric, rectangular simulation box extended only in z-direction, which drastically reduced the amount of explicit water molecules. The resulting box dimensions of 8.1 nm × 5.0 nm × 14.5 nm fulfilled the minimum image convention in all directions, also for the umbrella windows with maximum protofilament-monomer separation in bulk. The total system contained around 60 000 atoms.

To generate starting conformations for umbrella sampling, the monomer was pulled out of the binding site along ζ_r in steps of

0.05 nm from 0.33 to 4.0 nm with 300 ps simulation time each and constant force along ζ_r with a high force constant of $k_c = 50\,000$ kJ/(mol nm²). Meanwhile, the native conformation and orientation of M12 was maintained with force constants of $k_c = 10\,000$ kJ/(mol nm²) and $k_a = k_o = 5000$ kJ/(mol·nm²), significantly stronger than the ones used in the production simulations. With the orthogonal restraints on the relative orientation and conformation of both the protofilament and the monomer M12, the pulling approach was suitable for generating undistorted starting conformations. The resulting overlap between neighboring umbrella windows enabled the use of Hamiltonian replica exchange along the coordinate ζ_r for enhanced convergence in the production simulation. Based on the pulling conformations, 35 umbrella windows were generated, covering a range from 0.33 to 4.0 nm along ζ_r. Sufficient window overlap was achieved by a window spacing of $\Delta\zeta_r = 0.05$ nm and a force constant $k_u = 4000$ kJ/(mol·nm²) below distances of 0.93 nm, $\Delta\zeta_r = 0.1$ nm and $k_u = 800$ kJ/(mol·nm²) below 1.98 nm and $\Delta\zeta_r = 0.2$ nm below 4.0 nm. Replica exchange trials were performed every 1000 steps between adjacent windows of same force constant k_u. Compared to standard umbrella sampling, considerably shorter simulations of 20 ns per window were sufficient, with the first 5 ns being considered as equilibration.

2.3.5 | Axial and orientational corrections

The correction for axial and orientational restraints in the binding site was initiated from the umbrella endpoint at $\zeta_r = 0.48$ nm (the optimum inter-strand spacing of the fibril). The COM distance restraints were removed and the axial and orientational restraints were released via free energy perturbation (FEP). Ten intermediate FEP windows were used, corresponding to a reduction of the restraint potentials in steps of 100 kJ/(mol·rad²), and 50 kJ/(mol·rad²) for the dihedral angle Ψ. A simulation time of 0.5 ns per window was chosen. For convergence estimation, FEP calculation was also performed the other way round: Starting from an equilibrated, unrestrained monomer M12 bound to a protofilament, first the orientational then
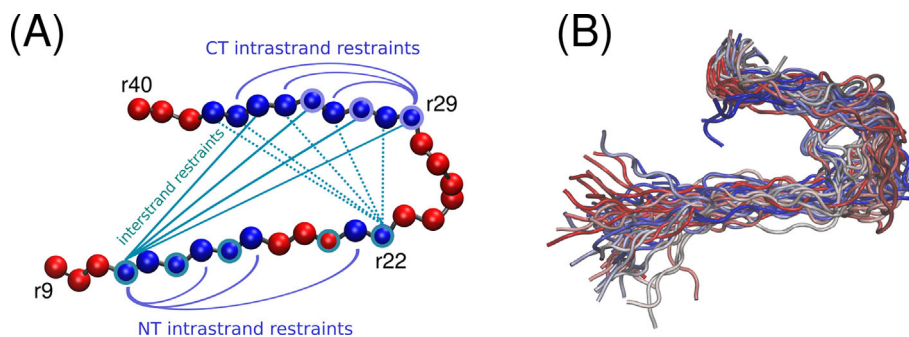

FIGURE 5 Definition of conformational distance-RMSD restraints. (A) Illustration of the combination of inter- and intra β-strand restraint atom pairs in order to maintain the Aβ$_{9-40}$ hairpin conformation. (B) The restraints were able to maintain the bound hairpin conformation in bulk solution. The snapshots resulted from a 20 ns simulation and were colored according to the simulation time, starting with blue colors and progressing to red colors.

the axial restraints were introduced, again with 10 intermediate windows each.

2.3.6 | Conformational PMFs

In order to correct for the conformational restraints on monomer M12, PMFs were calculated along the dRMSD coordinate ζ_c in binding site and bulk. The conformational PMF in the binding site was initiated from the bound monomer resulting from reverse FEP, with all axial, orientational and COM-restraints relaxed to zero. The relevant range of ζ_c was identified from unrestrained simulations of a free monomer both in bulk and binding site. Starting conformations were generated by pulling the monomer along ζ_c from 0.0 to 1.55 nm in steps of $\Delta\zeta_c = 0.05$ nm and constant force with a strong dRMSD force constant of $k_c = 5000$ kJ/mol. Each pulling step was simulated for 300 ps. In the production run, the resulting 32 umbrella windows were simulated for 30 ns each, with a weaker force constant of 800 kcal/mol for achieving sufficient window overlap. Replica exchange trials between adjacent windows were performed every 2 ps. To evaluate the convergence and quality of dRMSD restraints, a reverse PMF was initiated from the endpoint of the utmost umbrella window at $\zeta_c = 1.55$ nm. Reverse starting conformations were generated by pulling the monomer back to $\zeta_c = 0.00$ nm.

For calculating the conformational PMF of the monomer in bulk, a smaller and symmetric simulation box of 8.0 nm × 8.0 nm × 8.0 nm was constructed, which allowed to reduce the system size from 60 000 to 50 000 atoms. The box dimensions were ensured to satisfy the minimum image convention also for stretched monomer intermediates. The generation of starting conformations and the umbrella simulation setup corresponded to those in the binding site. Similarly, a reverse conformational PMF was calculated to evaluate the convergence and sampling quality of the dRMSD restraints. To compare the energetic landscape of conformations sampled with and without the presence of dRMSD restraints, an additional PMF was calculated with each window starting from a representative conformation sampled in the absence of restraints.

2.3.7 | Unrestrained simulations

Unrestrained simulations of the monomer were performed both in binding site and bulk to provide insight into favorable ranges of the coordinates ζ_z and ζ_c. This information served to validate the distance and conformational PMFs calculated during advanced sampling. In the binding site, the unrestrained monomer M12 was simulated for 100 ns at 300 K. For comparability of standard umbrella sampling and restraint umbrella sampling simulations, backbone-C_α position restraints were applied to M11 at the protofilament tip. As Aβ monomers were intrinsically disordered with a large conformational phase space in solution, the unrestrained reference distribution $\rho_{\text{free}}(\zeta_c)$ was created by combining the data from all bulk windows of standard umbrella sampling, which corresponded to 10 unfolding simulations of

1 μs at 300 K, starting from the native β-hairpin. To overcome barriers in phase space, the unfolding process in bulk was also simulated for 100 ns at elevated temperatures of 350 K and 400 K.

2.4 | PMF calculation, validation, and error estimation

PMFs were calculated using the weighted histogram analysis method (WHAM) implemented by Grossfield.[36] The tolerance parameter was set to 10^{-9}, ensuring sufficient convergence of the iterative algorithm. The course of the PMFs was tested for bias arising from the bin size, which resulted in an optimum number of 150 bins (300 for the larger data set of standard umbrella sampling). Conformational snapshots were extracted using cluster analysis[37] and corresponded to the maximum cluster representatives.

A common way of estimating statistical uncertainty of a PMF was denoted as subinterval method (SI) in the following.[17,38,39] The simulation data of each umbrella window were divided into successive, equally sized subintervals typically in the order of the equilibration time, and subset PMFs were then calculated for each data subinterval. At each bin of the PMF, upper and lower convergence estimates were determined by the largest deviation between subset PMFs and the total PMF. Due to data correlations, the resulting uncertainty was critically dependent on the exact size of the subintervals. The SI method was hence performed with two different subinterval sizes throughout this work.

The convergence of the conformational PMFs was of special focus in this work, as conformational free energy changes contributed significantly to the overall binding free energy of the intrinsically disordered Aβ peptide. To evaluate the sampling quality along the dRMSD restraint coordinate ζ_c, the PMFs in both binding site and bulk were compared to distributions obtained from unrestrained simulations, denoted as $\rho_{\text{free}}(\zeta_c)$. For further validation, the conformational PMFs were calculated both in forward and reverse direction. The forward direction corresponded to pulling the monomer M12 along ζ_c from the well-defined β-hairpin conformation at $\zeta_c = 0.0$ nm into the disordered bulk phase space at higher values of ζ_c. In contrast, sampling of the reverse path served to evaluate the capability of the dRMSD coordinate ζ_c to steer M12 from bulk into the β-hairpin conformation. A further quality control focused on the conformations of M12 sampled in the presence of dRMSD restraints: In order to estimate whether these resembled naturally occurring conformations, an additional PMF was calculated in bulk with starting structures extracted from unrestrained simulations.

In order to obtain a physically meaningful standard binding free energy, the free energy contributions of the correction steps were added to the restrained standard binding free energy (see Equation 1). Based on the uncertainties of the individual PMFs and FEP calculations, the method of Gaussian error propagation was applied to estimate an overall standard deviation corresponding to a 68%-confidence interval σ_f.[40] Let $f(x_1,...,x_N)$ denote a function of observables $x_1,...,x_N$ and $\sigma_1,...,\sigma_N$ the corresponding standard deviations. Via

error propagation, the standard deviation σ_f of the function $f(x_1,...,x_N)$ was calculated using the following expression:

$$\sigma_f = \sqrt{\sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial f(x_1,...,x_N)}{\partial x_i}\right)^2}. \quad (13)$$

The application of Gaussian error propagation required the statistical independence of the observables $x_1,...,x_N$, which was fulfilled in case of the independently calculated correction steps. Gaussian error propagation was also applied to the multi-step FEP calculations for correcting axial and orientational restraints in the binding site. In case of individual correction steps having asymmetric error ranges, Gaussian error propagation was applied separately to obtain an upper and lower error estimate for the resulting standard binding free energy.

3 | RESULTS

3.1 | Calculation of the Aβ₉₋₄₀ protofilament propagation free energy

In the following, we provide insights into the accuracy of calculated binding free energies for the growth propagation of Aβ₉₋₄₀ protofilament by monomer addition at the even (N-terminal) fibril tip. Table 1 shows a direct comparison of the absolute binding free energy

TABLE 1 Calculated standard binding free energies and contributions for (A) Standard umbrella sampling and (B) the WR-method.

(A)	Result [kcal/mol]	$\mathcal{E}_{SI}$ [kcal/mol]	T_{sim} (wind.) [ns]
ΔG_{PMF}	−10.90	±1.30	3300 (100)
ΔG_{bulk}	−2.13	—	—
ΔG_{bind}^0	−13.03	±1.30	3300 (100)
(B)	**Result [kcal/mol]**	**$\mathcal{E}_{SI}$ [kcal/mol]**	**T_{sim} (wind.) [ns]**
$\Delta G_{bind,restr}^0$	−19.30	±1.34	700 (20)
ΔG_a^{site}	+2.43	±0.13	5 (0.5)
ΔG_o^{site}	+2.45	±0.31	5 (0.5)
ΔG_c^{site}	+4.12	±0.06	960 (30)
ΔG_o^{bulk}	+6.25	—	—
ΔG_c^{bulk}	+13.23	$^{+0.67}_{-1.42}$	540 (30)
ΔG_{bind}^0	−8.82	$\sigma: {}^{+1.98}_{-1.54}$	2210
		max: $^{+2.51}_{-3.26}$	

Note: Uncertainty tolerances $\mathcal{E}_{SI}$ were calculated using the SI method. In the case of the WR-method, the overall uncertainty σ corresponded to a 68%-confidence interval derived from Gaussian error propagation. The maximum error estimate in table (B) corresponded to a maximum error accumulation of the individual contributions. Furthermore, the table listed the computational cost in terms of the simulation time associated with each free energy contribution and the simulation time per umbrella or FEP window (see round brackets).

resulting from unrestrained standard umbrella sampling and the WR-method. Table 1 shows the uncertainty estimates $\mathcal{E}_{SI}$ calculated with the subinterval (SI) method and further provides information on the computational cost associated with each free energy contribution. In case of the WR-method, the overall standard deviation σ of the standard binding free energy was calculated using Gaussian error propagation of the individual contributions, corresponding to a 68%-confidence interval.[40]

From unrestrained standard umbrella sampling, the standard binding free energy of protofilament propagation was calculated from the free energy difference between bound and unbound state ($\Delta G_{PMF} = -10.9 \pm 1.30$ kcal/mol) and the transfer from standard state volume $V^0 = 1.66$ nm³ to the sampled bulk volume of 60.75 nm³. The latter term contributed a gain in translational entropy and hence free energy of $\Delta G_{bulk} = -2.1$ kcal/mol (for details on the calculation of ΔG_{PMF} und ΔG_{bulk}, see SI section "Unrestrained binding free energy calculation using standard umbrella sampling"). In summary, the calculated binding free energy resulting from unrestrained standard umbrella sampling amounted to -13.0 ± 1.30 kcal/mol,[17] which deviates from the experimental reference of -8.6 kcal/mol for seeded Aβ₁₋₄₀ fibril propagation.[21] By contrast, the WR-method yielded close agreement with experiment with a standard binding free energy of -8.8 kcal/mol and a standard deviation of around $\sigma_{SI} = \pm 2.0$ kcal/mol. This showed that the deviation of standard umbrella sampling from the experimental reference was caused by insufficient sampling and not by shortcomings of the atomistic force field. In case of the WR-method, the resulting uncertainty σ_{SI} was comparable to values of $1 - 2$ kcal/mol obtained from advanced standard binding free energy calculations of protein–protein, protein-peptide and DNA-peptide association, which also included applications of the WR-method.[39,41] A standard deviation of 2 kcal/mol for the current system was hence considered to be reasonable.

To ensure that the calculated standard binding free energy of -8.8 ± 2.0 kcal/mol was not the result of a cancellation of errors in the individual correction steps, the complete calculation was repeated using a slightly different set of reference angles (see Figure 4B in Methods). The very similar standard binding free energy of -8.3 ± 2.1 kcal/mol supported the applicability of the WR-method to Aβ protofilament propagation (for detailed results see SI Tables S2 and S3).

3.2 | Comparison of potential-of-mean-force along separation coordinate

The distance PMFs of both sampling methods were compared in Figure 6. $W(\zeta_z)$ denoted the PMF resulting from unrestrained umbrella sampling along the separation coordinate ζ_z. $W(\zeta_r)$ denoted the restrained PMF within the WR-method. Distance distributions sampled within the umbrella windows were checked for sufficient overlap (see SI Figure S2). The course of both PMFs was classified into three regions: The binding site around the absolute minimum, a steeply increasing intermediate region, and the bulk region, where the
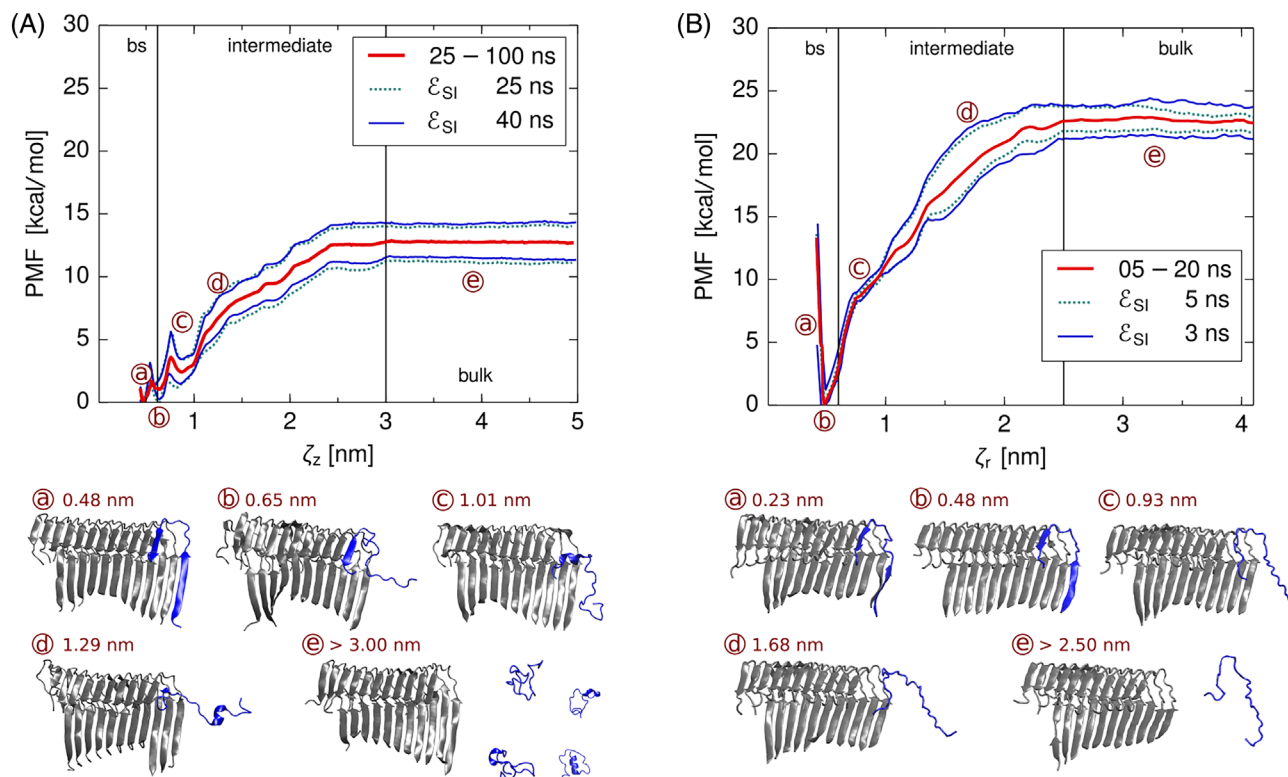
FIGURE 6 Distance PMFs $W(\zeta_z)$ and $W(\zeta_r)$ for protofilament propagation, calculated with (A) standard umbrella sampling along ζ_z ($W(\zeta_z)$) and (B) restrained umbrella sampling along ζ_r according to the WR-method ($W(\zeta_r)$). The well depth of the restrained PMF $W(\zeta_r)$ was almost twice as high due the artificial reduction of entropy orthogonal to ζ_r. For both PMFs, the absolute minimum at $\zeta_{z/r} = 0.48$ nm was in agreement with NMR predictions. Upper and lower convergence estimates (blue and dotted green curves) were calculated using the SI method by dividing the equilibrated data into subintervals of (A) 25 and 40 ns and (B) 3 and 5 ns. For each PMF, selected maximum cluster representatives (a–e) served to illustrate the differences in accessible orthogonal phase space with and without restraints.

PMFs reached a constant plateau within statistical noise. The latter indicated the absence of interactions between protofilament and monomer M12. For unrestrained standard umbrella sampling (Figure 6A), the PMF well depth was quantified to 12.8 kcal/mol. The absolute PMF minimum in the binding site indicated an optimum peptide inter-strand spacing of 0.48 nm, in good agreement with NMR studies.[23,24,26] Using the SI method, the PMF uncertainty was estimated to ±1.5 kcal/mol, with data subintervals of 25 and 40 ns length providing consistent results. Note, however, that this only corresponded to a data-inherent uncertainty, while the true uncertainty was expected to be considerably larger due to limited sampling of the large conformational phase space of the intrinsically disordered Aβ monomer M12. The large conformational variability was illustrated by maximum cluster representatives in Figure 6A, (a–e): β-hairpin conformations in the binding site region were followed by intermediate, predominantly stretched conformations, which enabled further contact to the protofilament tip. The two small barriers beyond the binding site corresponded to the detachment of first one, then both β-strands from the protofilament tip. In bulk, the monomer adopted an ensemble of intrinsically disordered, coil-like structures.

Application of the WR-method yielded a restrained distance PMF $W(\zeta_r)$ with a well depth of 22.7 kcal/mol. The latter was almost twice as high compared to $W(\zeta_z)$ resulting from standard umbrella sampling due to the artificial reduction of orientational and conformational

entropy (see Figure 6B). As in case of umbrella sampling, the absolute minimum of $W(\zeta_r)$ corresponded to the optimum inter-strand spacing of 0.48 nm.[23,24,26] This indicated that sampling of the binding site region was not severely hampered by the restraints on M12, which was an important criterion for the successful application of the WR-method. The uncertainty range of $W(\zeta_r)$ of ±1.4 kcal/mol resulting from the SI method was comparable to the one of $W(\zeta_z)$, but was achieved after only 20 instead of 100 ns simulation time per window. The faster convergence in a considerably shorter simulation time was related to the artificial reduction of orthogonal phase space. As illustrated in Figure 6B, the axial, orientational and conformational restraints on M12 allowed only small fluctuations around the native β-hairpin conformation and its orientation relative to the protofilament tip. Compared to umbrella sampling, the restraints impeded a stepwise strand detachment of the monomer away from the protofilament. This resulted in a steep increase of $W(\zeta_r)$ beyond the binding site until cooperative strand detachment occurred. As intermediate stretched conformations were also inhibited by the restraints, the bulk region of the PMF was reached earlier at 2.5 nm, compared to 3.0 nm in the case of $W(\zeta_z)$ resulting from standard umbrella sampling.

For an accurate standard binding free energy, the fast convergence of the restrained PMF $W(\zeta_r)$ was of major relevance, as the associated restrained standard binding free energy $\Delta G^0_{bind,restr} = -k_B T \ln\left(I^* S^* C^0\right) = -19.3$ kcal/mol was the largest energetic

contribution within the WR-method (see Table 1B). Calculation details for the Boltzmann-weighted integral I^* and the orthogonally accessible surface area S^* at a reference bulk radius $r^* = 3.5$ nm were provided in SI Table S4. The values of I^*, S^* and $\Delta G^0_{bind,restr}$ were furthermore confirmed to be independent of the definition of the binding site width and the bulk radius r^* (see SI Table S5).

The difference of 3.4 kcal/mol between the PMF well depth (-22.7 kcal/mol) and the restrained binding free energy $\Delta G^0_{bind,restr} = -19.3$ kcal/mol corresponded to a loss of translational entropy, as the restrained bulk volume $S^* dr$ was smaller than the standard state volume V^0. This could be seen by relating the difference of 3.4 kcal/mol to the logarithm $-k_B T \ln\left(\frac{S^* dr}{V_0}\right)$, which yielded an accessible bulk volume of $5.74 \cdot 10^{-3}$ nm^3 $\ll V^0 = 1.66$ nm^3.

3.3 | Axial and orientational restraints

The free energy loss associated with the axial and orientational restraints in the binding site was calculated using multi-step FEP according to Equations (6) and (7). In bulk, the corrections were calculated numerically from Equation (8). The axial restraints in the binding site contributed a free energy correction of $\Delta G^{site}_a = 2.43 \pm 0.13$ kcal/mol, with forward and reverse FEP yielding consistent results (for calculation details see SI Table S6). The bulk contribution of the axial restraints was implicitly

accounted for in the restrained standard binding free energy $\Delta G^0_{bind,restr} = -k_B T \ln\left(I^* S^* C^0\right)$, as the strength of the axial restraint potential determined the orthogonally accessible surface area S^* at the bulk radius r^*. The orientational free energy corrections accounted for 6.25 kcal/mol in bulk and 2.45 ± 0.31 kcal/mol in binding site, the latter being consistent for both forward and reverse FEP (for calculation details see SI Table S7). The net contribution of $\Delta G^{bulk}_o - \Delta G^{site}_o = 3.8$ kcal/mol was hence dominated by the loss of orientational entropy in bulk.

Note that ΔG^{site}_a and ΔG^{site}_o were strongly dependent on the choice of the reference restraint angles, whereas the bulk contributions were only dependent on the strength of the restraint potentials due to bulk isotropy. Using the most favorable angles of the conformationally *restrained* instead of the *unrestrained* monomer M12 in the binding site as reference (see Figure 4B) resulted in reduced contributions of $\Delta G^{site}_a = 0.69 \pm 0.05$ kcal/mol and $\Delta G^{site}_o = 1.88 \pm 0.05$ kcal/mol, in combination with a slight increase in $W(\zeta_r)$ (see SI Table S2).

3.4 | Conformational restraints

In order to correct for the conformational restraints on monomer M12, PMFs were calculated along the dRMSD coordinate ζ_c in both binding site and bulk (see Figure 7 and histograms in SI Figure S3).
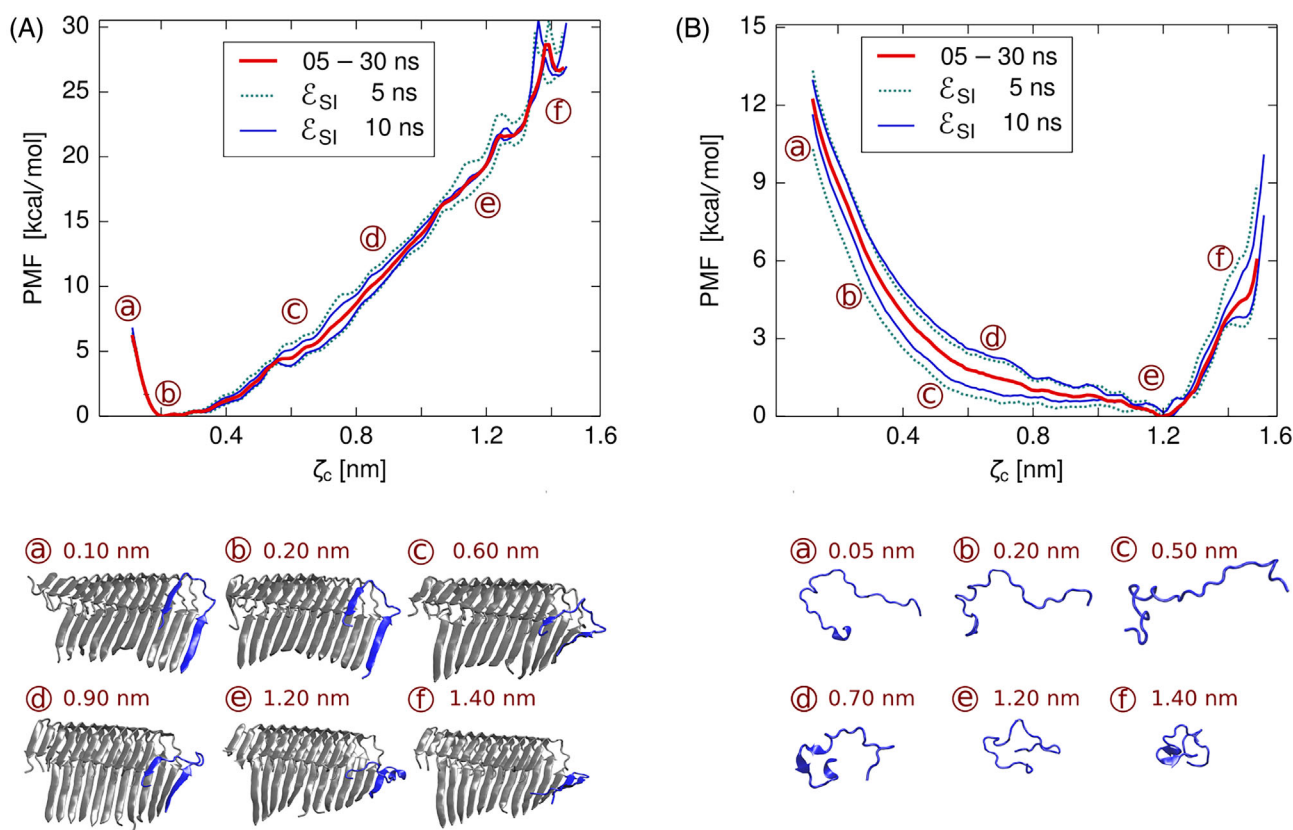

FIGURE 7 Conformational correction PMFs (red) along the dRMSD coordinate ζ_c for the monomer M12 in (A) the protofilament binding site and (B) bulk. Note the different scales of the PMF axes. Upper and lower convergence estimates (blue and dotted green curves) were calculated using the SI method by dividing the equilibrated data into subintervals of 5 and 10 ns. Selected maximum cluster representatives (a–f) for each PMF visualized the confining effect of the binding site on the monomer's phase space.

The loss of free energy associated with the introduction of restraints was calculated from Boltzmann-weighted integrals over the PMFs (see equation 5). For validating the sampling along ζ_c, the course of the PMFs was compared to unrestrained distributions $\rho_{free}(\zeta_c)$ (see Figure 7 and original histograms in SI Figure S4). In the binding site, the distributions were obtained from an unrestrained 100 ns NPT simulation, while the unrestrained conformational phase space of M12 in the unbound state was evaluated from the bulk windows of the unrestrained umbrella sampling simulations. Further convergence estimation and evaluation of the dRMSD quality was performed by calculating reverse PMFs both in binding site and bulk as well as a PMF based on freely sampled bulk conformations.

The presence of the binding site drastically confined the conformational phase space of M12 to fluctuations around the bound β-hairpin conformation. This was reflected in a deep PMF minimum around dRMSD values of $\zeta_c = 0.2 - 0.4$ nm, in overall good agreement with the free distribution $\rho_{free}(\zeta_c)$ (see Figure 7A and SI Figure S4D). The steep, approximately linear increase beyond $\zeta_c = 0.4$ nm corresponded to a deformation of the loop and inner strand regions, followed by the highly unfavorable, irreversible detachment of the shorter CT-strand. This was in agreement with the unrestrained simulation of M12 in the binding site, which was absent of an irreversible partial or complete β-strand detachment. The associated free energy correction due to conformational restraints on M12 in the binding site was a minor contribution of 4.1 kcal/mol (see Table 1).

In contrast to the binding site, the conformational PMF in bulk displayed a broad well absent of a sharply defined minimum, reflecting the large conformational phase space accessible to the intrinsically disordered Aβ peptide in bulk (see Figure 7B). Starting at $\zeta_c = 0.0$ nm, the course of the PMF resembled the unfolding process observed for Aβ hairpins in solution (see the time course of ζ_c in SI Figure S4). Hairpin conformations were highly unfavorable ($\zeta_c \approx 0.0 - 0.5$ nm). Instead, rapid transitions into transient stretched intermediates occurred ($\zeta_c \approx 0.5 - 0.8$ nm), followed by the convergence to coil-like

conformations of varying compactness. The distribution $\rho_{free}(\zeta_c)$ obtained from unrestrained simulations converged to values of $\zeta_c \approx 0.9 - 1.4$ nm with a significant maximum around $\zeta_c \approx 1.2 - 1.4$ nm (SI Figure S4), which was roughly coincident with the absolute minimum of the conformational PMF at $\zeta_c \approx 1.2$ nm. The sharp PMF increase beyond $\zeta_c \approx 1.2$ nm was possibly due to high values of ζ_c forcing the monomer into too compact structures with steric clashes (see cluster (f) in Figure 7B).

During calculation of the restrained distance PMF $W(\zeta_r)$ within the WR-method, the conformational restraints on the monomer in bulk corresponded to a shift from coil to hairpin phase space from dRMSD values of $\zeta_c \approx 1.2$ nm to $\zeta_c \approx 0.1 - 0.2$ nm. Correspondingly, this was associated with a considerable free energy loss of $13.2^{+0.7}_{-1.4}$ kcal/mol. The resulting net contribution of $\Delta G_c^{bulk} - \Delta G_c^{site} = 9.11$ kcal/mol was about half the size of the restrained standard binding free energy of -19.30 kcal/mol, indicating that conformational rearrangements were associated with a large opposing contribution of higher magnitude than the total standard binding free energy of -8.8 ± 2.0 kcal/mol.

3.5 | Validation of conformational PMFs

The overall good agreement between the conformational PMFs in binding site and bulk with the unrestrained distributions $\rho_{free}(\zeta_c)$ indicated the suitability of the dRMSD coordinate ζ_c for calculating free energies associated with conformational transitions. The conformational PMFs were calculated based on pulling the monomer from β-hairpin to coil phase space, denoted as forward direction in the following. To further evaluate the quality of ζ_c, reverse PMFs were simulated both in binding site and bulk (see blue curves in Figure 8A, B). The latter were initiated by pulling the monomer M12 from a coil conformation back to the β-hairpin phase space. To estimate whether the sampled conformations resembled those in the absence of restraints, an additional PMF was calculated with starting
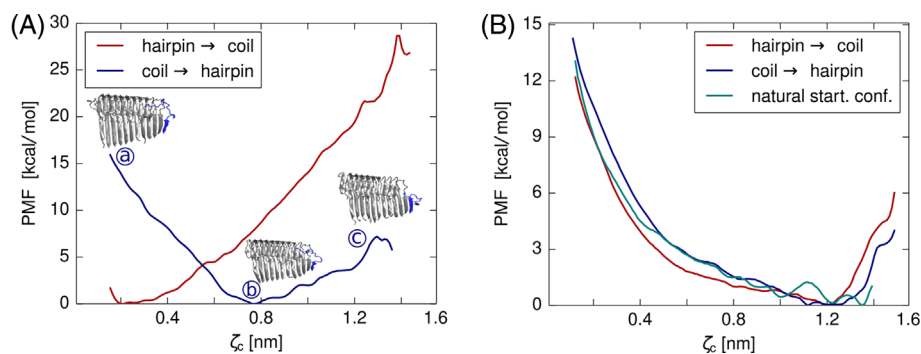

FIGURE 8 Validation of the conformational PMFs (red) by calculating reverse PMFs (blue) in (A) binding site and (B) bulk. While the original PMFs were based on pulling the monomer M12 from hairpin to coil phase space, the reverse PMFs were based on pulling M12 back to hairpin phase space. In bulk, an additional PMF (green) was calculated starting from unrestrained conformations. The reverse PMF in binding site resulted in a shifted absolute minimum, as the reverse pulling did not result in the native orientation of M12 relative to the protofilament tip, as illustrated by maximum cluster representatives (A–C). The PMFs in bulk showed overall good agreement, indicating a good sampling quality along the dRMSD coordinate ζ_c.

conformations extracted from unrestrained simulations (see green curve in Figure 8B).

In the binding site, the forward and reverse PMFs differed significantly. The reverse pulling did not result in the native orientation and CT-strand pairing of M12 relative to the protofilament tip, making low values of ζ_c highly unfavorable. However, the conformation of M12 at low values of ζ_c closely resembled a β-hairpin conformation, indicating that the dRMSD coordinate ζ_c was in principle able to steer M12 back into the β-hairpin phase space. This was further confirmed by the reverse PMF in bulk, which was in close agreement with the forward direction. The latter was particularly notable, as the reverse path lead from a disordered phase space back into the narrowly defined β-hairpin region. Furthermore, the PMF starting from naturally sampled conformations was in overall good agreement with the forward and reverse PMF in bulk, indicating the sampling of realistic conformations in the presence of dRMSD restraints.

3.6 | Minimum necessary simulation time

The unrestrained standard umbrella sampling simulations resulted in a standard binding free energy of -13.0 ± 1.3 kcal/mol, which deviated from the experimental reference of -8.6 kcal/mol.[21] These findings suggested that considerably more simulation time than 100 ns per window (3.3 μs in total) would be required to sample the large conformational phase space of the intrinsically disordered Aβ-peptide and to thereby enhance the accuracy of free energy calculation using unrestrained standard umbrella sampling.

In contrast, the WR-method achieved close agreement with experiment with a standard binding free energy of -8.8 ± 2.0 kcal/mol. Furthermore, the total simulation of around 2.2 μs was shorter compared to the standard umbrella sampling simulations. However, around 2 μs may still be computationally too expensive for systematic studies on biomedically relevant parameters of protofilament propagation such as quantifying the influence of Aβ mutations. For such applications,

however, the current simulations could serve as a priori knowledge for optimizing computational cost without loss of accuracy, as depicted in Table 2 and as discussed in the following (for calculation details see SI Table S8).

The 10 ns simulation time spent on the axial and orientational corrections could be reduced to 4 ns, although this was negligible compared to the time-consuming distance and conformational PMFs with simulation times of 700, 960, and 540 ns each. The calculation of cumulative PMFs based on an increasing percentage of the simulation data provided insight into the minimum simulation time required for convergence. As illustrated in Figure 9, all cumulative PMFs converged well within 20 or 30 ns per window, respectively, indicating that the total simulation time of 2.2 μs corresponded to an upper limit. As such, the distance PMF $W(\zeta_r)$ reached convergence already after 15 instead of 20 ns per window (see Figure 9A). Furthermore, the calculation of the constant bulk region could be reduced to a protofilament-monomer distance of around 3.2 nm without loss of statistical significance. In total, this resulted in a reduction of computational cost from 700 to 465 ns for the restrained distance PMF. The associated restrained standard binding free energy of $\Delta G_{bind,restr}^0 = -19.4 \pm 1.3$ kcal/mol was in close agreement with $\Delta G_{bind,restr}^0 = -19.3 \pm 1.3$ kcal/mol calculated from the full data set.

The introduction of conformational restraints on the monomer in the binding site only reduced fluctuations around the native β-hairpin conformation. In contrast, the highly unfavorable strand detachment beyond $\zeta_c = 0.0 - 0.5$ nm was irrelevant for the Boltzmann-weighted integrals over the conformational PMF (see Equation 5), which enabled to reduce the number of umbrella windows from 32 to only around 11. Furthermore, sufficient convergence was reached after already 15 instead of 30 ns per window (see Figure 9B). In total, this resulted in a drastic reduction of total simulation time from 960 to 150 ns with only minor loss of accuracy ($\Delta G_c^{site} = 4.07 \pm 0.17$ kcal/mol compared to 4.12 ± 0.06 kcal/mol from the full data set).

In contrast, the introduction of conformational restraints on the monomer in bulk was related to the shift from coil to hairpin phase

TABLE 2 Comparison of full and minimum simulation times for the WR-method.

	Complete data set		Minimum data set	
	[kcal/mol]	T_{sim} (wind.) [ns]	[kcal/mol]	T_{sim} (wind.) [ns]
$\Delta G_{bind,restr}^0$	-19.30	700 (20)	-19.36	465 (15)
ΔG_a^{site}	$+2.43$	5 (0.5)	$+2.43$	2.0 (0.2)
ΔG_o^{site}	$+2.45$	5 (0.5)	$+2.45$	2.0 (0.2)
ΔG_c^{site}	$+4.12$	960 (30)	$+4.07$	150 (15)
ΔG_o^{bulk}	$+6.25$	—	$+6.25$	—
ΔG_c^{bulk}	$+13.23$	540 (30)	$+12.95$	360 (20)
ΔG_{bind}^0	$-8.82^{+1.98}_{-1.54}$	2210	$-9.10^{+1.99}_{-1.69}$	979

Note: For both the complete and the minimum data set, the table lists the calculated standard binding free energies and contributions as well as the corresponding computational cost in terms of the total simulation time and simulation time per window (in round brackets). Error estimates of the individual contributions were listed in SI Table S8. The overall uncertainty σ for ΔG_{bind}^0 corresponded to a 68%-confidence interval derived from Gaussian error propagation.
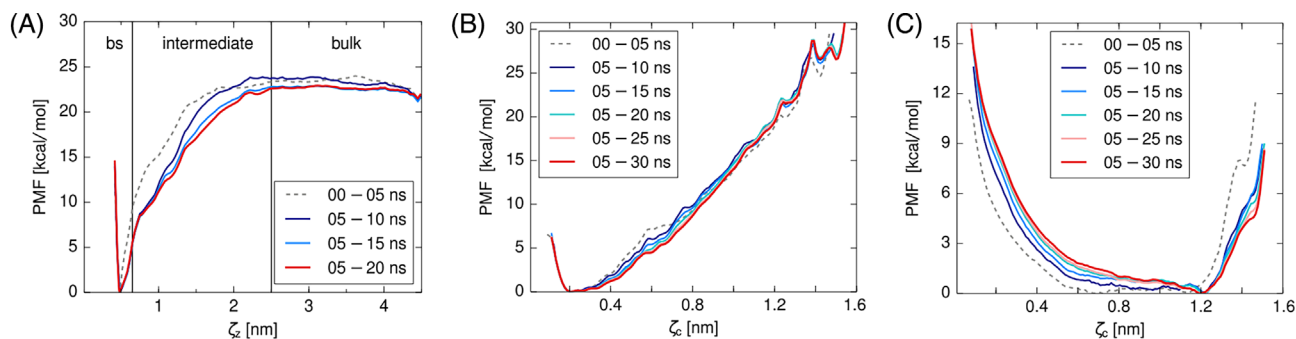
FIGURE 9 Cumulative representation of (A) the distance PMF $W(\zeta_r)$ and the conformational correction PMFs along ζ_c in (B) binding site and (C) bulk. The first 5 ns were omitted as equilibration (gray dashed curves). All PMFs showed good convergence within 20 or 30 ns simulation time per window, respectively (red curves). For future applications, the simulation time could hence be reduced to 15 ns for the distance PMF and conformational PMF in binding site (light blue curves) and 20 ns for the conformational PMF in bulk (green curve).

space and hence necessitated sampling of the full PMF range. Due to the complex conformational phase space, sufficient convergence required at least 20 ns simulation time per window (see Figure 9C). Still, compared to the previously used 30 ns per window, this corresponded to a reduction of simulation time from 540 to 360 ns, with the loss of accuracy being small compared to the overall uncertainty ($\Delta G_c^{bulk} = 12.95_{-1.59}^{+1.08}$ kcal/mol compared to $\Delta G_c^{bulk} = 13.2_{-1.4}^{+0.7}$ kcal/mol from the full data set).

In summary, the PMF for separation/association constituted the computationally most expensive part, closely followed by the conformational PMF in bulk (see Table 2). With the a priori knowledge from the present work, the total simulation time could be reduced by around half, from 2.2 μs to 1.0 μs, for future applications. The standard binding free energy and standard deviations obtained from the reduced versus the complete data set amounted to −9.1 ± 2.0 kcal/mol and −8.8 ± 2.0 kcal/mol, respectively, which was in close agreement and without relevant loss of accuracy.

4 | DISCUSSION

Accurate and efficient binding free energy calculations offer the possibility to understand the molecular mechanism of amyloid propagation and may also help to understand the influence of amino acid mutations of amyloidogenic peptides or proteins. This is especially important and of high medical relevance for the amyloid formation of Aβ-peptides that are involved in Alzheimer's disease. As a model system, we used the propagation of an Aβ$_{9-40}$ protofilament by an additional peptide binding to the filament tip for which experimental data were available.

Our results showed that the introduction of additional restraining potentials during umbrella sampling, as proposed by Woo and Roux[22], enhanced the accuracy of the calculated binding free energy compared to unrestrained standard umbrella sampling. A detailed analysis of all involved calculation steps provided further confidence in the convergence and allowed us to reduce the required simulation time to around 1.0 μs.

In addition, the restraining approach provided a separation of the total binding free energy into contributions from axial, orientational and conformational degrees of freedom. Although these contributions depend on the choice of reference states (e.g., the reference angles in the axial and orientational degrees of freedom) and force constants, it is still possible to discuss the corresponding free energy contributions. For example, we found small contributions of $1-3$ kcal/mol due to restrictions in orientational and axial degrees of freedom upon protofilament propagation, depending on the choice of restraint force constants and the reference angles in the bound state. Changes in these reference angles modulated both the corresponding correction terms and the distance PMF, that is, the separation contribution, to the binding free energy while the conformational free energy contribution was not directly affected.

In our calculations, the change of the conformational free energy to restrict the ensemble of unbound, intrinsically disordered Aβ$_{9-40}$ peptides to a narrow ensemble around the β-hairpin conformation of the bound state made up the largest contribution that opposed Aβ$_{9-40}$ binding and growth propagation. For the calculation of the conformational free energy contribution it is often difficult to achieve convergence. However, especially in the bulk we found in the present case good agreement between forward and backward free energy simulations. Its overall contribution (9.1 kcal/mol) was larger than the total Aβ$_{9-40}$ binding free energy (8.8 kcal/mol) and almost half of the separation free energy contribution (19.3 kcal/mol).

As a disadvantage of the restraining approach, the introduction of additional restraining potentials lead to an artificial sampling pathway, that does not represent a realistic dissociation or association process during protofilament propagation. Here, unrestrained umbrella sampling provided more suitable insight into conformational transitions and molecular interactions.[17] However, much longer simulation times of 3.3 μs were required to sample the large conformational space of the intrinsically disordered Aβ monomer in each umbrella window. Even with these extensive simulation times, however, insufficient sampling lead to deviations of experimental and simulated binding free energies (−8.6 kcal/mol vs. −13.0 ± 1.3 kcal/mol).

Note, an additional drawback of the geometric restraining approach is the possible sterical overlap between partners along a selected separation coordinate and direction. We recently introduced the possibility to reduce such sterical problems by adding a softcore scaling of overlapping segments during intermediate separation steps with the segments fully switched on at the beginning and endpoint of the separation.[42] This approach was shown to give rapidly converging PMF calculations along various different separation directions of binding partners even involving significant overlap.

Future applications of the WR-approach may include the quantitative study of $A\beta$ isoforms and mutations as well as the influence of salt, denaturants and inhibitors on the propagation step. Of particular interest is the comparison between $A\beta_{40}$ and the more toxic isoform $A\beta_{42}$, which is known to possess higher aggregation and oligomerization propensity.[43–46] Within the statistical uncertainty of around 2.0 kcal/mol, the quantification of mutations with a significant stabilizing or destabilizing effect may be practicable, such as the replacement of hydrophobic contacts by charged residues in the hydrophobic protofilament core.[47] Further studies may also include mutations with significant conformational impact on the bulk ensemble of the intrinsically disordered $A\beta$ peptide. Examples are well-known pathogenic mutations with a stabilizing effect on the bend in the loop region or hydrophobic mutations with enhanced β-sheet propensity.[48]

AUTHOR CONTRIBUTIONS

Christina V. Frost: Investigation; writing – original draft; visualization; validation; methodology; formal analysis; data curation; software. **Nadine Schwierz:** Conceptualization; writing – review and editing; validation; supervision; data curation. **Martin Zacharias:** Resources; supervision; project administration; writing – review and editing; validation; writing – original draft; conceptualization; funding acquisition.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26683.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Martin Zacharias https://orcid.org/0000-0001-5163-2663

REFERENCES

1. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*. 2006;75:333-366.
2. Fändrich M, Nyström S, Nilsson KPR, Böckmann A, LeVine H, Hammarström P. Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. *J Intern Med*. 2018;283(3):218-237.
3. Sawaya MR, Hughes MP, Rodriguez JA, Riek R, Eisenberg DS. The expanding amyloid family: structure, stability, function, and pathogenesis. *Cell*. 2021;184(19):4857-4873.
4. Eisenberg DS, Sawaya MR. Structural studies of amyloid proteins at the molecular level. *Annu Rev Biochem*. 2017;86:69-95.
5. Iadanza MG, Jackson MP, Hewitt EW, Ranson NA, Radford SE. A new era for understanding amyloid structures and disease. *Nat Rev Mol Cell Biol*. 2018;19(12):755-773.
6. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med*. 2016;8(6):595-608.
7. Xiaoguang D, Wang X, Geng M. Alzheimer's disease hypothesis and related therapies. *Transl Neurodegen*. 2018;7(1):2.
8. Cummings J. Lessons learned from Alzheimer disease: clinical trials with negative outcomes. *Clin Transl Sci*. 2018;11(2):147-152.
9. Folch J, Ettcheto M, Petrov D, et al. Review of the advances in treatment for Alzheimer disease: strategies for combating β-amyloid protein. *Neurología*. 2017;33(1):47-58.
10. Leide Caroline Picanco dos Santos, Priscilla F Ozela, Maiara de Fatima de Brito Brito, Abraao A Pinheiro, Elias C Padilha, Francinaldo S Braga, Carlos Henrique Tomich de Paula da Silva, Cleydson Breno Rodrigues dos Santos, Joaquin M C Rosa, and Lorane Izabel da Silva Hage-Melim. Alzheimer's disease: a review from the pathophysiology to diagnosis, new perspectives for pharmacological treatment. *Curr Med Chem*, 25(26):3141–3159, 2018.
11. Tycko R. Physical and structural basis for polymorphism in amyloid fibrils. *Protein Sci*. 2014;23(11):1528-1539.
12. Fändrich M, Schmidt M, Grigorieff N. Recent progress in understanding Alzheimer's β-amyloid structures. *Trends Biochem Sci*. 2011;36(6):338-345.
13. Fändrich M, Schmidt M. Methods to study the structure of misfolded protein states in systemic amyloidosis. *Biochem Soc Trans*. 2021;49(2):977-985.
14. Scheres SHW, Zhang W, Falcon B, Goedert M. Cryo-EM structures of tau filaments. *Curr Opin Struct Biol*. 2020;64:17-25.
15. Owen MC, Gnutt D, Gao M, et al. Effects of in vivo conditions on amyloid aggregation. *Chem Soc Rev*. 2019;48(14):3946-3996.
16. Barz B, Liao Q, Strodel B. Pathways of amyloid-β aggregation depend on oligomer shape. *J Am Chem Soc*. 2018;140(1):319-327.
17. Schwierz N, Frost CV, Geissler PL, Zacharias M. Dynamics of seeded aβ40-fibril growth from atomistic molecular dynamics simulations: kinetic trapping and reduced water mobility in the locking step. *J Am Chem Soc*. 2016;138(2):527-539.
18. Han W, Schulten K. Fibril elongation by aβ17–42: kinetic network analysis of hybrid-resolution molecular dynamics simulations. *J Am Chem Soc*. 2014;136(35):12450-12460.
19. Schwierz N, Frost CV, Geissler PL, Zacharias M. From aβ filament to fibril: molecular mechanism of surface-activated secondary nucleation from all-atom MD simulations. *J Phys Chem B*. 2017;121(4):671-682.
20. Sasmal S, Schwierz N, Head-Gordon T. Mechanism of nucleation and growth of aβ40 fibrils from all-atom and coarse-grained simulations. *J Phys Chem B*. 2016;120(47):12088-12097.
21. O'Nuallain B, Shivaprasad S, Kheterpal I, Wetzel R. Thermodynamics of aβ (1–40) amyloid fibril elongation. *Biochemistry*. 2005;44(38):12709-12718.
22. Woo H-J, Roux B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc Natl Acad Sci*. 2005;102(19):6825-6830.
23. Aneta TP, Ishii Y, Balbach JJ, et al. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci*. 2002;99(26):16742-16747.

24. Petkova AT, Yau W-M, Tycko R. Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry*. 2006; 45(2):498-512.

25. Buchete N-V, Tycko R, Hummer G. Molecular dynamics simulations of Alzheimer's β-amyloid protofilaments. *J Mol Biol*. 2005;353(4): 804-821.

26. Sciarretta KL, Gordon DJ, Petkova AT, Tycko R, Meredith SC. Aβ40-lactam (d23/k28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry*. 2005;44(16):6003-6014.

27. Buchete N-V, Hummer G. Structure and dynamics of parallel β-sheets, hydrophobic core, and loops in Alzheimer's aβ fibrils. *Biophys J*. 2007; 92(9):3032-3039.

28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(2):926-935.

29. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26(16):1701-1718.

30. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19): 8577-8593.

31. Hess B. P-lincs a parallel linear constraint solver for molecular simulation. *J Chem Theory Comput*. 2008;4(1):116-122.

32. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984;81(8):3684-3690.

33. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys*. 1984;81(1):511-519.

34. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A*. 1985;31(3):1695-1697.

35. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys*. 1981;52(12):7182-7190.

36. Grossfield A. WHAM: The weighted histogram analysis method, version 2.0. 9 (Accessed November 15, 2013). 2013. https://membrane.urmc.rochester.edu/content/wham.

37. Shao J, Tanner SW, Thompson N, Cheatham TE. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J Chem Theory Comput*. 2007;3(6): 2312-2334.

38. Zeller F, Zacharias M. Evaluation of generalized born model accuracy for absolute binding free energy calculations. *J Phys Chem B*. 2014; 118(27):7467-7474.

39. Ulucan O, Jaitly T, Helms V. Energetics of hydrophilic protein–protein association and the role of water. *J Chem Theory Comput*. 2014;10(8): 3512-3524.

40. Demtröder W. *Mechanics and Thermodynamics*. Springer; 2017.

41. Gumbart JC, Roux B, Chipot C. Standard binding free energies from computer simulations: what is the best strategy? *J Chem Theory Comput*. 2013;9(1):794-802.

42. Reif MM, Zacharias M. Improving the potential of mean force and nonequilibrium pulling simulations by simultaneous alchemical modifications. *J Chem Theory Comput*. 2022;18(6):3873-3893.

43. Dong S, Duan Y, Yinghe H, Zhao Z. Advances in the pathogenesis of Alzheimer's disease: a re-evaluation of amyloid cascade hypothesis. *Transl Neurodegen*. 2012;1(1):1-12.

44. Haass C, Selkoe DJ. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nat Rev Mol Cell Biol*. 2007;8(2):101-112.

45. McGeer PL, McGeer EG. The amyloid cascade-inflammatory hypothesis of Alzheimer disease: implications for therapy. *Acta Neuropathol*. 2013;126(4):479-497.

46. Straub JE, Thirumalai D. Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu Rev Phys Chem*. 2011;62(1):437-463.

47. Adler J, Scheidt HA, Krüger M, Thomas L, Huster D. Local interactions influence the fibrillation kinetics, structure and dynamics of aβ(1–40) but leave the general fibril structure unchanged. *Phys Chem Chem Phys*. 2014;16:7461-7471.

48. Chong S-H, Yim J, Ham S. Structural heterogeneity in familial Alzheimer's disease mutants of amyloid-beta peptides. *Mol Biosyst*. 2013; 9(5):997-1003.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.