



Article

# A Hybrid Hamiltonian for the Accelerated Sampling along Experimental Restraints

Emanuel K. Peter and Jiří Černý \*

Institute of Biotechnology of the Czech Academy of Sciences, BIOCEV, Průmyslová 595,  
252 50 Vestec, Prague West, Czech Republic; e.peter@fz-juelich.de

\* Correspondence: jiri.cerny@ibt.cas.cz; Tel.: +420-325-873-738

Received: 17 December 2018; Accepted: 10 January 2019; Published: 16 January 2019



**Abstract:** In this article, we present an enhanced sampling method based on a hybrid Hamiltonian which combines experimental distance restraints with a bias dependent from multiple path-dependent variables. This simulation method determines the bias-coordinates *on the fly* and does not require *a priori* knowledge about reaction coordinates. The hybrid Hamiltonian accelerates the sampling of proteins, and, combined with experimental distance information, the technique considers the restraints adaptively and in dependency of the system's intrinsic dynamics. We validate the methodology on the dipole relaxation of two water models and the conformational landscape of dialanine. Using experimental NMR-restraint data, we explore the folding landscape of the TrpCage mini-protein and in a second example apply distance restraints from chemical crosslinking/mass spectrometry experiments for the sampling of the conformation space of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A). The new methodology has the potential to adaptively introduce experimental restraints without affecting the conformational space of the system along an ergodic trajectory. Since only a limited number of input- and no-order parameters are required for the setup of the simulation, the method is broadly applicable and has the potential to be combined with coarse-graining methods.

**Keywords:** enhanced molecular dynamics simulations; protein folding

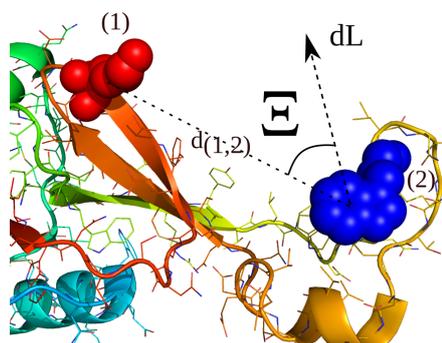
## 1. Introduction

Proteins fold into a unique three-dimensional structure after they are translated by the Ribosome. They execute their function through a variety of conformational transitions upon changes in their chemical environment, such as the binding of small molecules, changes in the electrolyte concentration, or the presence of interacting proteins in biological processes [1–3]. A number of experimental techniques, namely X-ray crystallography, solution NMR, cryo-electron microscopy, and structural mass-spectrometry have been developed to investigate the structure and the conformational changes of proteins [4–6]. Molecular Dynamics simulations (MD) [7] are a valuable tool for complementing the experimental results and to elucidate time-resolved atomistic pictures of configuration spaces of proteins upon mutation or the binding to a specific ligand [8]. Parallelization and the development of efficient algorithms improved the performance of MD by many orders of magnitude [9–12]. The MD-method largely contributed to the current understanding of protein dynamics and the conformational space in a vast amount of processes. At the same time, enhanced sampling methods, namely accelerated MD [13], Gō-modeling [14,15], Metadynamics [16–20], Monte-Carlo methods [21,22], umbrella sampling, and related techniques such as local elevation [23–28], replica exchange MD [29–31], replica exchange MD combined with umbrella sampling or other improvements in the sampling [32–46], transition path sampling techniques, and statistical sampling methods such as Milestoning [47–50], accelerated adaptive

path-sampling methods [51,52], accelerated methods based on path-integrals [53,54], multiple time-stepping techniques [44,55,56], self consistent field MD approaches [57], enhanced sampling combined with machine learning [58–60], and coarse-graining techniques [61–69] improved the efficiency of MD by many orders of magnitude. Already at the rise of the MD-method as an important complementary tool, a number of developments have been made to connect the MD-approach with experimental information [70]. A number of early developments used NMR-restraints for the structural determination of proteins from solution NMR-NOESY data [71,72] and NMR-shifts [73], which also has been implemented into enhanced sampling methodologies [74–78]. Alternative restraint techniques have been developed to sample proteins from X-ray small angle scattering experimental data [79–81] and cryo-electron microscopy data [82]. A more detailed development evaluated the influence of the applied bias-potential and its impact on the sampling of the underlying NMR-ensemble, which in fact is a time-averaged dataset representing an ensemble of conformations [83–87], which also appeared for biasing methodologies for the simulation of distance-restraints determined in chemical crosslinking experiments combined with mass-spectrometry [87–89].

In two recent works, we introduced a renormalization formalism leading to a hybrid Hamiltonian  $H(C)$ , which introduces additional biasing Hamiltonians  $H(B)$  to enhance the sampling in simulations of protein folding and aggregation [90,91]. The novel hybrid Hamiltonian  $H(C)$  accelerates the sampling of the system and can shift the resulting statistical averages in the partition function [91]. To bias the simulation along  $\nabla H(B)$ , we developed a technique which uses a history-dependent path-definition  $L$  resulting in a biased action integral  $L_s$  and a biased increment  $dL_s(t)$  [92]. The enhanced sampling methodology accelerates the sampling of a system, while a minimal set of input parameters, no *a priori* information about reaction pathways or product states is required and the ergodicity of the dynamics is guaranteed. We extend the definition of  $\nabla H(B)$  to the sampling along multiple biasing increments at multiple time-scales to capture the dynamical heterogeneity of the systems of interest. We then implemented a second adaptive methodology to the restrained sampling along given experimental distance information from NMR-NOESY and chemical crosslinking/mass-spectrometry experiments, which applies the bias indirectly in the form of an overlapping fraction of the biasing increments with the experimental restraint. For a schematic representation see Figure 1. In specific cases, the combination of the adaptive biasing method along pathways  $L$  and the adaptive reweighting to sampling along given experimental information can achieve a better convergence to the underlying statistical average than the application of restraints in the potential energy space, while a minimal set of parameters *a priori* to the simulation is required. The novel approach also has the advantage that the restrained underlying partition is only dependent on the used parameter set, which is a strong difference to methods, where restraints are applied in the energy space.

We validate the parallel biasing technique on the dipole relaxation of two water models and the conformational landscape of dialanine. As a first application of the novel restrained sampling method, we use the method in folding simulations of TrpCage, where we apply the NMR-data from ref. [93] to sample the conformation ensemble of this protein in explicit and implicit solvent. As a second example, we apply the technique on the conformation ensemble of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A), where we use restraint information obtained in a combined chemical cross-linking and mass-spectrometry experiment [89]. Because no order parameters are required for the setup of the simulation as in umbrella sampling techniques [94], the restrained path-sampling method is broadly applicable and can also be used in combination with systematic coarse-graining techniques [95,96].



**Figure 1.** Schematic restraint vector  $r$ , an instantaneous path-increment  $dL$  and the corresponding angle  $\Xi$ , which re-weights the adaptive bias coordinate towards an experimental distance information  $d_{12}$  between a pair of amino-acids 1 and 2. In this article, we present a method for the coupling of experimental distance restraints along distances between 2 atoms ( $d_{12}$ ) and multiple path-dependent biasing increments  $\nabla H(B)$ . We define a novel restrained biasing Hamiltonian  $\nabla H(B)_R$  as the overlapping fraction of the path-dependent bias in the dynamical trajectory space with the experimental distance vector. We re-evaluate the un-biased Hamiltonian  $H(A)$  using a renormalization technique, which leads to an accelerating hybrid Hamiltonian  $H(C)$ .

## 2. Results and Discussion

### 2.1. Influence of the Renormalization Parameters on the Time-Dependent Relaxation Behavior in Simulations along Multiple Biasing Increments $dL_{ik}$

We discuss the effect of the multiple renormalized biases and the renormalization parameters  $\alpha_{md}$  and  $\alpha'$  on the dynamical relaxation behavior of a time-dependent quantity  $X$  describing a system, such as the relaxation of the time-dependent total dipole moment  $\mathbf{M}(t)$  of a water system. Any quantity  $X(t)$  in an unbiased simulation follows a time-correlation function  $A(X)(t)$ , which can be described by an expansion to  $M$  monoexponential decay processes with periods  $\tau_m$ , which is defined by the time-behavior of a quantity  $X$ , rate constants  $k_m$ , and prefactors  $A_{m_0}$ :

$$A(X)(t) = \sum_m^M A_{m_0} \exp(-k_m t). \quad (1)$$

In contrast to the equilibrium MD case, a bias coupled to a set of  $N_R$  biases to  $N$  atoms, with  $\tau_{ik} = \tau_{1_{ik}}$  and  $\tau_{ik} = \tau_{2_{ik}}$  leading to an actual acceleration in terms of a change in the time-correlation function. That results in a modified relaxation behavior, affecting all dynamical quantities (that can lead to accelerated folding times, modified diffusion constants, and re-orientation Kinetics of H-bonds or dipoles in the system. There is also an effect on quantities such as the static and dynamic dielectric properties), which we write as an heuristic equation (as described in our simulation results of the dielectric response of SPC/E and TIP3P water):

$$A_m(X)(t) = \sum_m^M \left( A'_{i_0} \sum_i^{N_R} A'_{im_0} \right) \times \left( \exp \left( - \left( k'_m + \sum_i^{N_R} k_{im} \right) t_{rel} \right) \right), \quad (2)$$

where  $N_R$  stands for the number of renormalized biases and  $k_{i_R}$  is the rate-constant within each bias with index  $l$ , and we note that the time  $t$  changes to a relative time  $t_{rel}$ , which scales linear with an acceleration factor  $\rho > 1$ :  $t_{rel} = \rho t$ , as long as the coupling to the underlying un-biased gradient remains sufficiently low ( $\beta'_{ik_{MD}} = \beta'_{ik} \leq 1 \times 10^{-4}$ ,  $\eta'_{md} = \eta' \approx 25$ ; see Section: Methods, E: 'Renormalization

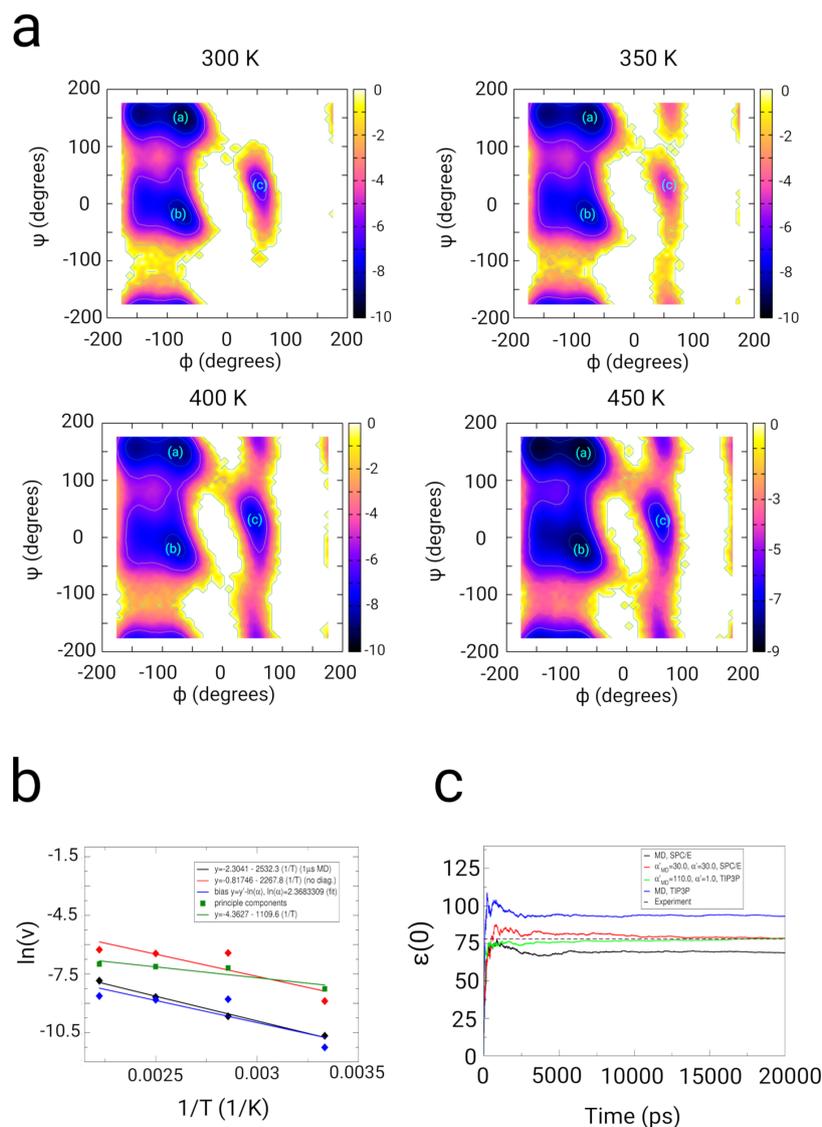
to the underlying Hamiltonian') in the enhanced sampling simulation. The relation (2) shows that depending on the magnitude and the parameters ( $\beta'_{md}$ ,  $\beta'$ ,  $\eta'$ ,  $\eta'_{md}$ ,  $N_R$ , and the coupling times  $\tau_1$ ,  $\tau_2$ ), the Hamiltonian of the system and the time-correlation behavior in the simulation are modified, while the processes depending from the parameters  $A$  and rate constants  $k$  still are described by modified monoexponential time-dependent decays, since the renormalization and the conditions on *adaptive bias MD* obey the principle of action as described in the Equation (19). That way, dynamical quantities such as dielectric quantities related to dipole fluctuations and in general fluctuation-dependent properties related to a linear response of the system can effectively be varied through the choice of the bias-parameters. Finally, we mention the boundary case of an infinite number of biases  $N_R$ . As stated in the Equation (1), the time-dependent property  $A(X)(t)$  will not converge within an imaginary infinite sampling period, if an infinite number of applied biases  $s$  would be used for this simulation. Thus, the number of applied biases has to be comparatively low in the range from 1 to  $N_R = 20$  biases. We validated the influence of the algorithm in simulations of TIP3P and SPCE/E water, where we measured the impact of the fluctuation dependent parameters on the dielectric properties of water in comparison with experiments [97–101] (see supplementary information: Section 2: Water simulations. Supplementary Tables S1 and S2, Figures S1 and S2). Using suitable parameters for  $\eta'_{md}$  and  $\eta'$ , we successfully could change the dielectric constant of SPC/E (MD:  $\epsilon(0) \approx 70$ ) and TIP3P water (MD:  $\epsilon(0) \approx 100$ ) to the value observed in experiments at approximately 78 (see Figure 2c). Using suitable parameters for  $\tau_1$  (adaptive bias MD),  $\tau_2$  (path-sampling),  $\eta'_{md}$  and  $\eta'$ , the dynamic relaxation of the dipole moment of water could also be driven towards the experimental dielectric relaxation spectra.

## 2.2. Dialanine

### 2.2.1. Results

We used dialanine as a test system to examine the effect of the algorithm on a simple and small peptide. Dialanine has emerged as a prominent system for the validation of enhanced sampling algorithms [16,102,103] and has been used to study the kinetics of transitions along the reactive coordinates [47,104,105] (the dihedral angles  $\Phi$ ,  $\Psi$ , and  $\chi$ ). In validation simulations of dialanine, we applied the principal components of the multiple pathways derived from orthogonal vectors  $dL_{\perp ik}$  to the sampling of dialanine (see Section: Methods). In the simulation at 300 K using the optimized sampling scheme along principal components, we observe an improved partition of the population densities in the partition at  $-50^\circ < \Psi < 50^\circ$ ,  $-100^\circ < \Phi < 0^\circ$  ( $-8 k_B T$ ) (minimum (b)),  $100^\circ < \Psi < 180^\circ$ ,  $-150^\circ < \Phi < 0^\circ$  ( $-9$  to  $-10 k_B T$ ) (minimum (a)), and the population at  $0^\circ < \Psi < 100^\circ$ ,  $-50^\circ < \Phi < 100^\circ$  (minimum (c)) (see Figure 2a). We compared MD- (1  $\mu$ s) and adaptive path-sampling simulations (50 ns) (path-sampling simulations without principal components, path-sampling simulations using principal components) at 4 temperatures ranging from 300 to 450 K. We analyzed the transition frequency of the  $\Phi$ -angle of dialanine as function of time and determined the escape frequency as function of the inverse temperature (see Figure 2b) [90]. In the path sampling simulations without the use of principal components, we find transition times  $\tau_\Phi$ : at T = 300 K  $\tau_\Phi = 7228$  ps, at T = 350 K  $\tau_\Phi = 614$  ps, T = 400 K  $\tau_\Phi = 634$  ps, and at T = 450 K  $\tau_\Phi = 522$  ps. The biased simulations using principal components results in transition times  $\tau_\Phi$ : at T = 300 K  $\tau_\Phi = 3873$  ps, at T = 350 K  $\tau_\Phi = 1325$  ps, T = 400 K  $\tau_\Phi = 1236$  ps, and at T = 450 K  $\tau_\Phi = 1083$  ps. The MD-results averaged over 1  $\mu$ s yields transition times at T=300 K  $\tau_\Phi = 42696$  ps, T = 350 K  $\tau_\Phi = 15759$  ps, T = 400 K  $\tau_\Phi = 5877$  ps, and T = 450 K  $\tau_\Phi = 2550$  ps [90]. From the linear regression along the 3 data-sets, we determine activation energies  $E_a$  of 21.05 kJ/mol for MD,  $E_a$  equal to 18.77 kJ/mol for the biased simulations and  $E_a$  equal to 9.2 kJ/mol in the biased simulation using the principal components (see Section 3). In the case of the accelerated simulations along principal components, we observe a shift in the transition frequencies at higher temperatures in contrast to standard MD and path-sampling leading to a shift in the value of  $E_a$ . At the same time, we yield an approximately identical partition

along the angles  $\Phi$  and  $\Psi$  in the case of the sampling along principal components. From the absolute shift of the transition frequencies, we determined an absolute acceleration factor  $\ln(\rho)$  for this transition equal to 2.39, leading to an acceleration factor  $\rho$  equal to 10.68 by which the biased simulation is faster than equilibrium MD.



**Figure 2.** Results from simulations of dialanine peptide in SPC/E water used for the validation of the sampling along  $N_R$  multiple biasing increments  $\nabla H(B)_{ab}$  ( $dL_{ab_{ik}}(t)$ , adaptive bias MD) and  $\nabla H(B)_{\sigma}$  ( $dL_{\sigma_{ik}}(L_{ik}(t))$ , path-sampling) to accelerate the sampling, which results in a biased action integral  $L_S$  and a modification of the un-biased Hamiltonian  $H(A)$ , which results in a hybrid Hamiltonian  $H(C)$ . (a) Free energy landscapes as function of backbone dihedral angles  $\Phi$  and  $\Psi$  from simulations at the temperatures 300 K, 350 K, 400 K, and 450 K using multiple path sampling in combination with optimization techniques (using principal components of the biasing Hamiltonian  $H(B)$ ), a number of  $N_R = 10$  biases,  $\beta'_{md} = \beta' = 10^{-3}$ ,  $\tau_1 = 0.5$  ps, and  $\tau_2 = 0.1$  ps. Energy values on the color bar are in units of  $k_B T$ . (b) Comparison of the logarithm of the transition frequency of the dihedral angle  $\Phi$  as function of the inverse temperature (300–450 K) in the biased (red and green curve) and 1  $\mu$ s MD-trajectories (black curve) (blue curve, fit of the biased data to the MD-data for the determination of the linear scaling factor  $\ln(\rho)$ ). In this Kinetic analysis, we determined an acceleration factor  $\rho$  equal to 10.68 by which the algorithm accelerates the sampling of dialanine peptide. (c) Static permittivities  $\epsilon(0)$  as function of MD-simulation time using different coupling  $\alpha'$  and  $\alpha_{md}$ -parameters in the individual simulations of SPC/E and TIP3P water compared with conventional MD simulations.

### 2.2.2. Discussion

In our recent work [90], we compared the quality of the sampling using the adaptive path MD, the path-sampling and the hybrid algorithm and found that the partition along the  $\Phi$  and the  $\Psi$ -angle from the 1  $\mu$ s with the hybrid- and the path-sampling simulation deviated by approximately  $\leq 1 k_B T$  in the high energy regions, corresponding to 12.5% in relation to the absolute scale of the free energy difference  $\Delta F$ . With the extended hybrid sampling methodology based on multiple pathways  $dL_{ik}$ , the renormalization and the propagation along principle components, we obtain the correct partition in the FEL along  $\Phi$  and  $\Psi$  with approximately identical free energy values. We also obtain quantitatively correct variations in the transition frequencies of the  $\Phi$ -angle in comparison with 1  $\mu$ s MD simulations [92].

### 2.3. Simulations with Restraints from NMR NOESY spectra: TrpCage Folding

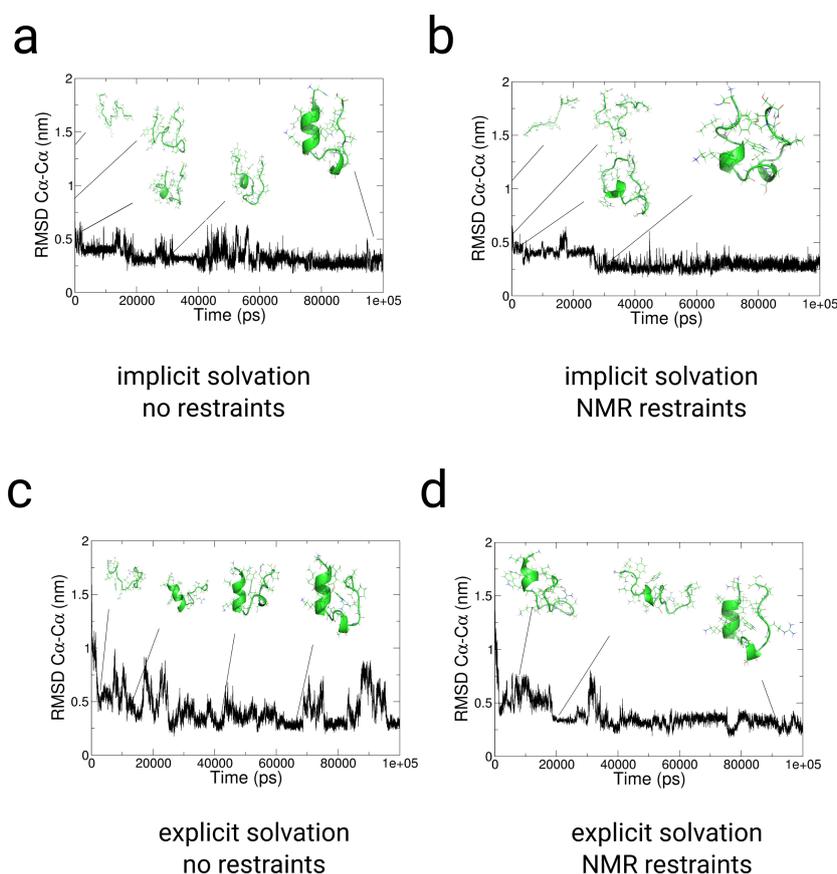
#### 2.3.1. Results

We used restraints from NMR-NOESY data from ref. [93] for folding simulations of TrpCage in implicit and explicit solvent (See the supplementary information for the restraint data used for the simulations). In the implicit solvent simulations with and without restraints, we observe a fast collapse from the extended conformation to a near native conformation at  $RMSD_{C\alpha-C\alpha} \approx 0.5$  nm within few ps (see Figure 3a,b). The subsequent 2 trajectories show a comparatively fast conversion to the folded state within 20–25 ns to  $RMSD_{C\alpha-C\alpha} < 0.24$  nm. The trajectories indicate a relative independence from the choice of the ranges of  $\tau_1$  and  $\tau_2$ , while lower coupling times lead to a higher propensity for the native structure of TrpCage (see Figure 3b). In the explicit solvent simulation without restraints, the folding-pathway of the peptide is initiated with the formation of a bend in the region from Lys8 to Gly10 at  $t = 3.8$  ns, followed by the formation of a coiled structure at 6.3 ns in which Leu7, Arg16, Asp9, and Pro18 are interacting non-specifically. At a MD-time of 12 ns, we observe the formation of an  $\alpha$ -helix between residues Asn1 and Pro12, corresponding to the native N-terminal  $\alpha$ -helix. After the process of helix formation, the backbone of the N-terminus between residues 1 and 8 remains at an  $RMSD_{\Phi} < 20^\circ$ ,  $RMSD_{\Psi} < 20^\circ$ , which indicates that the structure remains close to the native state within this region already after 12 ns till the end of the simulation (see Figure 3c). At later stages of the simulation, we observe 2 jumps in the  $RMSD_{C\alpha-C\alpha}$  from  $\approx 0.36$  nm to a value  $> 0.7$  nm, followed by a collapse of the peptide to  $RMSD_{C\alpha-C\alpha} < 0.25$  nm at 23 ns. Within the remaining simulation time, we observe a population of states in the range of  $0.2 < RMSD_{C\alpha-C\alpha} < 0.8$  nm, while the system mostly resides between  $0.2 < RMSD_{C\alpha-C\alpha} < 0.4$  nm. In contrast to the path-sampling result without restraints, the simulation using NMR-NOESY restraint data in explicit solvent shows a slower decay to a near native conformation within the first 20 ns. During this period, the peptide populates unfolded configurations at  $RMSD_{C\alpha-C\alpha} > 0.5$  nm till a partially folded state is accessed in which Trp6 is exposed to the solvent, and parts of the  $\alpha$ - as well as the 3–10 helix are formed (see Figure 3d). The initial population of unfolded states is followed by a re-opening of the peptide, which is expressed by a jump in the  $RMSD_{C\alpha-C\alpha} > 0.5$  nm and a collapse to  $RMSD_{C\alpha-C\alpha} < 0.25$  nm at approximately 38 ns. In the subsequent trajectory, the protein populates states between  $0.3 > RMSD_{C\alpha-C\alpha} > 0.17$  nm.

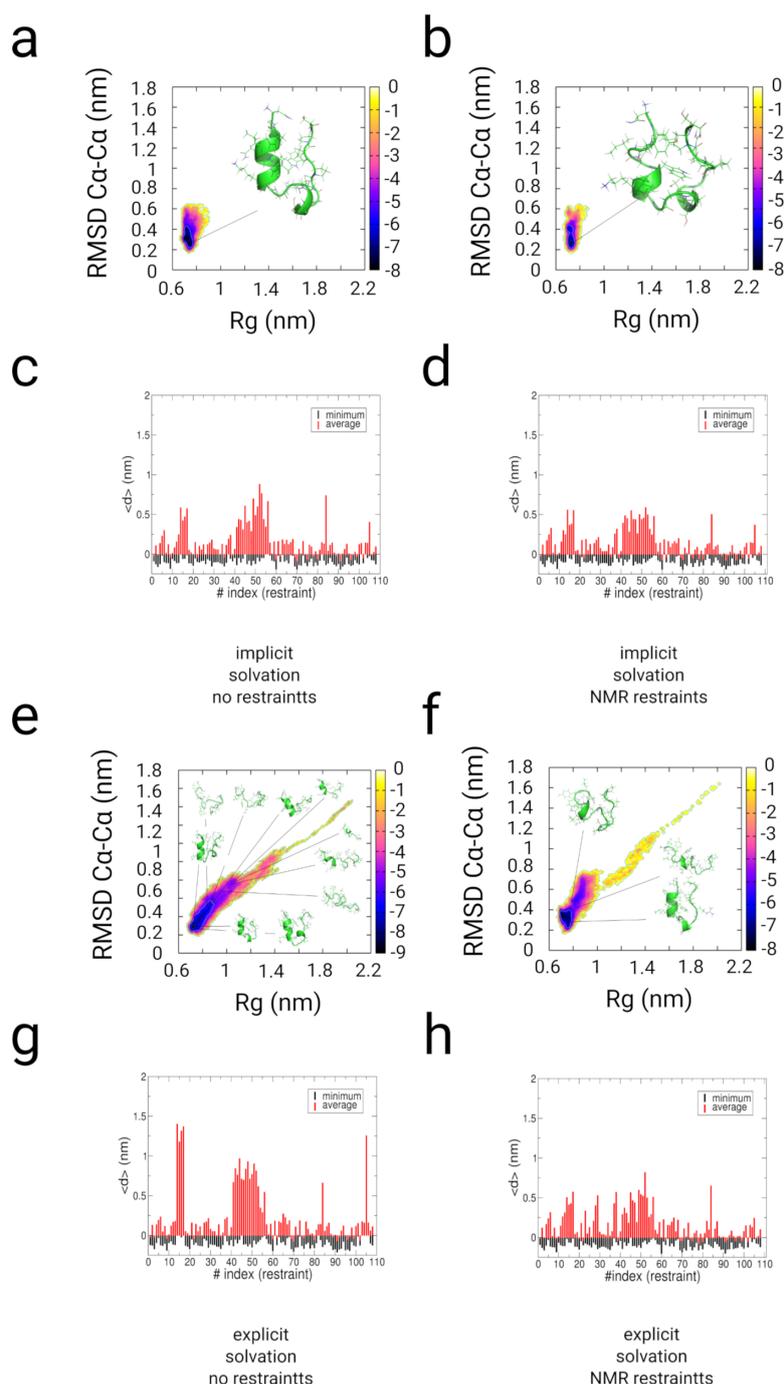
The free energy landscapes and the cluster population analysis of the implicit and explicit-solvent trajectories of TrpCage indicate that the native state of TrpCage represents the main conformer in the restrained simulations and to a smaller extent in the path-sampling simulation without experimental restraints (see Figures S4–S7 and Figure 4a,b,e,f). In the implicit solvent simulations, the native state is populated with 98.2% and 82.7% ( $RMSD_{C\alpha-C\alpha} < 0.25$  nm). In the path-sampling simulation in explicit solvent without restraints, the native state is populated with a propensity of 57.1%, while the restrained simulation yields a population of approximately 80% ( $RMSD_{C\alpha-C\alpha} < 0.25$  nm). All simulations populate a native basin at ( $RMSD_{C\alpha-C\alpha} < 0.25$  nm at free energy values of approximately  $-8$  to  $-9 k_B T$ ). We observe the population of states at  $RMSD_{C\alpha-C\alpha} > 0.4$  nm. In the case of the folding simulation without experimental restraints, the unfolded region is wider and ranges from

$1.1 > R_g > 0.8$  nm, while the restrained simulation in explicit solvent populates states at  $0.9 > R_g > 0.8$  nm ( $-6$  to  $-7 k_B T$ ).

We tested the fulfillment of the distance restraint in the measurement of the difference of the inter-atomic distances of each individual restraint as function of the restraint-index and the experimental values (see Figure 4c,d,g,h and Figure S3). In the distance profiles as function of time and the partitions, we observe values equal to and lower than zero for each of the restraints. We observe larger fluctuations with values larger than 0.5 nm for the restraints between Pro19/Tyr3 (restraint-index: 14–17) and the restraints between Pro12-Pro18/Trp6 (restraint-index: 41–52), which is indicative of a larger flexibility in that range of distances due to several re-openings of the  $\alpha$ -helix–Poly-Pro contact in all simulations and the open state in the early stages of the explicit solvent simulations (see Figure 4c,d,g,h). In the explicit solvent simulations, the application of restraints shifts the partition towards a higher degree of the fulfillment of the distance restraints in contrast to the path-sampling simulation where no experimental restraints are applied. The results in the fulfillment of the restraints between Pro12-Pro18/Trp6 (restraint-index: 41–52) and Pro19/Tyr3 (restraint-index: 14–17) indicate that the application of our restrained MD-method improves the sampling of NMR-related configurations by approximately 20–25% in comparison with the folding simulation using the multiple-path sampling method without NMR restraints.



**Figure 3.** Results from path-sampling simulations with and without restraints from NMR-NOESY measurements of TrpCage starting from an extended conformation [93] in implicit (a,b) and explicit solvent (c,d). Panels (a,c) Results from un-restrained folding simulations of TrpCage in implicit and explicit solvent. Panels (b,d) Results from folding simulations of TrpCage with NMR restraints in implicit and explicit solvent. (a–d)  $RMSD_{C\alpha-C\alpha}$  to the backbone of NMR-model #1 of PDB: 1l2y. In each of the cases, we observe folding to  $RMSD_{C\alpha-C\alpha}$  lower than 0.23 nm. The application of the set of restraints (b,d) changes the partition of conformations and prevents spontaneous unfolding to configurations at  $RMSD_{C\alpha-C\alpha}$  greater than 0.4 nm as we observe in path-sampling simulations without restraints (a,c).



**Figure 4.** Results from enhanced sampling simulations along multiple pathways with and without restraints from NMR-NOESY experiments on TrpCage. We started the simulation from an extended conformation [93] in implicit (a–d) and explicit solvent (e–h). (a,c,e,g) Results from path-sampling simulations without restraints in implicit and explicit solvent (using the multiple path sampling methodology without experimental restraints). (b,d,f,h) Results from restrained simulations using the experimental NMR restraints for TrpCage. (a,b,e,f) Free energy landscapes as function of  $RMSD_{C\alpha-C\alpha}$  and radius of gyration ( $R_g$ ). (c,d,g,h) Minimal and average value of the difference of the distance between restraint indexed atoms and the experimental value. The populations in the free energy landscapes of TrpCage are shifted in the restrained simulations with a significantly lower propensity of the unfolded ensemble at  $RMSD_{C\alpha-C\alpha}$  greater than 0.4 nm. The presence of the explicit solvent stabilizes near native configurations and leads to a better convergence of the simulation to the set of experimental restraints. Energy values on the color bars are in  $k_B T$  units.

### 2.3.2. Discussion

The ensemble of pathways of folding of the TrpCage peptide contains a variety of conformational states. Our previous studies on folding of that peptide indicated prior formation of the 3–10 helix initiated by the fast appearance of a turn in the same region [90,92,106,107]. After that first event, we observed that the folding pathways can differ in the formation of first tertiary contacts between the central tryptophan and the poly-proline helix or the first formation of the N-terminal  $\alpha$ -helix, which has been identified as a major rate-limiting step in the literature [108–114]. The sequence of events observed in our previous studies is in agreement with our results, while we do not observe a long-lived molten globule state as part of the folding pathway, which has been reported in two independent works [106,112]. The Kinetically trapped molten globule state is part of the ensemble of folding pathways of TrpCage, while we anticipate that an energy barrier separates that conformation from the set of major pathways of folding. Specific Monte Carlo moves or the choice of a set of biases can guide the system towards that molten globule state [90]. We observe that the application of the NMR-restraints improves the sampling of NMR-related configurations by approximately 20–25% in comparison with the multiple-path sampling simulation without restraints. Our observations indicate that the contact between Trp6/Tyr3 and the Poly-Proline helix is not as stable as indicated in the NMR-structural ensemble (PDB: 1l2y) and open-states as well as partially unfolded conformations might also be part of the actual NMR-ensemble. That result is also underlined by the free energy partitions as well as our analysis of the folding pathways, which are in agreement with previous investigations using different simulation methodologies [90,92,106,107,112,115,116]. From the energy differences of the unfolded state between the implicit and the explicit solvent simulation, we anticipate a stabilization free energy of the unfolded state of TrpCage due to the solvation of approximately  $2\text{--}4 k_B T$ .

### 2.4. Simulations with Restraints from Chemical Crosslinking Data: NKR-P1A

#### 2.4.1. Results

As second application, we simulated NKR-P1A in implicit and explicit solvent using the chemical crosslinking distances and distance-vectors as restraints (see supplementary information, data from ref. [89]). NKR-P1A is an important activating receptor expressed at the surface of natural killer cells [117]. We used the X-ray structure (PDB: 3m9z) as starting structure for the restrained simulations. This structure consists of a central segment of a mixed  $3\beta + 1\alpha + 2\beta + 1\alpha$ -fold between residues Leu92/Thr159, Glu185/Tyr215, which are linked via disulfide bridges at Cys122-Cys210, Cys189-Cys202, and Cys94-Cys105. A flexible loop with partial  $\alpha$ -helical content is located between Thr159 and Glu185. In a cross-linking experiment and another recent NMR-study, it has been shown that this flexible region is bound to the protein core, while it might still persist as a loop in its secondary structure [89,118]. Another study indicated the role of the loop region as an aggregation interface with interacting proteins [119]. We applied the same set of chemical cross-linking vectors with a target value  $d_0 = 0.75$  nm using the crystal structure as a starting configuration. A value for  $d_0$  equal to 0.75 nm leads to an attractive bias vector for all the chemical restraint values larger than that parameter and aims on the formation of approximate contacts between the 2 indexed aminoacids.

We simulated the system in implicit and explicit solvent. In the  $RMSD_{C\alpha-C\alpha}$  to the final structure of the simulation, we observe a decay from 1 nm to a collapsed state at 0.5 nm within 15 ns in the implicit solvent simulation and within approximately 38 ns in the simulation using explicit solvent (see Figure 5a,c). We then find another drop in the  $RMSD_{C\alpha-C\alpha}$  from 0.5 to approximately 0.2–0.3 nm at approximately 42 ns in the implicit solvent simulation and at approximately 44 ns in the simulation in explicit solvent. We analyzed the free energy landscapes as function of the  $RMSD_{C\alpha-C\alpha}$  to the final structure and the radius of gyration ( $R_g$ ) (see Figure 5b,d). In general, we find 5 different states of NKR-P1A throughout the simulation in implicit and explicit solvent. The first state occupies a conformation, in which the loop is flexible and extended. This state is similar to the X-ray structure

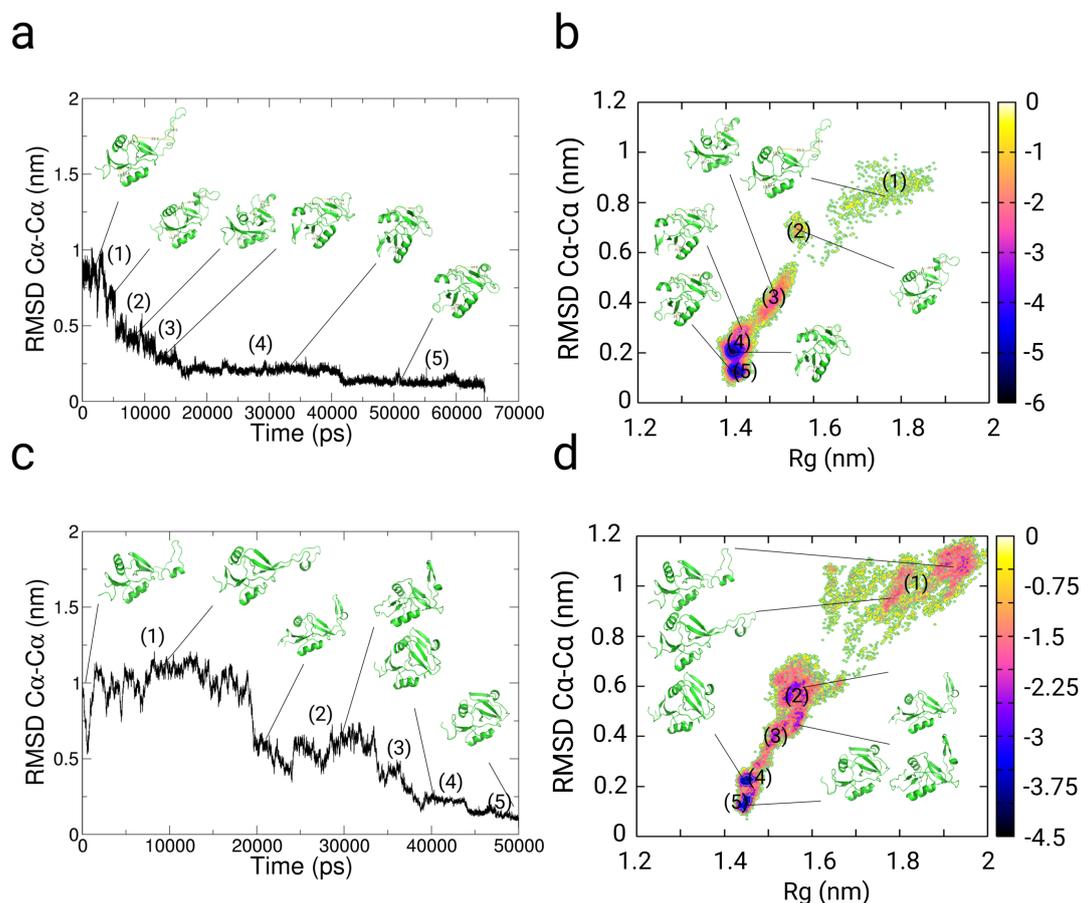
and populates values in  $RMSD_{C\alpha-C\alpha} \approx 0.8\text{--}1.1$  nm and  $R_g \approx 1.6\text{--}1.7$  nm in both simulations (see Figures 5a,c and 6a,b). In the explicit solvent simulation, this conformation also includes  $\beta$ -strand formation in one segment of the loop-region between residues Met163 and Asn174 (see Figure 6b). The second state at  $RMSD_{C\alpha-C\alpha} \approx 0.5\text{--}0.7$  nm and  $R_g \approx 1.45\text{--}1.65$  nm is characterized by a contact between Tyr158, Ile180, and Gly182 at the loop region with residues Phe199, Glu200, and Ser201 at the protein core. At the same time, parts of the loop region rotate around the  $\beta$ -strand axis, while parts of the loop segment are still exposed to the solvent, which is not as strongly visible in the implicit solvent simulation (see Figures 5a,c and 6c). The states (3), (4), and state (5) represent conformations in which the loop region is bound to the protein core at different internal coordinates at the interface with the 2  $\beta$ -region of the protein core (orange color, see Figure 6c,d). The state (3) is located at  $RMSD_{C\alpha-C\alpha} \approx 0.4$  nm and  $R_g \approx 1.55$  nm, where the loop-region is further rotated around the strand-axis with the same contact-order to the protein-core as in state (2), while residues Asn169, Ile168 at the turn of the strand-region in the loop are closer to the 2 $\beta$  segment of the protein core due to the rotation. In the explicit solvent simulation, the states (4) and (5) are shifted in the  $R_g$ -value by 0.025 nm in comparison with the simulation in implicit solvent. In both cases, we observe that both states (4) and (5) contain the strongest level of attraction with energies of  $-4.5 k_B T$  ( $-6 k_B T$  in the implicit solvent). In the implicit solvent simulation, the un-bound states (1, 2) are not stabilized as well as in the explicit solvent simulation, leading to a higher free energy of the bound states (3, 4, 5). The states (4, 5) differ in their location along the 2 $\beta$  segment (orange) at the protein core. In general, we observe contacts between Trp165, Trp167 (loop) and Val197 (core), Ile168 (loop) and Leu142 (core), Trp165 (loop) and Phe199 (core). The strand (Asn164-Asn174) within the loop region is stabilized on this segment. In the state (5), the strand reorients and shifts towards one of the  $\alpha$ -helical segments (Glu137-Ser144), forming a contact between residues Asn169 (loop) and Gln135 (core) (see Figures 5b,d and 6d).

To summarize, in the measurement of the distances between the restraint-indexed atoms, we observe fulfillment of 9 of the chemical crosslinking distances already at the beginning of the simulation (see Figure S8, and Figure 7a,c). For the restraint-distances with index #2 (residues Lys179-Lys196) and #10 (residues Asp176-Lys196), we observe populations with positive values in the partition of the difference ( $d(t) - d_0(\text{exp})$ ) and the distances over time (see Figure 7). Both restraints are a measure for the orientation of the loop-region to the protein core. The states (1)–(3) populate states with distances larger than 2.0/1.2 nm given as chemical cross-linking distance. The restraints with indexes #1, #3–#6, #8, #9, and #11 populate minima below the chemical cross-linking distance throughout the trajectory, which represent re-orientations within the protein-core. For restraint #7 (Glu147-Lys148), we observe the population of 2 states. Finally, we state that we observe a comparatively low level of fulfillment of the restraint #10 in the time-trace of the simulation, where the difference value converges to a value of approximately 0.7–0.8 nm (see Figure 7b,d and Figure S8).

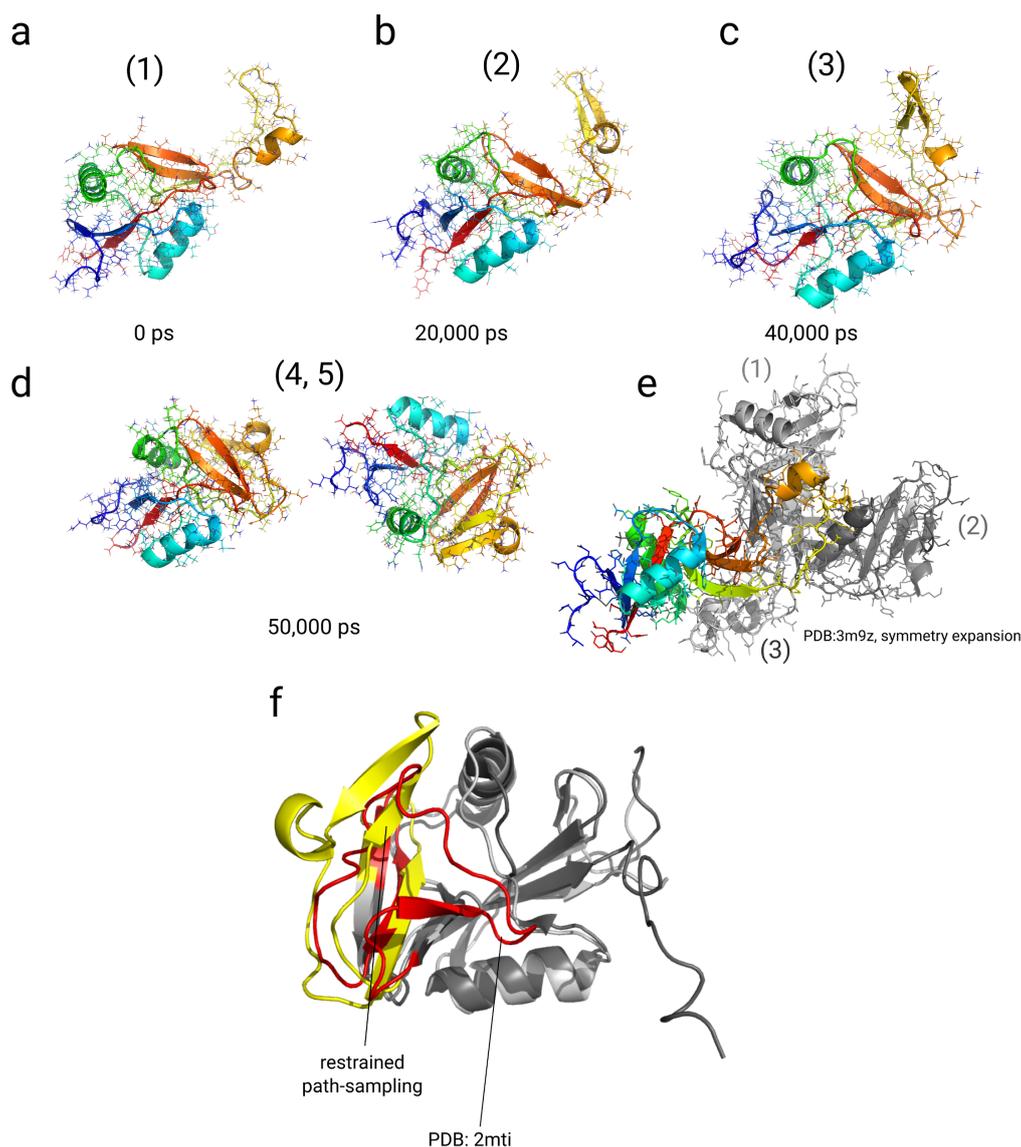
#### 2.4.2. Discussion

The simulations have relaxed the structures into a conformation in which the loop region is bound to the protein core, and parts of the loop form a  $\beta$ -stranded conformation. In general, we observe the population of 5 different states, while the first state corresponds to the experimental X-ray structure. In the X-ray structure, the open state of the loop-region reflects a bound state to other monomer within the protein crystal, where the loop binds to 2 other monomer units (see Figure 6e). That state might resemble a conformation, in which NKR-P1A binds to other proteins in immune-signaling pathways [119]. The orientation of the loop-region and its partial  $\beta$ -strand secondary structure found in our simulations has a different orientation than in 2 other studies [89,118]. In the 2 structural models, the loop-region is also in contact with the protein core, but it has another orientation as well as a smaller  $\beta$ -strand content. In the 2 modeling studies based on NMR- and chemical restraint data, the loop region binds the protein core in a salt-bridge between Asp176 and Lys196 [89], while an NMR-study indicated a similar packing of that region [118]. Our model differs in the  $RMSD_{C\alpha-C\alpha}$  equal to 0.56 nm from the NMR-structure (See Figure 6f) [118]. Using the chemical restraint data, our algorithm reaches

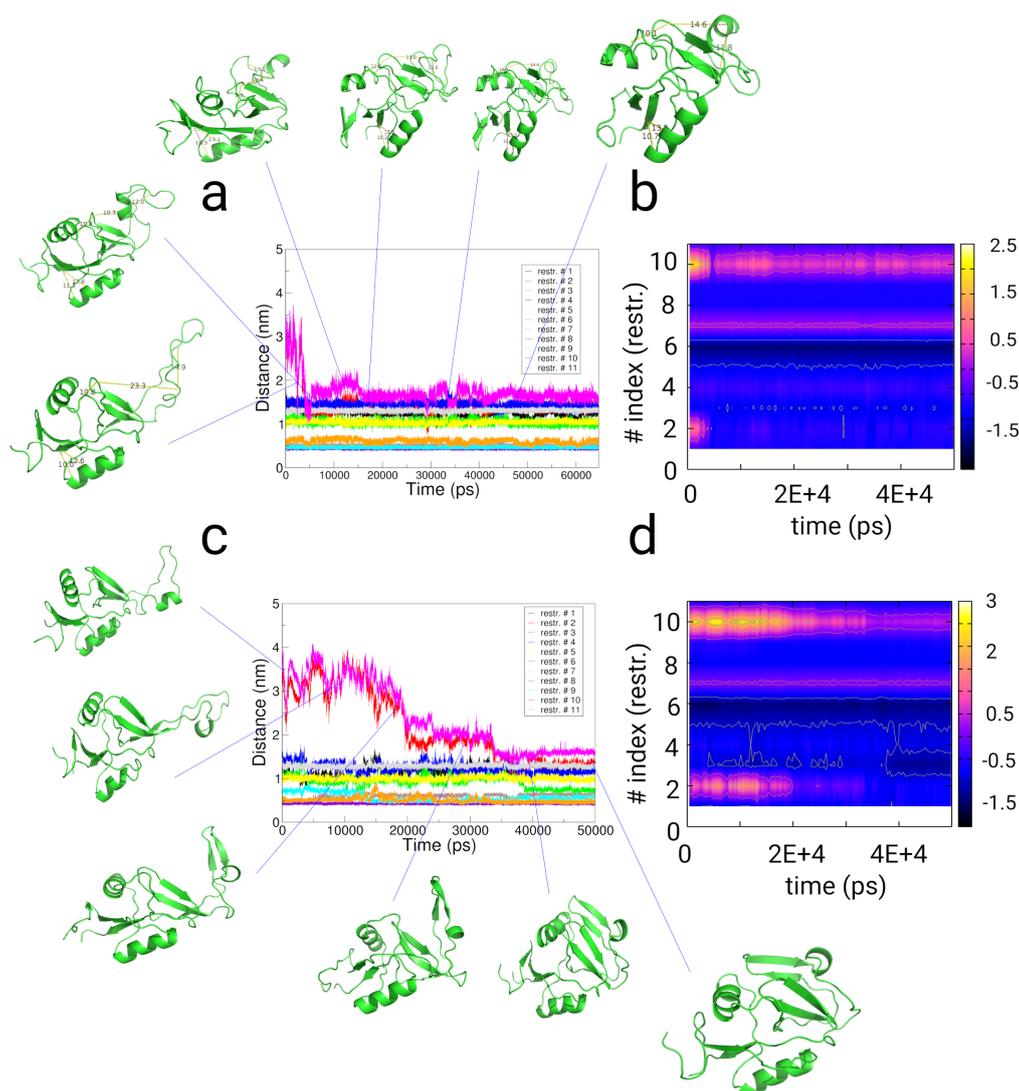
a configuration with a potentially better packing of the loop region than the models of the same protein described previously, which underlines the importance of a dynamical sampling along the chemical restraint data. From the energy differences of the unbound state between the implicit and the explicit solvent simulation, we anticipate a stabilization free energy of the open state of NKR1 $\alpha$  by the solvent of approximately  $2 k_B T$ .



**Figure 5.** Results from restrained simulations of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A) along multiple path-dependent biasing variables with chemical cross-linking restraints from ref. [89] (PDB: 3m9z as starting structure). **(a,b)** Results from enhanced restrained simulations using the path-dependent biasing approach in implicit solvent. **(c,d)** Results from enhanced restrained simulations along multiple path-dependent biases in explicit solvent. **(a,c)**  $RMSD_{C\alpha-C\alpha}$  to the final structure of each simulation. **(b,d)** Free energy landscape as function of  $RMSD_{C\alpha-C\alpha}$  and the radius of gyration ( $R_g$ ). Energy values on the color bar are in units of  $k_B T$ . The protein relaxes quickly out of the extended conformation in both simulations within few ns and adopts 5 different compact states, where the loop region (residues: Arg157-Ser188) are packed to the  $\beta$ -strand interface (residues: Ser188-Asn203) and parts of the second N-terminal helix (residues: Gln135-Ile145) of the monomer.



**Figure 6.** Results from restraint multiple path-sampling simulations of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A) using chemical cross-linking restraints from ref. [89] in explicit solvent (PDB: 3m9z as starting structure). Conformations at different stages of the restrained simulation (0 ps (a), 20,000 ps (b), 40,000 ps (c), and 50,000 ps (d)). (e) Symmetry expanded structure of the X-ray structure (PDB: 3m9z) with monomers in contact with the loop region (residues: Arg157-Ser188) of the protein. Monomer (1) (light grey) shares a contact with its 3  $\beta$ -strands (residues: Tyr149-Ser201) with the loop region in the crystal structure. Monomer (2) (dark grey) is in contact with its second  $\alpha$ -helix (Gln135-Lys146) with the loop region. (f) Structural overlay of NKR-P1A structure after 50 ns in the explicit restrained simulation (protein core in grey color, loop regions in yellow color) with the experimental NMR-structure (protein core in dark grey color, loop region in red color) (PDB: 2mti [118]). The  $RMSD_{C\alpha-C\alpha}$  between the 2 structures equals 0.56 nm.



**Figure 7.** Results from restrained simulations of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A) using chemical cross-linking restraints from ref. [89] (PDB: 3m9z as starting structure). (a,b) Results from implicit solvent restrained simulations. (c,d) Results from restrained simulations in explicit solvent. (a,c) Distances along 11 distance restraints as function of simulation time given by chemical crosslinking experiments as given in ref. [89]. (b,d) Difference between the distance and the chemical restraint distance value as given in ref. [89] as function of simulation time and the restraint index. For the restraints #2 and #10, we find conformations, which also populate states with positive values in the difference between the distance and the restraint value.

### 3. Methods

In a recent work, we introduced a path-variable approach for the sampling along 2 biasing increments derived from the history-dependent momenta  $p$  and displacements  $dq$  [92]. We implemented an adaptive bias MD and a path-sampling component for the accelerated sampling of protein folding and aggregation. In this work, we developed 3 extensions of the previous work: (1) We extended the time-periods for the sampling to the simulation along multiple time-ranges to capture the heterogeneity of structural transitions in biomolecular systems. (2) We implemented a renormalization scheme to facilitate the sampling within the underlying ensemble in dependency of a linear coupling factor. (3) We combined the first 2 methodologies with the simulation along experimental restraint vectors using a novel coupling scheme, in which the accelerating bias along multiple time-ranges is adaptively reweighted. Each of the 3 components are directly linked to each other in a consistent

definition of a bias-gradient, while a limited number of input parameters is required. We start with a description of the restraint modeling and compare with the modeling in the energy space. Then we introduce the restrained Hybrid Hamiltonian and the renormalization followed by a description on the impact of the restraint vector on the resulting averages. As a last part, we introduce the enhanced sampling methodology along multiple path-dependent variables and define the bias gradients used in the restrained sampling (see also the Supplementary Material for a derivation of each segment: Section I, Methods).

### 3.1. Theory

#### Adaptive Restraint Modeling

Restraint modeling combined with MD-simulations is a standard tool used for NMR-structure determination, protein structure refinement and applications such as the structural modeling of a system on which restraints are imposed by chemical crosslinking/mass-spectrometry experiments. We note that the restraint modeling strongly depends from the quality of the used experimental data-set, where divergence in the data as for intrinsically disordered proteins would lead to misguided simulations. In the case of a defined experimental dataset containing a detailed residue-based distance information, a very general way to impose the restraint information on the system is the definition of a restraint in the potential energy space  $V$  of the system, where a classical potential  $V_{unbiased}$  is supplemented by an additional restraint potential  $V_r$  to fulfill an underlying set of experimental information about the system:

$$V = V_{unbiased} + V_r, \quad (3)$$

leading to forces:

$$\nabla V = \nabla V_{unbiased} + \nabla V_r, \quad (4)$$

which guides the system towards a potential energy minimum within the given set of restraint parameters, if it is possible for the system to relax into that state. The given set of restraints leads to a faster relaxation of the system towards the potential experimental configurations and the direct comparison of structural information based on experimental input, which directly can connect experiments with simulations on an atomistic length scale. However, an approach imposing direct information on a system containing a system-specific large number of internal degrees of freedom, can guide towards potentially wrong conformations of the system. For example, the novel restrained Hamiltonian can define new local metastable traps in which the system remains for long periods, although such a trap does not correspond to the true experimental ensemble of conformations.

Especially in chemical-crosslinking/mass spectrometry experiments, the experimental information is represented by a distance value  $d_0$ , which corresponds to the geometrical length of the crosslinking molecule between 2 residues. In that case, the restraint between the 2 experimentally crosslinked residues aims on the fulfillment of the condition that the distance  $d$  between the 2 residues fulfills:

$$d \approx d_0 \pm \Delta d, \quad (5)$$

where  $\Delta d$  represents the structural variability of the inter-residue distance  $d$  within the ensemble of simulated restrained conformations. That means that the distance information contained in the crosslinking experiment is far from a discrete distance value, but represents an average value  $\langle d_0 \rangle$  with an unknown system specific variation. The same formalism in general holds for NMR-NOESY distance information, where a variability  $\Delta d$  of the simulated conformations around the distance  $d_0$  also has to be taken into account. Especially, if restraints from chemical crosslinking experiments are used, the variability  $\Delta d$  can be very high, due to the chemical flexibility of the linkers, which motivates our development of an adaptive reweighting of a path-dependent bias in dependency of an angular overlap of the given restraint vector between 2 residues and the adaptive path-dependent bias. That approach has the advantage, that the fulfillment of the restraint corresponding to the equilibrium

value  $d_0$  is not explicitly imposed on the system by a restraint potential  $V_r$  as function of  $d_0$ , but as function of a new reweighted hybrid Hamiltonian, which only depends from the un-biased forcefield parameter set.

### 3.2. Hybrid Hamiltonian

In the following, we define the unbiased Hamiltonian  $H(A)$ :

$$H(A) = T(A) + V(A), \quad (6)$$

where  $T$  stands for the Kinetic and  $V$  is the potential energy. The biasing Hamiltonian  $H(B)$  consists of the biased Kinetic  $T(B)$  and potential energy terms  $V(B)$  in analogy to the definition of  $H(A)$ . We add that biasing Hamiltonian  $H(B)$  to  $H(A)$  using a renormalization technique [90], which results in a new Hamiltonian  $H(C)$ :

$$H(C) = \frac{H(A)}{1 + \alpha_{md}} + \frac{\alpha' |H(A)|}{|H(B)|} H(B), \quad (7)$$

where we distinguish between  $\alpha_{md}$  as coupling parameter renormalizing the un-biased Hamiltonian and  $\alpha'$ , which defines the coupling of the bias to the system. We define:

$$\alpha_{md} = \eta_{md} \beta_{md} (1 - \xi), \quad (8)$$

and

$$\alpha' = \eta' \beta (1 - \xi), \quad (9)$$

where  $\xi$  stands for a uniform random number,  $\beta$ ,  $\beta_{md}$ ,  $\eta'$ , and  $\eta_{md}$  are simulation parameters.

### 3.3. Restrained Hybrid Hamiltonian

The modification of the bias-function to a restrained Hamiltonian  $H(B)_R$  along experimental restraints leads to a re-definition of the hybrid Hamiltonian to a restrained energy function:

$$H(C)_R = \frac{H(A)}{1 + \alpha_{md}} + \frac{\alpha' |H(A)|}{|H(B)_R|} H(B)_R, \quad (10)$$

where the new hybrid Hamiltonian  $H(C)_R$  consists of a reweighted experimental restraint component  $r$ . In principle, the total energy remains approximately un-affected, while other properties can be introduced through the bias  $H(B)$ . We consider a set of distance restraints  $r$  as an external experimental input for the generation of a biasing energy function  $H(B)_R$ . The energy  $H(B)$  consists of an adaptive bias MD component  $H(B)_{ab}$  and a path-sampling part  $H(B)_\sigma$ :

$$H(B) = H(B)_{ab} + H(B)_\sigma. \quad (11)$$

We then define a restraint bias-gradient  $\nabla H(B)_R$ , which only acts on the system through the cosine of the angle  $\Xi$  between the distance-restraint vector and the biasing increments  $\nabla H(B)$ , in a way that the applied bias changes with its orientation relative to the distance vector defining the restraint  $r$ . Therefor, we calculate the cosine  $\cos(\Xi)$  between the bias  $s$  and  $q_{12}$  using :

$$\cos(\Xi) = \frac{\nabla H(B) \cdot q_{12}}{|\nabla H(B)| |q|}, \quad (12)$$

that the resulting bias-gradient changes its sign from  $-\frac{1}{2}\pi < \Xi < \frac{1}{2}\pi$  to  $\frac{1}{2}\pi < \Xi < \frac{3}{2}\pi$ , while the increment in the path remains constant as in the formalism described before. We mention that the cosine fulfills the property that for the angles  $\frac{1}{2}\pi < \Xi \leq \pi$  and  $\pi \leq \Xi < \frac{3}{2}\pi$  the sign of the applied bias changes towards the direction of the distance restraint. We define the restraint by a distance vector

between 2 atoms expressed as  $q_{12} = q_1 - q_2$  and modify the applied bias  $\nabla H(B)$  at time  $t$  to obtain a restraint bias  $\nabla H(B)_R$  of the form:

$$\nabla H(B)_R = \nabla H(B) \cos(\Xi) \frac{(d_{12} - d_0)}{|d_{12} - d_0|}, \quad (13)$$

where  $d_0$  stands for the equilibrium distance of the restraint (see Figure 1). Equation (10), then, can be expressed as:

$$H(C, \Xi)_R = \frac{H(A)}{1 + \alpha_{md}} + \frac{\alpha' |H(A)|}{|H(B, \Xi)_R|} H(B, \Xi)_R. \quad (14)$$

The methodology has the following advantages:

1. The momentum conservation is still obeyed and no additional energy is applied on the system, since the absolute value of  $|\nabla H(B)_R|$  varies from 0 to  $|H(B)|$ , due to  $0 \leq |\cos(\Xi)| \leq 1$  for all possible values of  $\Xi$ .
2. The effective biasing increments  $\nabla H(B)_{ab}$  ( $dL_{ab_{ik}}(t)$ , adaptive bias MD) and  $\nabla H(B)_{\sigma}$  ( $dL_{\sigma_{ik}}(L_{ik}(t))$ , path-sampling) are modified for the atoms on which the restraint  $r$  is applied. That means that the path-increments are changed to  $\nabla H(B)_{ab_R}$  and  $\nabla H(B)_{\sigma_R}$  using a flexible bias vector in dependency of the vectorial quantity of  $\nabla H(B)$  and the restraint  $r$ , while the absolute increments applied per unit-time remain identical:  $|\nabla H(B)| = |\nabla H(B)_R|$ .

### 3.4. Restrained Partitions

If a restraint potential  $V_r$  would be added in conventional sampling along given restraints  $r$ , the novel restrained partition function  $Z_R$  is defined by:

$$Z_R = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{k_B T} (T(A)+V(A)+T_r+V_r)} dr^1 dr^2 \dots dr^N dp^1 dp^2 \dots dp^N, \quad (15)$$

where  $T_r$  is the restrained Kinetic energy. Using our newly developed restrained Hamiltonian  $H(C, \Xi)_R$ , the restrained partition results in:

$$Z'_R = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{k_B T} (H(C, \Xi)_R)} dr^1 dr^2 \dots dr^N dp^1 dp^2 \dots dp^N, \quad (16)$$

where the hybrid Hamiltonian  $H(C, \Xi)$  is defined in the Equation (14), which is a fundamental difference to the partition as described in the first equation, because the scalar quantity of the energy  $H(C, \Xi)$  is only affected in dependency of the coupling factors  $\alpha_{md}$  and  $\alpha'$ . The shifted average  $\langle X \rangle_R$  through the application of a restraint in the energy space is then written as:

$$\langle X \rangle_R = \frac{X e^{-\frac{1}{k_B T} (T(A)+V(A)+T_r+V_r)}}{Z_R}, \quad (17)$$

while the novel sampling methodology using a sampling along reweighted path-dependent bias is defined as:

$$\langle X \rangle_{R'} = \frac{X e^{-\frac{1}{k_B T} (H(C, \Xi)_R)}}{Z'_R}. \quad (18)$$

The 2 last expressions (17) and (18) also indicate the limitations and advantages of both different restrained sampling techniques: In the first example, where the restraint is applied in the energy space using  $V_r$ , a new potential is added to the system, which can improve the underlying parameter set significantly and leads to an accelerated convergence to a relevant conformation. Examples for applications of that approach can be found in the widely used Gō-modeling approaches [15,120]. However, it might also lead to an 'over-riding' of the original parameter set  $V(A)$ , in a way that  $V_r$  becomes the dominant part in the Hamiltonian. That would mean that essential conformations as

part of  $\langle X \rangle_R$  could not be accessed. In the most extreme case, the potential  $V(A)$  is 'over-ridden' by  $V_r$  that the resulting conformers are not part of the experimental ensemble. In contrast, the approach represented by the Equation (18) has the following limitations and advantages: (1) It requires longer simulation periods for the system to relax into the area of fulfillment of the restraint, since the actual bias is only added in dependency of a necessarily small coupling value  $\alpha$ . However, the underlying partition corresponds to the Hamiltonian defined by the forcefield parameters. (2) The resulting average contains primarily only the data in a direct dependency of the used forcefield parameter set due to the definition of the hybrid Hamiltonian. That can result in a limitation of the resulting statistical averages, but has the advantage that the sampled configurations correspond to an approximate equilibrium state.

### 3.5. Path-Sampling Bias: $\nabla H(B)$

#### 3.5.1. Theory

We define the bias Hamiltonian  $H(B)$  used for the accelerated sampling along multiple path increments  $dL_{ik}$  (for pathways  $i$  and atom-indices  $k$ ) as well as its modification to its principal components, which is adaptively changed into a bias  $H(B)_R$  in dependency of a distance restraint  $r$  given by experimental data. We consider that the simulated system in an equilibrium simulation propagates along a pathway with the general condition that the reduced action  $L$  as function of momentum  $p$  and positions  $q$  remains constant [121,122]:

$$L = \oint pdq = \text{const.} . \quad (19)$$

Along a time-dependent MD trajectory of a system exposed to non-zero fluctuations of its momenta  $dp$ , we rewrite the Equation (19) as a time-integral:

$$L = \int_{t=t_0}^{t_1} \left( \frac{d(p(t)dq(t))}{dt} \right) dt' = \text{const.} , \quad (20)$$

where  $t_0$  stands for the start and  $t_1$  for the end of the simulated MD trajectory. The time-dependent integral is expressed as:

$$L = \int_{t=t_0}^{t_1} \left( \frac{dp(t)}{dt} dq(t) + p(t) \frac{dq(t)}{dt} \right) dt' = \text{const.} . \quad (21)$$

We then define a differential  $dL(t)$  for a microscopic system, in which fluctuations of the momenta  $dp(t)$  occur. That local change in  $L(t)$  at the time  $t$  is defined by:

$$\frac{dL(t)}{dt} = \frac{d}{dt}(p(t)dq(t)) = \frac{dp(t)}{dt} dq(t) + p(t) \frac{dq(t)}{dt} . \quad (22)$$

We obtain the following differential at time  $t$ :

$$dL(t) = p(t)dq(t) + dp(t)dq(t) = (p(t) + dp(t))dq(t) . \quad (23)$$

The expressions in the Equations (20)–(23) relate to the standard MD-case with fluctuations in the momenta  $dp(t)$  and displacements  $dq(t)$  at times  $t$ . Any arbitrary biasing technique developed to accelerate a MD simulation adds an instantaneous increment  $dL_s(t)$  to  $dL(t)$ :

$$dL'(t) = dL(t) + dL_s(t) , \quad (24)$$

where we obtain a modified increment  $dL'(t)$  and additional changes in the momenta resulting from applied bias-energies affect the instantaneous action of a system in order to reach a faster convergence to the statistical average, i.e., the free energy landscape [16,47,74]. In our recent work, we defined 2

biasing increments  $dL_s(t)$ :  $dL_{ab}(t)$  (adaptive bias MD) (in the original work, we refer to the variable  $dL$ ) and  $dL_{\sigma}(L_{ik}(t))$  (path-sampling) depending from 2 coupling time intervals  $\tau_1$  (adaptive bias MD) and  $\tau_2$  (path-sampling) in which the gradient has been evaluated [92]:

$$dL_s(t) = dL_{ab}(t) + dL_{\sigma}(L_{ik}(t)). \quad (25)$$

We extend the formalism to the sampling within multiple biases defined by the parameter  $N_R$  standing for the number of multiple biases. We redefine the expression (20) to a multiple sampling in multiple bias-paths along  $N_R$  multiple biases. The 2 path increments  $dL_{ab_{ik}}(t)$  and  $dL_{\sigma_{ik}}(L_{ik}(t))$  are derived from the adaptive bias MD and the path-sampling components of the hybrid algorithm [92]. Both components are added to the un-biased path  $\int_{t_0}^{t_1} dL(t)dt'$  such that the principle of conservation of the action integral is obeyed, and the action integral has to remain constant upon addition of the bias in order to sample the system along an equilibrium trajectory. Taking into account that especially protein systems, and more general aqueous solutions behave heterogeneous related to their relaxation behavior (depending on their actual state), an extension to multiple biases, which is coupled to a finite set of different relaxation times  $\tau_{1_{ik}}$  and  $\tau_{2_{ik}}$  for  $N_R$  biases with index  $i$  and  $k$  individual atom indices, leads to a more efficient sampling of dynamically heterogeneous systems. That way, we apply that individual number of  $N_R$  biases within each individual bias  $i$ , for which a bias  $s$  is re-evaluated within periods  $\tau_{1_{ik}}$  and  $\tau_{2_{ik}}$  for the atom with index  $k$  and is applied over the same period. We define the individual time periods  $\tau_{1_{ik}}$  and  $\tau_{2_{ik}}$  over  $N_R$  multiple pathways with the atom-index  $k$  as:

$$\tau_{1_{i=N_R,k}} = \sum_{i=1}^{N_R} \tau_{1_k}, \quad (26)$$

and

$$\tau_{2_{i=N_R,k}} = \sum_{i=1}^{N_R} \tau_{2_k}. \quad (27)$$

Although we apply constant values  $\tau_1$  and  $\tau_2$  over all atoms, we introduce the notation with an additional index  $k$ , which would mean that each individual atom  $k$  can potentially be coupled to an individual value  $\tau_{1_{ik}}$  or  $\tau_{2_{ik}}$ , which might lead to a higher accuracy to capture individual relaxation times of each atom in the system.

### 3.5.2. Adaptive Bias MD and Path-Sampling

In the *adaptive bias MD* section of the algorithm we define the first bias component  $\nabla H(B)_{ab}$ . We introduce  $N_R$  multiple biases in which the bias is re-evaluated within periods of  $\tau_{1_{ik}}$  for the bias with an index  $i$  and atom  $k$ . We define the corresponding force  $F_b(t)$  at time  $t$ , and use time-derivative of  $\nabla H(B)_{ab}$ :  $\frac{d}{dt} H(B)_{ab} = H(\dot{B})_{ab}$ . Referring to the previous work, we call that biasing increment  $dL_{ab}$ . We define the bias along a history-dependent pathway in adaptive bias MD, and introduce a coarsening along multiple finite time increments  $\tau_{1_{ik}}$  to coarse-grain the dynamics and to increase the computational efficiency, which leads to an expression for the corresponding force in adaptive bias MD:

$$\begin{aligned} \nabla H(B)_{ab}(\tau_1) &= \frac{dH(\dot{B})_{ab}}{d\tau_1} d\tau_1 \\ &= \sum_i^{N_R} \sum_k^N \left[ \gamma'_{ik}(t) \frac{d}{d\tau_{1_{ik}}} \left( \frac{d}{dt} [(p_k(t) + dp_k(t)) dq_k(t)] \right) d\tau_{1_{ik}} \right. \\ &\quad \left. + \frac{d\gamma'_{1_{ik}}(t)}{d\tau_{1_{ik}}} d\tau_{1_{ik}} \left( \frac{d}{dt} [(p_k(t) + dp_k(t)) dq_k(t)] \right) \right], \end{aligned} \quad (28)$$

where  $\gamma''_{ik}$  stands for the fluctuation range with the dimension of a length ( $[nm^{-1}]$ ) and  $\xi$  is a normally distributed random number with a weight equal 1 (we used a constant value  $\gamma''_{ik} = 10^{-4}$  in all simulations). For the *path-sampling* component  $\nabla H(B)_\sigma$ , we use a definition of the reactive coordinate  $\sigma_{ik}(t)$  to determine the biasing segment  $\nabla H(B)_\sigma$  as a function of the increment  $dL_{\sigma_{ik}}(L_{ik})(t)$ , which we define as [92]:

$$\sigma_{ik}(t) = dL_{\sigma_{ik}}(L_{ik})(t), \quad (29)$$

A history dependent bias potential  $\Phi_{ik}^t$  is added in intervals of  $\tau_{2_{ik}}$ :

$$\Phi_{ik}^t = -\frac{\partial}{\partial \sigma_{ik}} W_{ik} \sum_{t \leq t_b} \prod_{ik} \exp \left( -\frac{|\sigma_{ik} - \sigma_{ik}^{t-\tau_{2_{ik}}}|}{2\delta\sigma_{ik}^2} \right), \quad (30)$$

where the height  $W_i$  and the width  $\delta\sigma_i$  are conventionally parameters chosen to provide computational efficiency and an efficient exploration of the free energy  $\mathcal{F}(\nabla H(B)_\sigma)$ . We define the bias component as:

$$\nabla H(B)_\sigma(\tau_{2_{ik}}) = \nabla_{\sigma_{ik}(t)} \Phi_{ik}^t. \quad (31)$$

That formulation constantly drives the system to explore new configurations along the variable  $L_{ik}(t)$  and prevents the system to revisit conformers, which have been sampled previously.  $W$  and  $\Delta E$  are constants [123]. In order to accelerate the sampling along  $H(B)$  and  $H(B)_R$ , we re-evaluate the principal components of  $H(B)$  in order to sample the system along its slowest modes of the 2 segments  $H(B)_{ab}$  and  $H(B)_\sigma$ . Therefore, we diagonalize the matrices  $dL_{\sigma_{ik}}(L_{ik})$  and  $dL_{ab_{ik}}(t)$ . The corresponding eigenvectors with the smallest eigenvalue represent the slowest modes and are applied to the system [124].

### 3.6. Bias Gradients

The un-restrained total bias gradient  $\nabla H(B)$  is then defined as:

$$\nabla H(C) = \frac{\nabla H(A)}{(1 + \alpha_{md})} + \nabla H(B)_{ab} \alpha' \frac{|\nabla H(A)|}{|\nabla H(B)_{ab}|} + \nabla H(B)_\sigma \alpha' \frac{|\nabla H(A)|}{|\nabla H(B)_\sigma|}, \quad (32)$$

while the restrained gradient is written as:

$$\begin{aligned} \nabla H(C, \Xi)_R &= \frac{\nabla H(A)}{(1 + \alpha_{md})} + \left( \nabla H(B)_{ab} \alpha' \frac{|\nabla H(A)|}{|\nabla H(B)_{ab}|} + \nabla H(B)_\sigma \alpha' \frac{|\nabla H(A)|}{|\nabla H(B)_\sigma|} \right) \\ &\times \cos(\Xi) \frac{(d_{12} - d_0)}{|d_{12} - d_0|}. \end{aligned} \quad (33)$$

### Algorithm

- Start loop over MD steps.
- Loop over  $N_R$  biases,  $N$  atoms.
- Determine the path-dependent increments  $dL_{ik}$  at time  $t$  for the *adaptive bias MD* segment at periods of  $\tau_1$  to define the biasing element  $\nabla H(B)_{ab}$  (see Equation (28)).
- Evaluate  $\Phi_{ik}^t$  to increase the history dependent bias at periods of  $\tau_2$  for the bias with index  $i$ , atom  $k$  for the definition of  $\nabla H(B)_\sigma$  (*path-sampling*) (see Equation (30)).
- Calculate gradients after renormalization in bias with index  $i$ , atom  $k$  (*path-sampling* and *adaptive bias MD*) (see Equation (7)).
- Determine the principal components of the 2 segments  $H(B)_{ab}$  and  $H(B)_\sigma$ .
- Evaluate overlap of bias  $\nabla H(B)$  with the given restraint to obtain  $\nabla H(B)_R$  (see Equation (12)).
- Add all gradients.

### List of parameters for the parallel sampling algorithm

- $N_R$ , number of parallel biases
- $\tau_1$ , time-period for the adaptive bias MD for  $N_R \times \tau_{1_{ik}}$  periods along parallel biases  $i$  for atoms with index  $k$ .
- $\tau_2$ , time-period for the path-sampling for  $N_R \times \tau_{2_{ik}}$  periods along parallel biases  $i$  for atoms with index  $k$ .
- $\gamma''_{ik}$ , fluctuation-width for the adaptive path-sampling algorithm to determine the deviation of the fluctuating bias around an average value  $\langle dL \rangle \approx 0$ . We used a constant value  $\gamma'' = 10^{-4}$
- $W$ , height of the Gaussians applied in the path-sampling. We applied a constant value  $W = 0.1$  kJ/mol ( $\Delta E = 1000$  kJ/mol).
- $\beta'_{md}$  and  $\beta'$ , renormalization parameters of the bias to the unbiased gradient (see the definitions of  $\alpha_{md}$  and  $\alpha'$ ).
- $\eta'_{md}$ , fluctuation range of the unbiased gradient (see the definitions of  $\alpha_{md}$  and  $\alpha'$ ).
- $\eta'$ , fluctuation range of the bias related to the unbiased gradient.
- The list of atom-indices per residue are added to a list with a corresponding equilibrium distance  $d_0$ . For the simulations of chemical crosslinking, we added all atom-indices for the first residue and one atom-index for the second residue.

### 3.7. Parameter Selection

In the novel sampling method along multiple time-scales, a system coupled with the parameters  $N_R = 5$ ,  $\tau_1 = 1$  ps,  $\tau_2 = 2.5$  ps is sampled with path dependent biases coupled to the times  $\tau_{1_{ik}} = \{1.0, 2.0, 3.0, 4.0, 5.0\}$  ps and  $\tau_{2_{ik}} = \{2.5, 5.0, 7.5, 10.0, 12.5\}$  ps. We used characteristic time-periods ranging from 10 ps to 1 ns. The 2 forms of the bias  $s$  are depending from 2 independent coupling times  $\tau_{1_{ik}}$  and  $\tau_{2_{ik}}$  ( $\tau_1$  and  $\tau_2$  corresponding to  $\tau_1$  for adaptive bias MD and the period  $\tau_2$  for the collection of path-sampling coordinates). For the sampling of a small and fast fluctuating system like dialanine with the 2 order parameters represented by its dihedral angles, the choice of the  $\tau$  parameters has to be sufficiently low to capture the dihedral rotations on the time-scale of less than 1 ps. The true dynamics of a full rotation are slower, but the collection of coordinates facilitating such a transition have to be on the order of less than 10 ps (5–7.5 fs). For the sampling of the folding dynamics of a small peptide, approximately identical parameters can be used. We anticipate that larger systems containing even slower degrees of freedom, such as domain rearrangements need also a larger number  $N_R$  in the range from 10 to 20. The 2 separate increments  $dL_{\sigma_{ik}}(L_{ik}(t))$  and  $dL_{ab_{ik}}$  are evaluated within 2 separate modules called adaptive bias MD and path-sampling, which we called hybrid path-sampling algorithm. The magnitude of the added bias  $\nabla H(B)$  derived from multiple path-dependent increments depends from the coupling parameters  $\beta'_{md}$ ,  $\beta'$  and the fluctuation parameters  $\eta'_{md}$ ,  $\eta'$ . Considering the fact that bonded interactions in the biomolecular forcefield can contribute with gradients  $> 10^4$  kJ/mol/nm, a coupling with a factor with a magnitude of  $\beta'_{md} = \beta' = 10^{-4}$  corresponds to the order of magnitude of typical non-bonded interactions.

### 3.8. Program and System Preparation

We implemented the method into the GROMACS-4.5.5 code in the routine `/src/kernel/md.c` [9]. The distance restraints used for the simulations are given to the program `gmx_mdrun` using a separate parsing routine. We used the AMBER99-SB forcefield for describing the interactions of the proteins and the SPC/E-water model for the explicit solvent simulations [125]. We applied Particle Mesh Ewald electrostatics with a cut-off of 1.0 nm. Van der Waals interactions have been calculated with the same cut-off. We used a time-step of 1.0 fs. The neighborlist has been updated every second time-step. We used the standard generalized Born solvent (GBSA) parameters for the implicit solvent simulations of TrpCage, where electrostatics were calculated using a twin-range cut-off of 1.0/1.2 nm (the long-ranged interaction calculated every second time-step). For the simulations of dialanine,

we centered the extended peptide (Ace-Ala-NMe) in a box with dimensions  $2.27 \times 2.27 \times 2.27 \text{ nm}^3$  and added 371 SPC/E waters. We applied  $\tau_1 = 7.5 \text{ ps}$ ,  $\tau_2 = 5.5 \text{ ps}$ ,  $N_R = 30$  and  $\beta_{md} = 1 \times 10^{-4}$ ,  $\beta' = 2.5 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 1.0$ ). We simulated the system at the 4 temperatures 300, 350, 400, and 450 K. For the explicit solvent simulations of TrpCage, we centered the extended peptide (NLYIQWLKDGGPSSGRPPPS [93]) 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. For the explicit solvent simulations of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A), we centered the experimental structure (PDB: 3m9z [117]) in a box with dimensions  $6.8260 \times 4.2300 \times 5.0470 \text{ nm}^3$  and filled the box with 6990 SPCE/E waters. We added 9 sodium ions to neutralize the system.

For the 2 simulations of TrpCage in implicit solvent we applied  $\tau_1 = 48.5 \text{ ps}$ ,  $\tau_2 = 25.5 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 1.0$ ), and  $\tau_1 = 4.85 \text{ ps}$ ,  $\tau_2 = 2.25 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 1.0$ ), and  $N_R = 5$  multiple biases (we applied the restraint data set from the data-set presented on [www.rcsb.org](http://www.rcsb.org) on the TrpCage protein [93]. See the supplementary information, Section 3). For the path-sampling simulation of TrpCage in explicit solvent without restraints, we applied  $\tau_1 = 2.50 \text{ ps}$ ,  $\tau_2 = 1.55 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 5.0$ ), and 5 multiple biases. For the path-sampling simulation of TrpCage in explicit solvent with restraints, we applied  $\tau_1 = 4.85 \text{ ps}$ ,  $\tau_2 = 2.55 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 1.0$ ), and  $N_R = 5$  multiple biases. For the simulations of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A) in implicit solvent, we applied  $\tau_1 = 48.0 \text{ ps}$ ,  $\tau_2 = 25.0 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 5.0$ ), and  $\tau_1 = 4.8 \text{ ps}$ ,  $\tau_2 = 2.5 \text{ ps}$ ,  $\beta'_{md} = \beta' = 1 \times 10^{-4}$  ( $\eta'_{md} = \eta' = 5.0$ ), and  $N_R = 5$  multiple biases (we applied the restraint data set from the data-set presented in ref. [89]. See the supplementary information). For the simulation of the Killer Cell Lectin-like Receptor Subfamily B Member 1A in explicit solvent, we applied  $\tau_1 = 2.8 \text{ ps}$ ,  $\tau_2 = 1.5 \text{ ps}$ ,  $\beta'_{md} = \beta' = 10^{-5}$  ( $\eta'_{md} = \eta' = 1.0$ ), and  $N_R = 5$  multiple biases.

## Analysis

For the analysis of the dielectric behavior of SPC/E and TIP3P water, we refer to the supplementary information. For the cluster analysis of the folding simulations of TrpCage, we used the *gmx\_cluster* module using an root-mean square deviation (RMSD) based clustering approach to the backbone of each conformer ( $RMSD_{C\alpha-C\alpha}$ ) with a value of 0.25 nm as criterion for one conformer to belong to one specific cluster [9,126]. We analyzed the transition time of the  $\Phi$ -angle of Dialanine peptide as the average time  $\tau_a$  needed to transition between the region  $\Phi < 0^\circ$  to the region  $\Phi > 0^\circ$ . We then determined the frequency using  $\nu_a = 1/\tau_a$  (see Figure 2b). For the determination of the activation energy  $E_a$ , we used the relations:

$$\nu_a = A_a \exp\left(-\frac{E_a}{RT}\right), \quad (34)$$

and

$$\ln \nu_a = \ln A_a - \frac{E_a}{RT}. \quad (35)$$

We defined the dihedral  $\Phi$  by the atoms: C-N-CA-C and  $\Psi$  using: N-CA-C-N of the Dialanine peptide. For the measurement of the free energies  $\Delta F$ , we used:

$$\Delta F = -K_B T \ln \frac{P}{P_{min}}, \quad (36)$$

where  $P_{min}$  stands for the minimal probability of the projection on 2 quantities. For the analysis of the fulfillment of a distance restraint, we determined the difference  $\Delta = d(t) - d_0$ , where  $d(t)$  stands for the time-dependent distance and  $d_0$  for the experimental reference value. We used the Equation (36) for the determination of the free energy differences for the probability to find the system along the reference distance between the actual distance and the experimental reference value. We used the

tools *gmx\_rms*, *gmx\_angle*, and *gmx\_mindist* for the calculation of the RMSD to the native structure, the dihedral angles, and the inter-atomic distances. For the determination of the linear acceleration factor  $\rho$ , we state that through the renormalization to the underlying Hamiltonian, the transition state for a transition of the system between 2 states remains unaffected, if the coupling factors  $\beta'_{md}$ ,  $\beta'$ ,  $\eta'_{md}$ , and  $\eta'$  remain within a range  $\ll 1$  [90]. If the transition state energy level  $\Delta E$  remains approximately unaffected by the applied bias [105,127], the resulting transition time  $t_{1,2}$  for a 2-state transition equals:

$$t_{1,2} = \frac{1}{r_t} \approx \frac{1}{r_u \times r_c}, \quad (37)$$

where  $r_u$  stands for the un-biased and  $r_c$  for the biased rate of the transition. We define a linear acceleration factor  $\rho$  which expresses the proportion of the un-biased time  $t_{1,2_u}$  to the time observed in the accelerated simulation  $t_{1,2}$  equals then:

$$\rho = \frac{r_c}{r_u} \approx \frac{t_{1,2_u}}{t_{1,2}}. \quad (38)$$

#### 4. Conclusions

In this article, we presented a methodology which defines a hybrid Hamiltonian  $H(C)$ , which combines an un-biased Hamiltonian  $H(A)$  with a biasing energy function  $H(B)$  derived from path-dependent variables. We combine the energy function  $H(B)$  with experimental distance restraints  $r$  to obtain a novel Hamiltonian  $H(B)_R$ . The energy function  $H(B)_R$  is adaptively reweighted in dependency of the angular overlap between the restraint vector and the experimental distance vector. The biasing function  $\nabla H(B)$  enhances the sampling along a number of  $N_R$  intrinsic pathways at multiple time-scales  $\tau_1$  (adaptive bias MD) and  $\tau_2$  (path-sampling) in order to capture all possible dynamic modes in the system. We validate the algorithm in simulations of SPC/E and TIP3P water and the conformational landscape of the dialanine dipeptide. We then apply the methodology to the sampling of experimental distance restraints derived from NMR-NOESY data or chemical crosslinking. In simulations of folding of TrpCage using NMR-NOESY experimental restraint data [93], we observe good agreement of the sampled conformation space with the experimental NMR-data. In a second example of restrained sampling of the Killer Cell Lectin-like Receptor Subfamily B Member 1A (NKR-P1A), we apply restraints from chemical crosslinking/mass spectrometry experiments [89]. In both examples, the conformers are in agreement with the applied restraint data. We conclude that the new sampling technique has the advantage of an accelerated sampling of any underlying system, while a minimal amount of *a priori* input is required. The adaptive restraint technique guides the system along experimental restraint vectors in a dynamic way leading to an ergodic sampling within the partition described by the underlying Hamiltonian. Only a small number of input- and no system-dependent order parameters are required for the setup of the simulation as for example in umbrella sampling techniques [94]. The method is broadly applicable and can be used to accelerate the sampling in coarse-grained simulations.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/20/2/370/s1>. See supplementary material for the simulations, where we investigated the impact of the enhanced sampling simulation parameters on the dielectric behavior of 2 water models and the restraint parameters used in the simulations of TrpCage and NKR-P1A.

**Author Contributions:** Conceptualization, E.K.P. and J.Č.; Methodology, E.K.P. and J.Č.; Software, E.K.P. and J.Č.; Validation, E.K.P. and J.Č.; Formal Analysis, E.K.P. and J.Č.; Investigation, E.K.P. and J.Č.; Resources, E.K.P. and J.Č.; Data Curation, E.K.P. and J.Č.; Writing—Original Draft Preparation, E.K.P. and J.Č.; Writing—Review & Editing, E.K.P. and J.Č.; Visualization, E.K.P. and J.Č.; Supervision, E.K.P. and J.Č.; Project Administration, E.K.P. and J.Č.; Funding Acquisition, J.Č.

**Funding:** This project was supported by the Institutional Research Project of the Institute of Biotechnology (RVO 86652036) and by projects from the European Regional Development Fund (BIOCEV CZ.1.05/1.1.00/02.0109) and the Czech National Infrastructure for Biological Data ELIXIR CZ (LM2015047 and CZ.02.1.01/0.0/0.0/16\_013/0001777).

**Acknowledgments:** Access to the computing and data storage facilities of MetaCentrum (LM2010005) is appreciated.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Weber, G. Ligand binding and internal equilibria in proteins. *Biochemistry* **1972**, *11*, 864–878. [[CrossRef](#)] [[PubMed](#)]
2. Baldwin, R.L. How Hofmeister ion interactions affect protein stability. *Biophys. J.* **1996**, *71*, 2056–2063. [[CrossRef](#)]
3. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)]
4. Lanucara, F.; Holman, S.W.; Gray, C.J.; Eyers, C.E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014**, *6*, 281–294. [[CrossRef](#)] [[PubMed](#)]
5. Billeter, M.; Wagner, G.; Wüthrich, K. Solution NMR structure determination of proteins revisited. *J. Biomol. NMR* **2008**, *42*, 155–158. [[CrossRef](#)] [[PubMed](#)]
6. Rose, P.W.; Prlić, A.; Bi, C.; Bluhm, W.F.; Christie, C.H.; Dutta, S.; Green, R.K.; Goodsell, D.S.; Westbrook, J.D.; Woo, J.; et al. The RCSB Protein Data Bank: Views of structural biology for basic and applied research and education. *Nucleic Acids Res.* **2015**, *43*, D345–D356. [[CrossRef](#)] [[PubMed](#)]
7. Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Clarendon Pr: Oxford, UK, 1987.
8. Adcock, S.A.; McCammon, J.A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, *106*, 1589–1615. [[CrossRef](#)] [[PubMed](#)]
9. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [[CrossRef](#)] [[PubMed](#)]
10. Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [[CrossRef](#)]
11. Brooks, B.; Brooks, C.; MacKerell, A.; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614. [[CrossRef](#)]
12. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [[CrossRef](#)] [[PubMed](#)]
13. Comer, J.; Gumbart, J.C.; Hémin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **2015**, *119*, 1129–1151. [[CrossRef](#)]
14. Shea, J.E.; Onuchic, J.; Brooks, C. Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12512–12517. [[CrossRef](#)]
15. Shea, J.E.; Brooks, C.L., III. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Phys. Chem. Rev.* **2001**, *52*, 499–535. [[CrossRef](#)] [[PubMed](#)]
16. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566. [[CrossRef](#)] [[PubMed](#)]
17. Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 5062–5067. [[CrossRef](#)] [[PubMed](#)]
18. Smiatek, J.; Heuer, A. Calculation of free energy landscapes: A histogram reweighted metadynamics approach. *J. Comput. Chem.* **2011**, *32*, 2084–2096. [[CrossRef](#)]
19. Giberti, F.; Savalaglio, M.; Parrinello, M. Metadynamics studies of crystal nucleation. *IUCr* **2015**, *2*, 256–266. [[CrossRef](#)]

20. Perego, C.; Savalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *J. Chem. Phys.* **2015**, *142*, 144113. [[CrossRef](#)]
21. Schug, A.; Wenzel, W.; Hansmann, U.H.E. Energy landscape paving simulations of the trp-cage protein. *J. Chem. Phys.* **2005**, *122*, 194711. [[CrossRef](#)]
22. Schug, A.; Herges, T.; Wenzel, W. Reproducible protein folding with the stochastic tunneling method. *Phys. Rev. Lett.* **2003**, *91*, 158102. [[CrossRef](#)] [[PubMed](#)]
23. Sørensen, M.R.; Voter, A.F. Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* **2000**, *112*, 9599. [[CrossRef](#)]
24. Montalenti, F.; Voter, A.F. Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics method. *J. Chem. Phys.* **2002**, *116*, 4819. [[CrossRef](#)]
25. Huber, T.; Torda, A.E.; van Gunsteren, W.F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput. Aided Mol. Des.* **1994**, *8*, 695–708. [[CrossRef](#)] [[PubMed](#)]
26. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919. [[CrossRef](#)] [[PubMed](#)]
27. Kong, X.; Brooks, C.L., III.  $\lambda$ -dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414. [[CrossRef](#)]
28. Knight, J.L.; Brooks, C.L., III. Multi-Site  $\lambda$ -dynamics for simulated Structure-Activity Relationship studies. *J. Chem. Theor. Comput.* **2011**, *7*, 2728–2739. [[CrossRef](#)]
29. Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **2000**, *329*, 261–270. [[CrossRef](#)]
30. Mitsutake, A.; Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J. Chem. Phys.* **2003**, *118*, 6664. [[CrossRef](#)]
31. Mitsutake, A.; Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. II. Application to a more complex system. *J. Chem. Phys.* **2003**, *118*, 6676. [[CrossRef](#)]
32. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058. [[CrossRef](#)]
33. Faller, R.; Yan, Q.; de Pablo, J.J. Multicanonical parallel tempering. *J. Chem. Phys.* **2002**, *116*, 5419. [[CrossRef](#)]
34. Whitfield, T.W.; Bu, L.; Straub, J.E. Generalized parallel sampling. *Phys. A* **2002**, *305*, 157–171. [[CrossRef](#)]
35. Jang, S.; Shin, S.; Pak, Y. Replica-exchange method using the generalized effective potential. *Phys. Rev. Lett.* **2003**, *91*, 058305. [[CrossRef](#)] [[PubMed](#)]
36. Liu, P.; Kim, B.; Friesner, R.A.; Berne, B.J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13749–13754. [[CrossRef](#)] [[PubMed](#)]
37. Liu, P.; Huang, X.; Zhou, R.; Berne, B.J. Hydrophobic aided replica exchange: An efficient algorithm for protein folding in explicit solvent. *J. Phys. Chem. B* **2006**, *110*, 19018–19022. [[CrossRef](#)]
38. Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified replica exchange simulation methods for local structure refinement. *J. Phys. Chem. B* **2005**, *109*, 8220–8230. [[CrossRef](#)]
39. Lyman, E.; Ytreberg, M.; Zuckerman, D.M. Resolution exchange simulation. *Phys. Rev. Lett.* **2006**, *96*, 028105. [[CrossRef](#)]
40. Liu, P.; Voth, G.A. Smart resolution replica exchange: An efficient algorithm for exploring complex energy landscapes. *J. Chem. Phys.* **2007**, *126*, 045106. [[CrossRef](#)]
41. Calvo, F. All-exchanges parallel tempering. *J. Chem. Phys.* **2005**, *123*, 124106. [[CrossRef](#)]
42. Rick, S.W. Replica exchange with dynamical scaling. *J. Chem. Phys.* **2007**, *126*, 054102. [[CrossRef](#)] [[PubMed](#)]
43. Kamberaj, H.; van der Vaart, A. Multiple scaling replica exchange for the conformational sampling of biomolecules in explicit water. *J. Chem. Phys.* **2007**, *127*, 234102. [[CrossRef](#)] [[PubMed](#)]
44. Brenner, P.; Sweet, C.R.; VonHandorf, D.; Izaguirre, J.A. Accelerating the replica exchange method through an efficient all-pairs exchange. *J. Chem. Phys.* **2007**, *126*, 074103. [[CrossRef](#)]
45. Zhang, C.; Ma, J. Simulation via direct computation of partition functions. *Phys. Rev. E* **2007**, *76*, 036708. [[CrossRef](#)]

46. Trebst, S.; Troyer, M.; Hansmann, U.H.E. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* **2006**, *124*, 174903. [[CrossRef](#)] [[PubMed](#)]
47. Bolhuis, P.G.; Dellago, C.; Chandler, D. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5877. [[CrossRef](#)] [[PubMed](#)]
48. Bello-Rivas, J.M.; Elber, R. Exact milestoning. *J. Chem. Phys.* **2015**, *142*, 094102. [[CrossRef](#)] [[PubMed](#)]
49. Ballard, A.J.; Jarzynski, C. Replica exchange with nonequilibrium switches. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12224–12229. [[CrossRef](#)]
50. Zhang, B.W.; Jasnow, D.; Zuckerman, D.M. The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* **2010**, *132*, 054107. [[CrossRef](#)]
51. a Beccara, S.; Škripić, T.; Covino, R.; Faccioli, P. Dominant folding pathways of a WW domain. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2330–2335. [[CrossRef](#)]
52. a Beccara, S.; Fant, L.; Faccioli, P. Variational scheme to compute protein reaction pathways using atomistic force fields with explicit solvent. *Phys. Rev. Lett.* **2015**, *114*, 098103. [[CrossRef](#)] [[PubMed](#)]
53. Elber, R. Computer simulations of long time dynamics. *J. Chem. Phys.* **2016**, *144*, 060901. [[CrossRef](#)] [[PubMed](#)]
54. Olender, R.; Elber, R. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *J. Chem. Phys.* **1996**, *105*, 9299–9315. [[CrossRef](#)]
55. Ma, Q.; Izaguirre, J.A. Targeted Mollified Impulse: A Multiscale Stochastic Integrator for Long Molecular Dynamics Simulations. *Multiscale Model. Simul.* **2003**, *2*, 1–21. [[CrossRef](#)]
56. Leimkuhler, B.; Margul, D.T.; Tuckerman, M.E. Molecular dynamics based enhanced sampling of collective variables with very large time steps. *Mol. Phys.* **2013**, *111*, 3579–3594. [[CrossRef](#)]
57. Richters, D.; Kühne, T.D. Linear-scaling self-consistent field theory based molecular dynamics: Application to C60buckyballs colliding with graphite. *Mol. Sim.* **2018**, *44*, 1380–1386. [[CrossRef](#)]
58. Chen, W.; Ferguson, A.L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018**, *39*, 2079–2102. [[CrossRef](#)] [[PubMed](#)]
59. Chiavazzo, E.; Covino, R.; Coifman, R.R.; Gear, C.W.; Georgiou, A.S.; Hummer, G.; Kevredikis, I.G. Intrinsic map dynamics exploration for uncharted effective free-energy landscapes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5494–E5503. [[CrossRef](#)]
60. Chen, M.; Yu, T.Q.; Tuckerman, M.E. Locating landmarks on high-dimensional free energy surfaces. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3235–3240. [[CrossRef](#)]
61. Calvo, F.; Doyle, J.P.K. Entropic tempering: A method for overcoming quasiergodicity in simulation. *Phys. Rev. E* **2000**, *63*, 010902. [[CrossRef](#)]
62. Morris-Andrews, A.; Rottler, J.; Plotkin, S.S. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.* **2010**, *132*, 035105. [[CrossRef](#)]
63. Ouldridge, T.E.; Louis, A.A.; Doyle, J.P.K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 085101. [[CrossRef](#)] [[PubMed](#)]
64. Naôme, A.; Laaksonen, A.; Vercauteren, D.P. A Solvent-Mediated Coarse-Grained Model of DNA Derived with the Systematic Newton Inversion Method. *J. Chem. Theor. Comput.* **2014**, *10*, 3541–3549. [[CrossRef](#)]
65. Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130–137. [[CrossRef](#)] [[PubMed](#)]
66. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936. [[CrossRef](#)]
67. Morris-Andrews, A.; Brown, F.L.; Shea, J.E. A Coarse-Grained Model for Peptide Aggregation on a Membrane Surface. *J. Phys. Chem. B* **2014**, *118*, 8420–8432. [[CrossRef](#)]
68. Monticelli, L.; Kandasamy, S.K.; Periole, X.; Larson, R.G.; Tieleman, D.P.; Marrink, S.J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theor. Comput.* **2008**, *4*, 819–834. [[CrossRef](#)]
69. Marrink, S.J.; Tieleman, D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822. [[CrossRef](#)]
70. Schwieters, C.D.; Bermejo, G.A.; Clore, G.M. Xplor-NIH for molecular structure determination from NMR and other data sources. *Prot. Sci.* **2018**, *27*, 26–40. [[CrossRef](#)]

71. Clore, G.M.; Gronenborn, A.M.; Kjaer, M.; Poulsen, F.M. The determination of the three-dimensional structure of barley serine proteinase inhibitor 2 by nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Prot. Eng.* **1987**, *1*, 305–311. [[CrossRef](#)]
72. Clore, G.M.; Nilges, M.; Sukumaran, D.K.; Brünger, A.T.; Karplus, M.; Gronenborn, A.M. The three-dimensional structure of alpha1-purothionin in solution: Combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **1986**, *5*, 2729–2735. [[CrossRef](#)]
73. Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **2010**, *18*, 923–933. [[CrossRef](#)] [[PubMed](#)]
74. Barducci, A.; Bonomi, M.; Parrinello, M. Linking well-tempered metadynamics simulations with experiments. *Biophys. J.* **2010**, *98*, L44–L46. [[CrossRef](#)] [[PubMed](#)]
75. Shen, R.; Han, W.; Fiorin, G.; Islam, S.M.; Schulten, K.; Roux, B. Structural Refinement of Proteins by Restrained Molecular Dynamics Simulations with Non-interacting Molecular Fragments. *PLoS Comput. Biol.* **2015**, *11*, e1004368. [[CrossRef](#)]
76. Islam, S.M.; Roux, B. Simulating the distance distribution between spin-labels attached to proteins. *J. Phys. Chem. B* **2015**, *119*, 3901–3911. [[CrossRef](#)]
77. Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6817–6822. [[CrossRef](#)]
78. Ma, T.; Zang, T.; Wang, Q.; Ma, J. Refining protein structures using enhanced sampling techniques with restraints derived from an ensemble-based model. *Prot. Sci.* **2018**, *27*, 1842–1849. [[CrossRef](#)]
79. Chen, P.-c.; Hub, J.S. Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data. *Biophys. J.* **2014**, *107*, 435–447. [[CrossRef](#)] [[PubMed](#)]
80. Hub, J.S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 18–26. [[CrossRef](#)]
81. Björling, A.; Niebling, S.; Marcellini, M.; van der Spoel, D.; Westenhoff, S. Deciphering solution scattering data with experimentally guided molecular dynamics simulations. *J. Chem. Theor. Comput.* **2015**, *11*, 780–787. [[CrossRef](#)]
82. Velásquez-Muriel, J.; Lasker, K.; Russel, D.; Phillips, J.; Webb, B.M.; Schneidmann-Duhovny, D.; Sali, A. Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 18821–18826. [[CrossRef](#)]
83. Park, H.; DiMaio, F.; Baker, D. The origin of consistent protein structure refinement from structural averaging. *Structure* **2015**, *23*, 1123–1128. [[CrossRef](#)] [[PubMed](#)]
84. Schmitz, U.; Ulyanov, N.B.; Kumar, A.; James, T.L. Molecular dynamics with weighted time-averaged restraints for a DNA octamer. Dynamic interpretation of nuclear magnetic resonance data. *J. Mol. Biol.* **1993**, *234*, 373–389. [[CrossRef](#)] [[PubMed](#)]
85. Scheek, R.M.; Torda, A.E.; Kemmink, J.; van Gunsteren, W.F. *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*; Plenum Press: New York, NY, USA, 1991; pp. 209–217.
86. Fennen, J.; Torda, A.E.; van Gunsteren, W.F. Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble. *J. Biomol. NMR* **1995**, *6*, 163–170. [[CrossRef](#)] [[PubMed](#)]
87. Hansen, N.; Heller, F.; Schmid, N.; van Gunsteren, W.F. Time-averaged order parameter restraints in molecular dynamics simulations. *J. Biomol. NMR* **2014**, *60*, 169–187. [[CrossRef](#)]
88. Merkley, E.D.; Rysavy, S.; Kahraman, A.; Hafen, R.P.; Daggett, V.; Adkins, J.N. Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Prot. Sci.* **2013**, *23*, 747–759. [[CrossRef](#)] [[PubMed](#)]
89. Rozbeský, D.; Man, P.; Kavan, D.; Chmelík, J.; Černý, J.; Bezouška, K.; Novák, P. Chemical cross-linking and H/D exchange for fast refinement of protein crystal structure. *Anal. Chem.* **2012**, *84*, 867–870. [[CrossRef](#)] [[PubMed](#)]
90. Peter, E.K.; Shea, J.E. An adaptive bias-hybrid MD/kMC algorithm for protein folding and aggregation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 17373–17382. [[CrossRef](#)]
91. Peter, E.K.; Černý, J. Enriched Conformational Sampling of DNA and Proteins with a Hybrid Hamiltonian Derived from the Protein Data Bank. *Int. J. Mol. Sci.* **2018**, *19*, 3405. [[CrossRef](#)]
92. Peter, E.K. Adaptive enhanced sampling with a path-variable for the simulation of protein folding and aggregation. *J. Chem. Phys.* **2017**, *147*, 214902. [[CrossRef](#)]

93. Neidigh, J.W.; Fesinmeyer, R.M. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430. [[CrossRef](#)]
94. Torrie, G.M.; Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation—Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [[CrossRef](#)]
95. Wang, H.; Junghans, C.; Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E* **2009**, *28*, 221–229. [[CrossRef](#)] [[PubMed](#)]
96. Izvekov, S.; Parrinello, M.; Burnham, C.J.; Voth, G.A. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.* **2004**, *120*, 10896–10913. [[CrossRef](#)] [[PubMed](#)]
97. Soper, A.K.; Phillips, M.G. A new determination of the structure of water at 25 °C. *Chem. Phys.* **1986**, *107*, 47. [[CrossRef](#)]
98. Mills, R. Self-diffusion in normal and heavy water in the range 1–45.deg. *J. Phys. Chem.* **1973**, *77*, 685. [[CrossRef](#)]
99. Price, W.S.; Ide, H.; Arata, Y. Self-Diffusion of Supercooled Water to 238 K Using PGSE NMR Diffusion Measurements. *J. Phys. Chem. A* **1999**, *103*, 448. [[CrossRef](#)]
100. Owen, B.B.; Miller, R.C.; Milner, C.E.; Cogan, H.L. The dielectric constant of water as a function of temperature and pressure. *J. Phys. Chem.* **1961**, *65*, 2065–2070. [[CrossRef](#)]
101. Braun, D.; Boresch, S.; Steinhauser, O. Transport and dielectric properties of water and the influence of coarse-graining: Comparing BMW, SPC/E, and TIP3P models. *J. Chem. Phys.* **2014**, *140*, 064107. [[CrossRef](#)]
102. Tobias, D.J.; Brooks, C.L., III. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. *J. Phys. Chem.* **1992**, *96*, 3864–3870. [[CrossRef](#)]
103. Swope, W.C.; Pitera, J.W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B.G.; Germain, R.S.; Rayshubski, A.; Zhestkov, Y.; Zhou, R. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a  $\beta$ -Hairpin Peptide. *J. Chem. Phys. B* **2004**, *108*, 6582–6594. [[CrossRef](#)]
104. Stelzl, L.S.; Hummer, G. Kinetics from Replica Exchange Molecular Dynamics Simulations. *J. Chem. Theor. Comput.* **2017**, *13*, 3927–3935. [[CrossRef](#)]
105. Tiwary, P.; Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602. [[CrossRef](#)]
106. Peter, E.K.; Shea, J.E. A hybrid MD-kMC algorithm for folding proteins in explicit solvent. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6430–6440. [[CrossRef](#)] [[PubMed](#)]
107. Peter, E.K.; Pivkin, I.V.; Shea, J.E. A kMC-MD method with generalized move-sets for the simulation of folding of  $\alpha$ -helical and  $\beta$ -stranded peptides. *J. Chem. Phys.* **2015**, *142*, 144903. [[CrossRef](#)] [[PubMed](#)]
108. Culik, R.M.; Serrano, A.L.; Bunagan, M.R.; Gai, F. Achieving secondary structural resolution in kinetic measurements of protein folding: A case study of the folding mechanism of Trp-cage. *Angew. Chem.* **2011**, *123*, 11076–11079. [[CrossRef](#)]
109. Meuzelaar, H.; Marino, K.A.; Huerta-Viga, A.; Panman, M.R.; Smeenk, L.E.J.; Kettelarij, A.J.; van Maarseveen, P.T.J.H.; Bolhuis, P.G.; Woutersen, S. Folding Dynamics of the Trp-Cage Miniprotein: Evidence for a Native-Like Intermediate from Combined Time-Resolved Vibrational Spectroscopy and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 11490–11501. [[CrossRef](#)]
110. Juraszek, J.; Bolhuis, P.G. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15859–15864. [[CrossRef](#)] [[PubMed](#)]
111. Juraszek, J.; Bolhuis, P.G. Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys. J.* **2008**, *95*, 4246–4257. [[CrossRef](#)] [[PubMed](#)]
112. Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5*, e1000452. [[CrossRef](#)]
113. Snow, C.D.; Zagrovic, B.; Pande, V.S. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2002**, *124*, 14548–14549. [[CrossRef](#)] [[PubMed](#)]
114. Ren, H.; Lai, Z.; Biggs, J.D.; Wang, J.; Mukamel, S. Two-dimensional stimulated resonance Raman spectroscopy study of the Trp-cage peptide folding. *Phys. Chem. Chem. Phys.* **2013**, *15*, 19457–19464. [[CrossRef](#)] [[PubMed](#)]
115. Neuweiler, H.; Doose, S.; Sauer, M. A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16650–16655. [[CrossRef](#)] [[PubMed](#)]
116. Qiu, L.; Pabit, S.A.; Roitberg, A.E.; Hagen, S.J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4  $\mu$ s. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953. [[CrossRef](#)]

117. Kolenko, P.; Rozbeský, D.; Vaněk, O.; Kopecký, V.; Hofbauerová, K.; Novák, P.; Pompach, P.; Hašek, J.; Skálová, T.; Bezouška, K.; et al. Molecular architecture of mouse activating NKR-P1 receptors. *J. Struct. Biol.* **2011**, *175*, 434–441. [[CrossRef](#)] [[PubMed](#)]
118. Rozbeský, D.; Adámek, D.; Pospíšilová, E.; Novák, P.; Chmelík, J. Solution structure of the lymphocyte receptor Nkrp1a reveals a distinct conformation of the long loop region as compared to in the crystal structure. *Proteins* **2016**, *84*, 1304–1311. [[CrossRef](#)] [[PubMed](#)]
119. Skálová, T.; Kotýnková, K.; Dušková, J.; Hašek, J.; Koval', T.; Kolenko, P.; Novák, P.; Man, P.; Hanč, P.; Vaněk, O.; et al. Mouse Clr-g, a ligand for NK cell activation receptor NKR-P1F: Crystal structure and biophysical properties. *J. Immunol.* **2012**, *189*, 4881–4889. [[CrossRef](#)] [[PubMed](#)]
120. Bernhardt, N.A.; Xi, W.; Wang, W.; Hansmann, U.H.E. Simulating Protein Fold Switching by Replica Exchange with Tunneling. *J. Chem. Theory Comput.* **2016**, *12*, 5656–5666. [[CrossRef](#)]
121. Kleinert, H. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 5th ed.; World Scientific: Singapore, 2009; pp. 1–1547.
122. Feynman, R.; Hibbs, A.R. *Quantum Mechanics and Path Integrals*; MacGraw Hill Companies: New York, NY, USA, 1965.
123. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603. [[CrossRef](#)]
124. Balsera, M.A.; Wriggers, W.; Ono, Y.; Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996**, *100*, 2567–2572. [[CrossRef](#)]
125. Kollman, P.A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Acc. Chem. Res.* **1996**, *29*, 461–469. [[CrossRef](#)]
126. Hess, B.; van der Spoel, D.; Lindahl, E. The Gromacs Development Team. 2012. Available online: [www.gromacs.org](http://www.gromacs.org) (accessed on 12 January 2018).
127. Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E* **1995**, *52*, 2893. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).