

Chapter 12

Recipes for Free Energy Calculations in Biomolecular Systems

**Mahmoud Moradi, Volodymyr Babin, Celeste Sagui,
and Christopher Roland**

Abstract

During the last decade, several methods for sampling phase space and calculating various free energies in biomolecular systems have been devised or refined for molecular dynamics (MD) simulations. Thus, state-of-the-art methodology and the ever increasing computer power allow calculations that were forbidden a decade ago. These calculations, however, are not trivial as they require knowledge of the methods, insight into the system under study, and, quite often, an artful combination of different methodologies in order to avoid the various traps inherent in an unknown free energy landscape. In this chapter, we illustrate some of these concepts with two relatively simple systems, a sugar ring and proline oligopeptides, whose free energy landscapes still offer considerable challenges. In order to explore the configurational space of these systems, and to surmount the various free energy barriers, we combine three complementary methods: a nonequilibrium umbrella sampling method (adaptively biased MD, or ABMD), replica-exchange molecular dynamics (REMD), and steered molecular dynamics (SMD). In particular, ABMD is used to compute the free energy surface of a set of collective variables; REMD is used to improve the performance of ABMD, to carry out sampling in space complementary to the collective variables, and to sample equilibrium configurations directly; and SMD is used to study different transition mechanisms.

Key words: Free energy, Replica exchange, Proline, Carbohydrate

1. Introduction

During the last decade, new algorithms combined with advances in sampling techniques and increasing computational power have enabled both accurate and robust (within a given model) calculations of molecular free energies. One can expect that future biomolecular simulations will make use of these new free energy methods and related technology with ever increasing frequency. The purpose of this chapter is to briefly review three complementary simulation techniques in current use for calculating biomolecular free energies

and to illustrate their working with two prototypical examples. Specifically, we will cover the adaptively biased molecular dynamics (ABMD) (1), replica-exchange molecular dynamics (REMD) (2), and steered molecular dynamic (SMD) (3) methods, as applied to the puckering of sugars and the conformational space of short polyproline peptides (4–6). We proceed to the examples after a brief overview of the methodology.

Consider a complex multi-particle system. In order to investigate its equilibrium properties, it is often convenient to identify a suitable *collective variable* $\sigma = \sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)$, which is understood to be a smooth function of the particle positions $\sigma : \mathbf{r}_1, \dots, \mathbf{r}_N \mapsto \mathbb{Q}$, and then to study its equilibrium probability density:

$$p(\xi) = \langle \delta[\xi - \sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)] \rangle, \quad \xi \in \mathbb{Q}. \quad (1)$$

Here, the angular brackets indicate an ensemble average. This probability density $p(\xi)$ provides useful information as to the relative stability of the states corresponding to different values of ξ (7). In practice the Landau free energy or *potential of mean force* (PMF)

$$f(\xi) = -k_B T \log p(\xi) \quad (2)$$

(here T is temperature and k_B the Boltzmann constant) is typically preferred over $p(\xi)$. Either $p(\xi)$ or $f(\xi)$ provides for a reduced-dimensionality description of the system in terms of the collective variable ξ with other degrees of freedom integrated out. Typically, the collective variable is chosen in such a way as to bring out the relevant physics in a natural and transparent manner. It is often associated with the slowest degrees of freedom of the original system, although this is not a formal requirement.

In most circumstances, the direct computation of the free energy from an equilibrium trajectory is impractical, because the sampling is limited to the vicinity of the local minima of $f(\xi)$. The problem may be resolved by means of *umbrella sampling* methods (8), in which the trajectory is biased in order to explore the desired range of ξ . Later, this bias is “undone,” and the free energy is recovered. Umbrella sampling schemes in combination with the so-called weighted histogram analysis method (WHAM) (9–11) offer a practical route to computing free energies. Most recently, free energy methods based on biased nonequilibrium simulations have become popular. These methods all estimate the PMF from an evolving trajectory (12, 13) and use that estimate to modify the system dynamics so as to ultimately flatten the effective free energy surface. Methods that explore this idea include the local elevation method (LEM) (14), the Wang-Landau approach (15), the non-equilibrium metadynamics (16, 17), etc. Collectively, all these methods may be considered as umbrella sampling in which the biasing potential eventually reproduces the initially unknown free energy surface in the long time limit, as illustrated in Fig. 1.

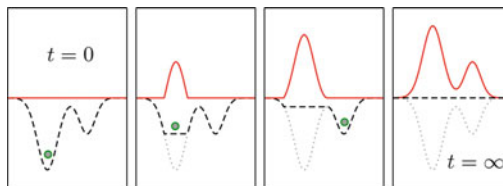


Fig. 1. A schematic illustration for the time-dependent biasing potential used to estimate an unknown free energy surface. Here, the *dotted line* represents the unknown free energy of the system as a function of a suitable collective variable. The solid line (*red*) represents the evolving biasing potential, and the *dashed line* the “effective” free energy of the system. The latter is given by the sum of the unknown free energy and the biasing potential. Initially, at $t = 0$, the effective free energy coincides with the free energy of the system. As time evolves towards $t \rightarrow \infty$, the effective free energy of the system is flattened and the biasing potential gives the negative of the free energy.

2. The Adaptively Biased Molecular Dynamics Method

The ABMD method is formulated in terms of the following equations:

$$m_a \frac{d^2 \mathbf{r}_a}{dt^2} = \mathbf{F}_a - \frac{\partial}{\partial \mathbf{r}_a} U[t|\sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)], \quad (3)$$

$$\frac{\partial U(t|\xi)}{\partial t} = \frac{k_B T}{\tau_F} G[\xi - \sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)], \quad (4)$$

where the first one contains the Newton’s equations that govern ordinary molecular dynamics (MD) (temperature and pressure regulation terms are not shown), augmented with an additional force coming from the time-dependent biasing potential $U(t|\xi)$ (with $U(t=0|\xi) = 0$), whose time evolution is given by the second equation. In the following, we refer to τ_F as *flooding time scale* and to $G(\xi)$ as the *kernel* (in analogy to the *kernel density estimator* widely used in statistics (18)). The kernel should be positive definite ($G(\xi) > 0$) and symmetric ($G(-\xi) = G(\xi)$). It may be viewed as a smoothed Dirac delta function. For large enough τ_F and small enough kernel width, the biasing potential $U(t|\xi)$ converges toward $-\beta\phi(\xi)$ as $t \rightarrow \infty$ (12, 13) (see Fig. 1).

Our numerical implementation of the ABMD method involves the use of cubic B-splines (19) (or products thereof) in conjunction with a biweight kernel for $G(\xi)$ (18). The method has moderate storage requirements and does not introduce any noticeable performance penalty. In addition, ABMD is characterized by only *two* control parameters: the flooding time scale τ_F and the kernel width $4\Delta\xi$.

We have also implemented two simple extensions to ABMD. The first is identical in spirit to the *multiple walkers metadynamics* (12, 20).

It amounts to carrying out several different MD simulations biased by the same $U(t|\xi)$, which evolves via

$$\frac{\partial U(t|\xi)}{\partial t} = \frac{k_B T}{\tau_F} \sum_{\alpha} G[\xi - \sigma(\mathbf{r}_1^{\alpha}, \dots, \mathbf{r}_N^{\alpha})], \quad (5)$$

where α labels different MD trajectories. A second extension is to gather several different MD trajectories, each bearing its own biasing potential and, if desired, its own distinct collective variable, into a generalized ensemble for “replica exchange” with modified “exchange” rules (21–23). Both extensions are advantageous and lead to a more uniform flattening of $U(t|\xi) + f(\xi)$. This enhanced convergence to $f(\xi)$ is due to the improved sampling of the “evolving” canonical distribution.

We have implemented the ABMD method in the AMBER package (24), with support for both replica exchange and multiple walkers. In pure “parallel tempering” replica exchange (same collective variable for all replicas), N_r replicas are simulated at different temperatures T_n , $n = 1, \dots, N_r$. Each replica has its own biasing potential $U^n(t|\xi)$, $n = 1, \dots, N_r$, that evolves according to its own dynamical equation. Exchanges between neighboring replicas are attempted at a prescribed rate, with an exchange probability given by (21)

$$w(m|n) = \begin{cases} 1, & \Delta \leq 0, \\ \exp(-\Delta), & \Delta > 0, \end{cases} \quad (6)$$

$$\begin{aligned} \Delta = & \left(\frac{1}{k_B T_n} - \frac{1}{k_B T_m} \right) (E_p^m - E_p^n) \\ & + \frac{1}{k_B T_m} [U^m(\xi^n) - U^m(\xi^m)] \\ & - \frac{1}{k_B T_n} [U^n(\xi^n) - U^n(\xi^m)], \end{aligned} \quad (7)$$

where E_p denotes the atomic potential energy. The biasing potentials are temperature-bound and converge in the $t \rightarrow \infty$ limit to the free energies at their respective temperatures.

In a *general* replica-exchange scheme, we allow different replicas to have different collective variables and/or temperatures. We also allow for either an evolving or a static biasing potential (the latter obviously includes the case of $U^n(t|\xi) = 0$) on a per-replica basis (when all biasing potentials are static, the method reduces to a particular variant of the “Hamiltonian Replica Exchange” described in (21)). Exchanges between random pairs of replicas are then attempted at a prescribed rate.

Finally, we note that in order to improve the accuracy of the free energies as calculated with the ABMD method, follow-up equilibrium umbrella sampling runs may be necessary. Such a procedure is very much in the spirit of adaptive umbrella sampling (25)

and makes use of the biasing potential $U(\xi)$ as is (26). With such runs, one calculates the biased probability density:

$$p^B(\xi) = \langle \Delta[\xi - \sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)] \rangle_B. \quad (8)$$

The idea here is that if, as a result of an ABMD run, $f(\xi) + U(\xi)$ is a constant, then the biased probability density $p^B(\xi)$ will be flat. In a typical simulation, however, this is not the case. Using the results of the follow-up equilibrium runs, one can correct the calculated free energy surface via:

$$f(\xi) = -U(\xi) - k_B T \ln p^B(\xi). \quad (9)$$

This corrective stage can also involve many independent equilibrium runs whose outcomes are “joined” using WHAM (9, 10). Finally, we note that approximate biasing potentials computed by the ABMD may be utilized for enhanced unbiased equilibrium sampling through the replica-exchange scheme (5, 27). All the samples generated by such replica-exchange runs can also be fed to WHAM to “correct” the free energies.

3. Replica-Exchange Schemes Utilizing Approximate Free Energies

Although the free energy maps can be used to characterize the properties of a system in terms of one or more collective variables, it is often desirable to sample unbiased equilibrium configurations directly. In this case, the approximate free energies computed via ABMD may be used to set up the Hamiltonian replica-exchange molecular dynamics (HREMD) simulations for enhanced equilibrium sampling (27). Typically, several replicas running at the lowest temperature T_0 , are introduced into the setup. One of the replicas is completely unbiased and therefore samples the unbiased Boltzmann distribution at $T = T_0$. The other replicas are subject to a reduced bias at T_0 . These “proxy” replicas ensure adequate exchange rates and hence enhance the mixing.

In the most general case, the different replicas enhance the crossing of barriers by means of a temperature ladder and/or biasing potentials: the former is not specific to any particular mode of the system, while the latter allows to emphasize some modes (e.g., cis/trans transitions of the prolyl bonds for the polyproline system to be discussed later). In many cases, both devices may be necessary for robust REMD simulations. The ergodic time scale of a REMD simulation is largely determined by two factors: the efficiency of the random walk between the replicas, and the fastest ergodic time scale present among the individual replicas. The first factor can be described in terms of the exchange rates between the pairs of replicas and, as long as the rates are sufficiently high, need not be considered. The replica characterized by the fastest

ergodic time scale is usually referred to as the “hottest” one. Typically, it is the replica running at the highest temperature. Its mixing properties can be assessed by simulating it alone, making sure that it is ergodic, and that its autocorrelation time is sufficiently small so as to generate enough independent samples in a feasible runtime. In practical setting, the number of replicas is typically limited, and hence, the “hottest” replica must be constructed with care. In principle it is always possible to use just the temperature alone (however, the higher the temperature of the “hottest” replica, the more replicas are needed to span the temperatures down to lowest T_0) or to use the biasing potentials alone (running all the replicas at the same temperature). The second scenario is not always practical, as it is considerably less generic than the first one. A combination of Hamiltonian and temperature REMD methods (HT-REMD (5, 28)) could provide a practical way to reduce the computational costs of the REMD sampling, as it permits one to facilitate the sampling in the “hottest” replica by using both means, thus allowing for better “tuning” of the entire setup.

4. Steered Molecular Dynamics

The SMD (3), is particularly useful for examining select pathways and mechanisms between two equilibrium states. It is based on the following notions. Consider a thermodynamic process that changes the system by means of altering a control parameter ξ , which is changed from ξ_0 to ξ_t over a time interval t . Because of the second law of thermodynamics, the average work W performed on the system cannot be less than the free energy difference:

$$\Delta F = f(\xi_t) - f(\xi_0) \leq W, \quad (10)$$

with the equality holding if the process is reversible. Thus, a nonequilibrium process can only provide us with an upper bound for ΔF . However, the Jarzynski (29) equality

$$\exp\left(-\frac{\Delta F}{k_B T}\right) = \left\langle \exp\left(-\frac{W}{k_B T}\right) \right\rangle \quad (11)$$

holds even when the transformation is irreversible (the angular brackets denote an average over all possible realizations of an external process that takes the system from its initial equilibrium state to a nonequilibrium state under the same external conditions as that of the final equilibrium state). Thus, in principle, ΔF can be estimated using *Nfinite-timenonequilibrium* simulations:

$$\Delta F \approx -k_B T \ln \left[\frac{1}{N} \sum_{i=1}^N \exp\left(-\frac{W_n}{k_B T}\right) \right]. \quad (12)$$

Here, W_n represents the work obtained for the n th realization of the process. Unfortunately, because the average of the exponential function is dominated by rare realizations, convergence is often poor unless N is very large. It is, however, possible to improve the estimates of the calculated free energy differences using modifications based on the maximum likelihood estimator (MLE) and Bennett's acceptance ratio (BAR) method, which are equivalent (30–32). These methods make use of both forward (F) and reverse (R) nonequilibrium pulling processes.

In practical implementations of the SMD method, a restraining harmonic potential is added to the potential energy:

$$U_{\text{SMD}}(\xi) = \frac{K}{2} [\sigma(\mathbf{r}_1, \dots, \mathbf{r}_N) - \xi]^2, \quad (13)$$

so that the system can be “steered” toward the states with the desired value of the collective variable $\sigma(\mathbf{r}_1, \dots, \mathbf{r}_N)$ by “pulling” on the ends of the spring $\xi(t)$. This harmonic potential should be evaluated in the stiff-spring limit, with $K \gtrsim U_{\text{max}}/(\Delta\xi)^2$, where U_{max} is the highest PMF barrier to be explored and $\Delta\xi$ sets the desired precision in ξ . This stiff-spring approximation ensures that the PMF as a function of ξ of the unbiased original system is well represented by the free energy of the system with the harmonic restraint. The latter can be conveniently thought of as an approximation of the delta function $\Delta(\xi - \sigma(\mathbf{r}_1, \dots, \mathbf{r}_N))$ in the partition function. Generally, a number of nonequilibrium “pulling” and “pushing” simulations are then carried out by changing the values of ξ over a prescribed pathway. The numerical values of the nonequilibrium work for the forward (W^F) and reverse (W^R) pullings are collected, and ΔF is estimated from the transcendental equation (30–32):

$$\sum_{i=1}^{n_F} \frac{1}{1 + n_F/n_R \exp(W_i^F - \Delta F)} - \sum_{i=1}^{n_R} \frac{1}{1 + n_R/n_F \exp(W_i^R + \Delta F)} = 0, \quad (14)$$

with $n_{F,R}$ denoting the numbers of forward and reverse simulations, respectively. As expected, we find that ΔF calculated with this two-sided method displays an improved convergence over estimates obtained with the straightforward one-sided application of the Jarzynski equality.

The SMD method is especially easy to implement with one-dimensional collective variables: one simply repeats the same run with different initial configurations and/or random seeds, and collects the work. The Jarzynski equality (or MLE-based formula if both forward and reverse runs are performed) is then used to estimate the free energy difference. Turning to systems described by more than one collective variable, the formalism additionally

requires one to average over all possible paths that join the points $\xi(0)$ and $\xi(t)$. This is rather difficult to achieve for larger systems, and therefore, only selected pathways are typically examined in a practical setting. For these systems, SMD can be also used as a tool to examine and compare pathways associated with different transition mechanisms in a *qualitative* way. One can always “handpick” pathways that appear to be the most interesting. Particular attention is typically given to the ones that require the least amount of work, since these are mostly to dominate the sum for ΔF in the Jarzynski equality. Another approach for investigating mechanisms is to steer the system along selected pathways over the two-dimensional free energy landscapes, which have been selected by other means such as the least free energy path (LFEP) method (33).

5. Examples

Having described the free energy methods and approaches in the previous sections, we illustrate their utility by considering two prototypical examples, involving the puckering of a select sugar molecule, and the conformational properties of several proline oligopeptides. All the computations were performed using the AMBER 10 (24) simulation package.

5.1. Puckering of β -D-Fructose

Here, we consider the calculation of the free energy landscape which characterizes the puckering of β -D-fructofuranose (fructose), shown in Fig. 2.

As per instruction figures should be in black and white but color ‘grey’ is mentioned in Fig. 2. Please check.

Computational studies of carbohydrates have evolved along several directions. Considerable work has been devoted to quantum chemical studies of the energetics of relatively small compounds (see, e.g., (34) and references therein). While very accurate, such *ab initio* calculations cannot typically handle explicit solvent over nanosecond time scale. Yet, it is known that the solvation plays a crucial role in the conformational preferences of carbohydrates (35, 36). Until recently, only coarse-grained models were able to breach the long time scales involved in the conformational sampling.

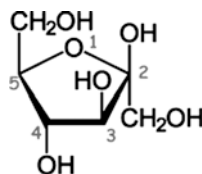


Fig. 2. Structure of the β -D-fructofuranose molecule. The atoms on the ring are numbered in grey, and used to compute the puckering coordinates as discussed.

However, methodological advances coupled to growing computational power have allowed the realization of all-atom, explicit solvent classical molecular dynamics over hundreds of nanoseconds (37). Naturally, there is a growing interest for development and improvement of the force fields (38). An accurate determination of the properties of the compounds in explicit solvent (39) is indispensable for the validation of such model potentials, and also interesting in its own right.

5.1.1. Collective Variables

The puckering states of a five-membered ring can be conveniently described in terms of just two numbers q_c and q_s (40) defined as follows. Consider the positions of the atoms of the ring $\mathbf{R}_1, \dots, \mathbf{R}_5$ translated to their geometrical center so that

$$\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 + \mathbf{R}_4 + \mathbf{R}_5 = 0.$$

One computes the vector \mathbf{n} normal to the plane of the ring:

$$\begin{aligned}\mathbf{R}_c &= \sum_{a=1}^5 \mathbf{R}_a \cos \left[\frac{2\pi(a-1)}{5} \right], \\ \mathbf{R}_s &= \sum_{a=1}^5 \mathbf{R}_a \sin \left[\frac{2\pi(a-1)}{5} \right], \\ \mathbf{n} &= \frac{\mathbf{R}_s \times \mathbf{R}_c}{|\mathbf{R}_s \times \mathbf{R}_c|},\end{aligned}$$

and uses it to calculate the displacements $z_a = \mathbf{n} \cdot \mathbf{R}_a$, $a = 1, \dots, 5$ for the vertices of the ring out of the plane. The displacements z_a are subject to three constraints (40), so that for the five-membered ring, only two independent combinations can be constructed. The puckering coordinates (q_c , q_s) used in this study are then defined as

$$q_c = \sqrt{\frac{2}{5}} \sum_{a=1}^5 z_a \cos \left[\frac{4\pi(a-1)}{5} \right] \quad (15)$$

$$q_s = \sqrt{\frac{2}{5}} \sum_{a=1}^5 z_a \sin \left[\frac{4\pi(a-1)}{5} \right] \quad (16)$$

It is customary (40) to use the puckering amplitude Q and phase angle φ instead of the coordinates (q_c , q_s) (i.e., $q_c = Q \cos \varphi$ and $q_s = Q \sin \varphi$). We chose to retain the coordinate representation because it is somewhat easier to deal with numerically.

5.1.2. Simulation Details

We used the GLYCAM 06 (38) parameters to model the fructose molecule along with the TIP3P (41) water model for explicit solvent simulations and the GB/SA (42–45) approximation for the implicit solvent simulations. The integrator time step was 1 fs. The Langevin thermostat (46) with collision frequency $\gamma = 1 \text{ ps}^{-1}$ was used to keep the temperature at 300 K. The lengths of all bonds containing

hydrogen were fixed via the SHAKE algorithm with the tolerance set to 10^{-6} . The GB/SA simulations were implemented using no cutoff on the nonbonded (van der Waals and electrostatics) interactions. For the explicit water simulations, we used periodic boundary conditions with fixed box size (30.5\AA in all three directions) chosen to render the density equal to $\approx 1\text{ g/cm}^3$. The particle mesh Ewald (PME) (47, 48) method was used with the direct space cutoff set to 9\AA and a $32 \times 32 \times 32$ grid for the smooth part of the Ewald sum. In all instances, the systems were thoroughly equilibrated, and new random number generator seeds were set for every run.

5.1.3. Calculation of the Free Energy Landscape

Simulations were started in implicit GB/SA solvent with “coarse” flooding given by $4\Delta q_s = 4\Delta q_c = 0.12\text{\AA}$ and $\tau = 1\text{ ps}$ for 1 ns . This short run turned out to be enough to cover the relevant region in the (q_c, q_s) space as shown in Fig. 3. Although the nonequilibrium nature of the run influences the accuracy of the resulting free energy in a rather complicated way, it is expected that the closer to equilibrium the system is, the more accurately the real, equilibrium free energy is approximated by the slowly growing biasing potential. Therefore, we carried out additional 5 ns long run using the same “spatial” resolution ($\Delta q_{c,s}$) but flooding the system at a slower pace with $\tau = 100\text{ ps}$. Again, it turned out that such run was long enough to repeatedly visit all the relevant points of the phase space, thereby correcting for the inaccuracies associated with the initial run.

In a typical ABMD calculation, the biasing potential “grown” in vacuum or with an implicit solvent generally becomes the starting point for considerably more demanding explicit solvent simulations. One generally continues with a “fine” flooding under explicit solvent and then uses these results for the corrective equilibrium sampling. For this particular system, however, we found only very minor differences between the implicit and explicit solvent results. We could therefore safely skip the flooding under the explicit solvent step and proceed directly with 100 ns long biased equilibrium run in

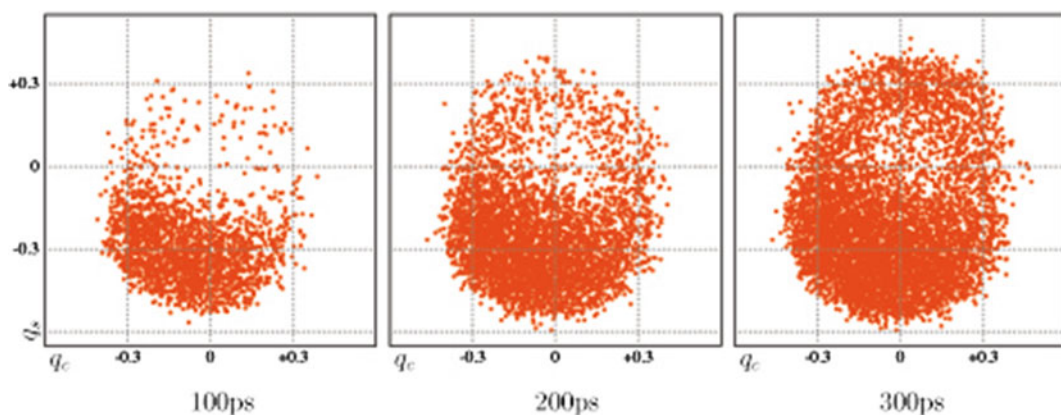


Fig. 3. Snapshots marking the progression of the ABMD simulation in the (q_c, q_s) plane.

explicit water. We collected the samples of (q_c, q_s) every 0.1 ps. While this is way too often for the biased density reconstruction, it allowed us to accurately compute the autocorrelation functions $q_{c,s}(0)q_{c,s}(t)$ of the puckering coordinates. The autocorrelations, as expected, decay exponentially, with a characteristic time of 10 ps for water (1 ps for GB/SA, 100 ps for ethanol). With these results, we used samples separated by 10 ps to reconstruct the biased *equilibrium* probability density as described in (18):

$$p^B(q_c, q_s) = \frac{1}{N_S h_c h_s} \sum_{l=1}^{N_S} K\left(\frac{q_c - q_c^l}{h_c}\right) K\left(\frac{q_s - q_s^l}{h_s}\right), \quad (17)$$

where $N_S = 10,000$ is the number of samples (q_c^l, q_s^l) , $l = 1, \dots, N_S$, $h_{c,s}$ are the window widths, and $K(\xi)$ are the kernels. We used Gaussian kernels

$$K(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\xi^2\right] \quad (18)$$

for both q_c and q_s . The smoothing widths $h_{c,s}$ along with the number of samples determine the precision of the resulting estimator. It is important to understand that the density is an infinite-dimensional object and hence cannot be reconstructed from any finite number of samples; the nonzero widths $h_{c,s}$ effectively “discretize” the density, i.e., they set the scale below which the density remains unknown. To estimate the widths, we followed Silverman’s guidelines (p. 56, (18)): we examined the plots of the Laplacian of the density for several decreasing values of the widths “by eye”; below some “optimal” choice of $h_{c,s}$ the Laplacian of the reconstructed density becomes noisy; values of $h_{c,s}$ slightly larger are then deemed to be “optimal.” In our particular case, we settled at $h_c = h_s = 0.1\text{\AA}$. Having the $h_{c,s}$ we computed the estimator using the first 5,000 samples and the last 5,000 samples and compared these two with the one based on all 10,000 samples: the absolute difference in the region $q_c^2 + q_s^2 < (3/5)^2$ did not exceed 0.1; this translates into an *at most* 0.1 kcal/mol difference between the corrections computed using the whole sample and either of its halves. We therefore conclude that the final free energy plot that is shown in Fig. 4 is accurate to at least 0.1 kcal/mol. The global minimum in the Fig. 4 is at the pseudo-rotation angle $\varphi = \tan^{-1} q_s/q_c \approx 250^\circ$ which corresponds to the E₃ (C3'-*exo*) conformation.

5.2. Proline Oligopeptides

Proline is the only natural amino acid in which the side chain is cyclized to the backbone, setting the backbone dihedral angle to $\phi = -75^\circ$. This restricts proline’s conformational freedom considerably, as compared to other amino acids. Polyproline peptides exhibit two well-characterized helical structures: (a) a left-handed polyproline helix (PPII), with all peptide bonds in the trans-isomer conformation (i.e., backbone dihedral angle $\omega = 180^\circ$), and (b) a more compact right-handed polyproline helix (PPI), with all the

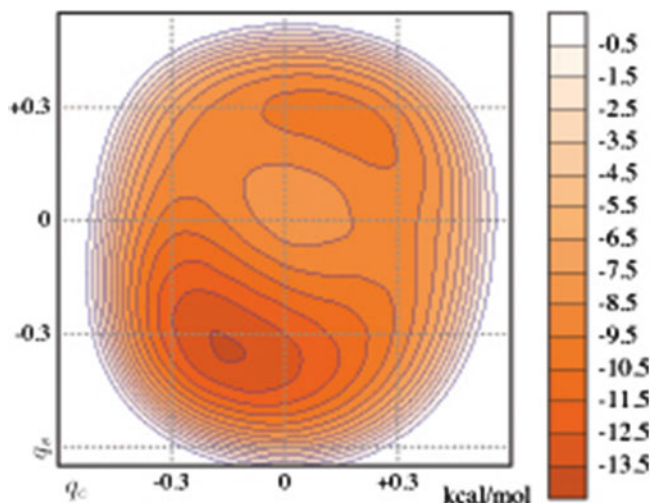


Fig. 4. Free energy landscape of β -D-fructofuranose in terms of the puckering coordinates (q_c , q_s) solvated by 920 TIP3P water molecules at $T = 300$ K.

peptide bonds in the cis-isomer conformation (i.e., backbone dihedral angle $\omega = 0^\circ$). Of the 20 natural amino acids, only proline is “comfortable” in the cis-isomer conformation. In addition, the cis-trans isomerization of X-Pro peptide groups is one of the rate-determining steps for folding and unfolding of various proteins (49–53). The probability distribution of cis/trans prolyl bonds is affected by neighboring amino acids (54, 55), pH and ionic strength (56), solvent (57–63), and chain length (64). Experimentally, it is known that the left-handed PPII structure is favored in water, benzyl alcohol, trifluoroethanol, organic acids, and most of the other solvents, while the PPI structure is favored in the presence of aliphatic alcohols like propanol (64–67).

Traditionally, the relative rigid structure of PPII has been used as a “molecular ruler” in structural biology, especially for the validation of spectroscopic rulers in Förster resonance energy transfer (FRET) experiments (68). However, recent studies have called this traditional picture of polyproline peptide as a rigid rod into question. (69–75). Specifically, recent calculations (4–6) based on the methodology outlined in this chapter indicate that pure PPI and PPII are just one of the possible minima of the phase space, with many other minima being associated with other stable or metastable structures which are characterized in terms of the cis or trans nature of the prolyl bonds. These results complement the PET experimental results (74), with regard to the existence of stable heterogeneous subpopulation of polyproline conformers. More importantly, by providing detailed free energy landscapes in terms of different collective variables it has become possible to characterize the probability distribution of the different isomers, along with the associated most probable transition pathways and mechanisms.

5.2.1. Simulation Details

From a computational point of view, the characterization of the conformations of proline-based peptides is rather difficult, since the cis/trans isomerization is much slower than the formation and rupture of hydrogen bonds in an α -helix coil transition. Indeed, characteristic time scales for prolyl isomerization range in tens to hundreds of seconds at room temperature (54, 76), leading to an energy barrier of the order of 10–20 kcal/mol (77–83). Due to these huge energy barriers, traditional MD simulations cannot explore the relevant conformational space and the whole range of cis/trans transitions. Hence, we made use of ABMD, HT-REMD, and SMD methods to explore the free energy landscapes of proline systems (4–6, 84). In particular, to exemplify the methodology, we discuss simulations of the polyproline pentamer (i.e., peptide *Ace* – (*Pro*)_{*n*} – *Nme* with *n* = 5 denoted as *P*₅) in an implicit water environment based on the Generalized Born approximation (85, 86), as well as simulations a corresponding polyproline nanomer (denoted *P*₉) in vacuo and in explicit hexane and 1-propanol.

Initial configurations in all cases consisted of the unfolded peptides, which were generated using LEAP. The *P*₅ simulations made use of the ff99SB version of the Cornell et al. force field (87), while *P*₉ simulations used ff99 (88). The integrator time step was set to 1 fs, and use was made of the Langevin thermostat (46) with a collision frequency of $\gamma = 1 \text{ ps}^{-1}$. The lengths of all bonds containing hydrogen atoms were fixed using the SHAKE algorithm, with the tolerance set to 10^{-5} . There was no cutoff associated with the nonbonded interactions in vacuo and implicit water simulations. For the *P*₉ explicit solvents simulations, the peptide was bathed separately in 426 molecules of the nonpolar solvent hexane, and in 912 molecules of 1-propanol. The explicit solvent calculations were carried out in a truncated octahedron cell using periodic boundary conditions, with a box length of 79.800 Å in all directions. The PME method (47, 48) was used with the cubic spline approximation, a cutoff of 8.0 Å, a nonbonded “skin” of 1.0 Å, a direct sum tolerance of 10^{-5} , and a grid size of 64 in each of the three-dimension.

5.2.2. Collective Variables

It is convenient to cast the collective variables of the proline systems in terms of changes associated with the cis/trans isomerization of the ω prolyl bonds (see Fig. 5), which also happens to be the most relevant slow mode of the system. Since ω varies between $\omega = 0^\circ$ (cis) and $\omega = 180^\circ$ (trans) one can “capture” different conformations of proline peptides with the following collective variable:

$$\Omega = \sum_{i=1}^n \cos \omega_n, \quad (19)$$

where the sum runs over all prolyl bonds. Clearly, for a proline-rich peptide with *n* prolyl bonds, Ω takes on the values $-n, -n+2, \dots, n$, and describes the net balance of the cis/trans states.

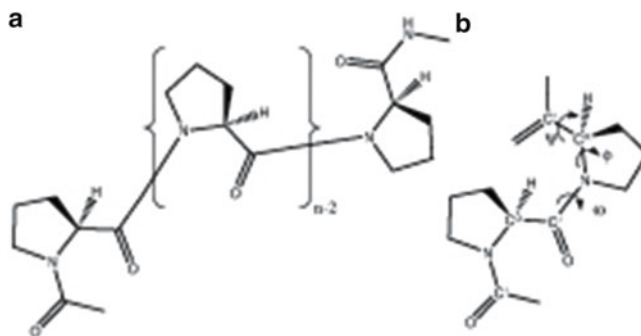


Fig. 5. (a) Schematic two-dimensional representation of Ace-(Pro)_n-Nme; (b) Definition of backbone dihedral angles ϕ , Ψ , and ω .

Although very useful, there are still a considerable number of degeneracies associated with this collective variable. To remove the most relevant ones (i.e., the ones associated with the prolyl bonds), we define additional collective variables that may be used in conjunction with Ω . For instance, one can consider the “interface” between bonds and define a collective variable Λ as

$$\Lambda = \sum_{i=1}^{n-1} \cos(\omega_i + \omega_{i+1}). \quad (20)$$

For a peptide with n prolyl bonds, there are $n - 1$ interfaces. If two neighboring bonds have the same dihedral angle ω , then their interface is $\cos(\omega_i + \omega_{i+1}) = +1$; otherwise, it is -1 . Correspondingly, Λ can take on any of the values $-n + 1, -n + 3, \dots, n - 1$. For $|\Omega| = n$, there is only one value of Λ ($\Lambda = n - 1$); for $\Omega = 0$ (that exists only for even values of n), there are $n - 1$ values of Λ ; for $|\Omega| = m$ (with $m \neq n$ and $m \neq 0$), there are $n - m$ values of Λ . Thus, a typical free energy landscape in the (Ω, Λ) plane has a triangular shape. Note that larger the numerical value of Λ , the more the structure corresponds to the ideal PPI or PPII helix, since the number of cis/trans interfaces is minimized.

5.2.3. Free Energy Maps

In this section, we discuss the free energy results obtained from the ABMD runs for the selected examples. Specifically, one-dimensional free energy profile based on Ω for P₅ and two-dimensional free energy maps as a function of the collective variables (Ω, Λ) for P₉ in different environments will be briefly discussed.

ABMD simulations for P₅ in implicit water (P₉ in vacuo) were carried out using 20 (24) replicas. Initial calculations of the free energy landscapes involved all the replicas at the same temperature $T = 1, 200$ K without any exchanges between the replicas. These short and inaccurate multiple-walker ABMD runs of 10 ns duration provided us with a rough estimate of the biasing potentials to be encountered during the next phase of the runs. The second

step in the simulation protocol involved long ABMD runs with replica exchange, using the results from the first step. However, this time, the temperature of all the replicas was varied from 300 to 1, 200 K. Specifically, these were set to 300, 322, 347, 373, 401, 432, 464, 499, 537, 578, 622, 669, 720, 774, 833, 896, 964, 1, 037, 1, 115, and 1, 200 K for P_5 in implicit water and 300, 318, 338, 359, 381, 405, 430, 457, 485, 516, 548, 582, 618, 656, 697, 740, 786, 835, 887, 942, 1, 001, 1, 063, 1, 129, and 1, 200 K for P_9 in vacuo. Each replica had its own biasing potential. Note that the temperature distribution between the lowest and highest values was chosen to be exponential, which usually leads to an optimal mixing (i.e., equal rates of exchange between the neighboring replicas). Although the number of replicas was somewhat influenced by peculiarities of our computer setup, the highest temperature chosen needs to be sufficiently high as to ensure flexibility in moving between different conformations (i.e., a small autocorrelation time). At the same time, because it determines the entire temperature distribution of the replicas, it needs to be small enough as to give a high rate of exchange between the replicas (i.e., good mixing). Both conditions need to be satisfied in order to ensure that the free energy results converge to the desired accuracy within a feasible runtime. Turning to the explicit solvent simulation, these were performed using the multiple-walker extension only. This was because test runs showed that parallel tempering runs were characterized by a very low measured rate of exchange between different replicas and therefore could not be considered useful.

In terms of the ABMD parameters, for the one-dimensional P_5 simulations, a kernel width of $4\Delta\xi = 0.2$ was used along with a flooding time scale of $\tau_F = 5.0, 25.0, 100.0$ ps with a runtime of 10 ns for each of the different τ_F simulations. For the two-dimensional free energy maps, a kernel width of $4\Delta\xi = 0.5$ was used for both collective variables. The flooding time scales were $\tau_F = 1.0, 2.0, 5.0$ ps, with each simulation lasting 20 ns. We note that the initial flooding time scale was chosen to be small enough as to generate rough estimates of the biasing potential in a short period of time. The larger flooding time scales were used to obtain more accurate results. However, the price one pays for this accuracy is that these simulations take considerably longer to reach the desired accuracy. Finally, we repeated the simulations using the calculated biasing potential for follow-up equilibrium umbrella sampling runs for another 30 ns. The kernel estimator method was then used to recover the biased probability density $p^B(\xi)$ (18).

Figure 6 shows the resulting free energy profile for P_5 in implicit water before (red) and after (black) the corrective runs. The difference is always smaller than 0.3 kcal/mol. The free energy difference between all-trans ($\Omega = -5$) and all-cis ($\Omega = 5$) structures was estimated to be 4.9 ± 0.3 kcal/mol. The structures associated with $\Omega = -3, -1, 1, 3$ have calculated free energies

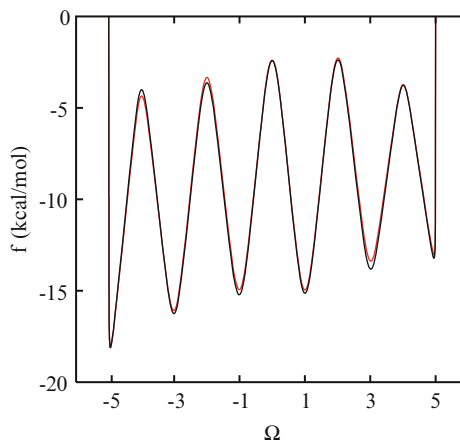


Fig. 6. Free energy profile of the polyproline pentamer peptide in implicit water as a function of Ω at $T = 300$ K.

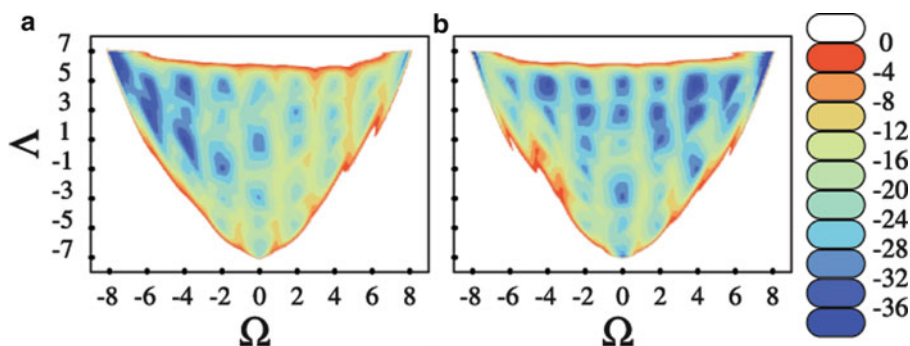


Fig. 7. Two-dimensional (Ω, Λ) free energy landscapes of the P_9 peptide in explicit solvents (a) hexane and (b) propanol.

of -3.0 , -2.0 , -1.9 , -0.6 kcal/mol, relative to the all-cis conformation. Overall, the results indicate that the more cis prolyl bond a structure has the greater its free energy.

We now discuss the (Ω, Λ) maps, shown in Figs. 7 and 8a. For these results, we used the $n=9$ prolyl bonds only to define the (Ω, Λ) collective variables, so that $(-8, 7)$ and $(8, 7)$ correspond to the PPII and PPI structures, respectively. Associated with $\Omega = -6$ are two values of Λ : the $\Lambda = 5$ structure involves a terminal cis bond, while $\Lambda = 3$ involves an inner cis bond. For $\Omega = 0$, there are $n - 2 = 6$ values of Λ , with the lowest $\Lambda = -7$ corresponding to a situation where all prolyl bonds alternate with respect to each other. Overall, the free energy minima on these maps are distributed in a triangular or wedge-like pattern.

The left (right) side of the triangle represents the trans (cis)-rich conformations. It is clear that the solvent has a great influence on the distribution of the conformers. Figure 7a shows the free energy map associated with a nanomer in the nonpolar solvent hexane.

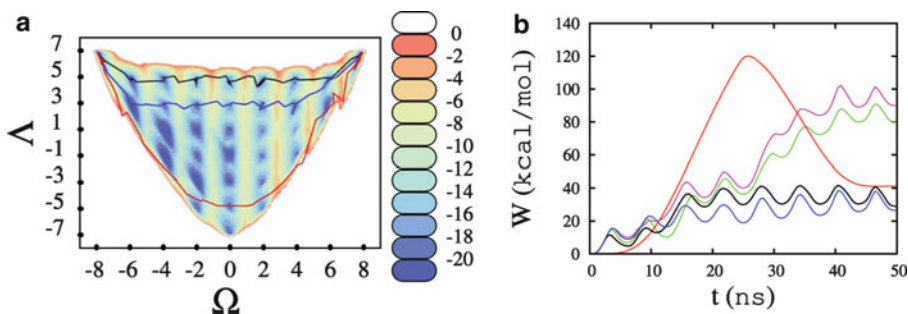


Fig. 8. **(a)** The (Ω, Δ) free energy landscapes in kcal/mol of a polyproline nanomer *in vacuo*, along with selected SMD pathways over these free energy landscapes. Here, the black, blue and red lines correspond to ZR, ZIO and “en-masse” pathways, respectively. **(b)** Here, we show the work (W) (required for a typical $PPII \rightarrow PPI$ transition) versus time for a nanomer *in vacuo*. Different mechanisms are characterized by different colors: black (ZR), blue (ZIO), pink (ZF), green (ZOI) and red (“en-masse”), respectively.

Here, the trans-rich structures dominate and the ideal PPII structure represents the global minimum with a free energy of about 20 kcal/mol lower than that of the ideal PPI structure. In the solvent 1-propanol, the proline nanomer has a very different behavior, and it is the cis-rich structures that are favored (Fig. 7b). In particular, the free energy difference between PPI and PPII was found to be ≈ -8 kcal/mol. We also note, that in general, the larger values of Δ are usually associated with a lower free energy. This trend is almost reversed for the proline nanomer *in vacuo*, as shown in Fig. 8a. The deeper minima are associated with lower Δ values, and neither PPII nor PPI represents the global minimum. The latter is associated with some of the more compact globular structures.

5.2.4. SMD Results

All the SMD simulations were carried out at $T = 300$ K, with a harmonic constant set to 100 kcal/mol. SMD simulations were used for two different purposes:

1. For a quantitative estimation of the free energy difference between the PPI and PPII structures for short peptides implicit water. Specifically, we discuss here results for the P_5 peptide.
2. For a qualitative comparison of the different mechanisms associated with the $PPII \rightarrow PPI$ transitions. Specifically, we discuss here results for P_9 *in vacuo*.

To estimate the free energy difference between PPII and PPI for P_5 using both the Jarzynski equality and BAR method, we carried out 100 forward and reverse nonequilibrium simulations by pulling on the Ω variable. Each of these simulations involved an initial equilibration run lasting 0.5 ns, followed by a 2.0 ns SMD run during which the Ω variable was pulled at a constant velocity. The simulations were carried out at $T = 300$ K, with the harmonic constant set to 100 kcal/mol.

In terms of the convergence, the results based on the Jarzynski equality turn out not to be particularly accurate. However, the results based on 100 runs using the MLE method yield a free energy difference between PPI and PPII of 4.6 kcal/mol, which is in good agreement with the 4.9 kcal/mol estimate obtained from the ABMD runs.

For a qualitative comparison of the different transition mechanisms associated with $PPII \rightarrow PPI$ transition of P_9 in vacuo, we steered the system via the successive flipping of the ω dihedral angles. Since it is not possible to exhaustively examine all possible transition pathways and mechanisms, we concentrated on a select group of mechanisms. Specifically, we focused on different zipper-like mechanisms, for which changes take place via the successive switching of neighboring prolyl bonds. These include: a zipper-like mechanism starting from the forward, acetylated end for the $PPII \rightarrow PPI$ transition which we label ZF (zipper-forward), such that the intermediate conformation in either direction would take the form $Ace - CCC \cdots TTT - Nme$; a zipper-like mechanism starting from the reverse, amidated end for the $PPII \rightarrow PPI$ transition which we label ZR (zipper-reverse), such that the intermediate conformation would take the form $Ace - TTT \cdots CCC - Nme$; a zipper-like mechanism starting from the center and moving outward for the $PPII \rightarrow PPI$ transition which we label ZIO (zipper-inside-outwards), such that the intermediate conformation would take the form $Ace - TT \cdots CCC \cdots TT - Nme$; a zipper-like mechanism starting from both ends, and working its way inward for the $PPII \rightarrow PPI$ transition which we label ZOI (Zipper-Outside-Inwards), such that the intermediate conformation would take the form $Ace - CC \cdots TTT \cdots CC - Nme$. Additionally, we considered mechanisms that involve changing the prolyl bonds in an “en-masse” mechanism where all the bonds are changed simultaneously. Clearly, there is some ambiguity in how the ZIO and ZOI mechanisms can be implemented. For example, one can flip either two bonds at a time *simultaneously*, or only one at a time. For the latter, the order of bond flipping provides for a further degree of freedom. In our simulations the alternative flipping of bonds was used.

Figure 8a plots the pathways associated with some of these mechanisms in the (Ω, A) space. Specifically, the black, blue, and red lines mark pathways taken by the ZR, ZIO, and “en-masse” mechanisms. From this map, it is clear that ZR mechanism always chooses the largest value of A , while for ZIO, the path is always one minimum below the ZR path, except for the end points. Although not shown here, ZF (ZOI) mechanism follows a path similar to ZR (ZIO) in the (Ω, A) space, due to the symmetries associated with both of the collective variables. In an “en-masse” mechanism, A does not stay close to its numerically largest value. The work performed by the system for these simulations is shown in

Fig. 8b. In contrast to the many short duration SMD runs performed for estimating the free energy differences, the results here are based on a few long runtime simulations lasting 50 ns.

In terms of the most to least amount of work required for the transition, the ordering is ZF, ZOI, “en-masse,” ZR, and ZIO. It is also of interest to figure out how the transition is likely to be initiated. For the $PPII \rightarrow PPI$ transition, the flipping at the amidated end and/or middle bonds appear to be good options. To probe this question from a different perspective, we performed SMD simulations for a PPII P_9 , in order to calculate the amount of work required to flip each of the bonds from trans to cis. The results indicate that the prolyl bonds at the acetylated and amidated ends require the least amount of work. However, the next preferred ones are not their neighbors. Rather, it is the middle bonds and their neighbors that require the least amount of work for flipping. Thus, it appears that while the first bond switch might well take place at either end, the subsequent bond switches favor a ZIO mechanism. This behavior suggests ZIO mechanism can be an appropriate mechanism for $PPII \rightarrow PPI$ transition, while ZOI is very unlikely. Turning back to Fig. 8b, it is clear that ZR mechanism requires less work than ZF, indicating that the intermediate structures would take the form of $Ace - TT \cdots CC - Nme$ rather than $Ace - CC \cdots TT - Nme$. This is consistent with the results from enzymatic hydrolysis kinetics experiments on polyproline in aqueous solution (89, 90), but opposite to the results from the NMR experiments on polyproline in D_2O solution (91).

5.2.5. Enhanced Equilibrium Sampling

In order to exemplify a population analysis carried out with REMD, we discuss here the cis versus trans population distribution for P_5 in implicit water. For such an analysis, the one-dimensional Ω free energy maps as obtained prior to the “umbrella correction” step (see Fig. 6) formed the basis of the HT-REMD runs. Initially, we considered running only temperature REMD. However, in assessing the mixing properties of the hottest replica at the highest temperature of $T = 1, 200$ K (by simulating it alone), we found that the temperature (without the biasing potential) was not enough to induce frequent cis/trans transitions. In other words, the hottest replica did not show enough flexibility and its autocorrelation time turned out to be too large (tens of nanoseconds). We also attempted to use biased replicas alone, all at the lowest temperature of $T = 300$ K. But again, the autocorrelation time turned out to be large. Therefore, we were forced to combine both features in our REMD simulations. We used the highest temperature replica with the same biasing potential as obtained from ABMD in order to ensure that (a) all possible cis/trans states get visited not in more than a few hundred picoseconds; and (b) the autocorrelation time is reasonably small (less than 2 ns) so as to produce a reasonable number of independent samples in a feasible runtime.

Four more replicas were added all at the lowest temperature $T = 300$ K, one with no biasing potential and three with the ABMD-resulted biasing potential of the lowest temperature replica, scaled-down by the factors of 0.49, 0.76, 0.9. The temperatures, the scaling factors, and the ratio of the temperature-varying versus Hamiltonian-varying replicas (20 vs. 4) were chosen in a way as to ensure a similar rate of exchange between all neighboring replicas. We ran 100 ns HT-REMD simulations, and collected the trajectory of the unbiased replica every 1 ps that resulted in 10^5 equilibrium samples. We note that the rest of the simulation parameters were identical to those used for the ABMD runs.

Concerning the cis/trans pattern, the all-trans structure proved to be the most probable conformation by a population of 55%, followed by structures with only one cis bond that form 28% of the whole population, combined. Labelling the proline residues P_1 , P_2 etc from the *Ace* – end of the peptide, we found that structures with only a single cis bond at *Ace* – P_1 , P_5 – *Nme*, P_1 – P_2 , P_2 – P_3 , or P_4 – P_5 have the populations of 11 %, 8 %, 5 %, 4 %, or 4 %, respectively. The rest of the structures with more than one cis bond form only 17% of the entire population. A residue-based analysis of the polyproline structure, as opposed to a sequence-

Table 1
Probability (as a percentage) of prolines in the trans (as opposed to cis) conformation (P(T)), to be in the F (as opposed to) region of the Ramachandran plot (P(F)) for P_5 in implicit water as obtained bu HT-REMD simulations. Conditional probabilities (as a percentage) of being in F conditioned on being either trans or cis are denoted as P (FIT) and P(FIC), respectively, are also listed. Finally, the probability (as a percentage) of being in a PPII or PPI configuration are also listed. These probabilities are all listed for the individual residues along with the average value for the peptide

bond	$p(T)$	$p(F)$	$p(F T)$	$p(F C)$	$p(PPII)$	$p(PPI)$
Ac- P_1	81	98	98	97	82	18
P_1 - P_2	88	98	98	99	85	12
P_2 - P_3	90	99	98	99	88	10
P_3 - P_4	94	56	53	80	49	5
P_4 - P_5	85	2	2	1	2	0
Average	88	71	70	75	62	9

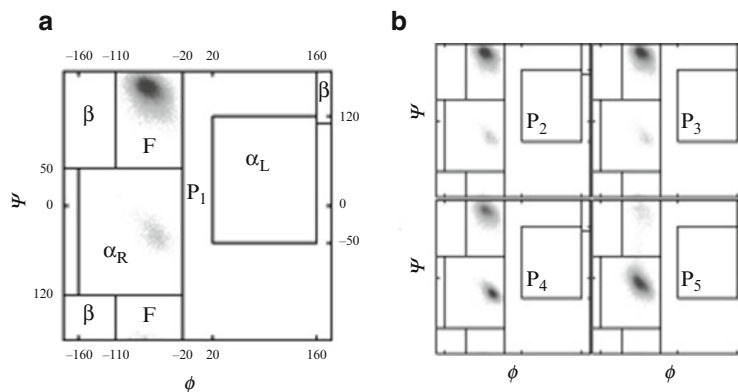


Fig. 9. Ramachandran plots of all proline residues for a polyproline pentamer peptide: (a) P_1 residue results along with the definition of different regions (b) results for $P_2, 3, 4, 5$ residues.

based analysis, is given in Table 1. In brief, one can determine the cis/trans content of each residue. The *in vacuo*P(T) results indicate that the trans content is about 88% on average, varying somewhat with the position of the residue. The highest cis content can be found at the acetylated and amidated ends by 19% and 15%, respectively.

Please provide the significance of bold in Table 1.

Having determined the coordinates for a large number of equilibrium structures, one can investigate not only the cis/trans patterns but also the Ramachandran plots of each residue (5, 84). Figure 9 shows Ramachandran plots for all P_5 residues. It turns out, that the dihedrals for the PPI and PPII structure fall into the F-region of the Ramachandran plots, as defined by the Zimmerman code (92). We have marked different regions of the Ramachandran plot in Fig. 9a as F , β , α_R , and α_L . The F region here is defined to be slightly broader than its original formulation, in order to encompass all the different fluctuations. Table 1 also gives the probability of finding the proline residue in the F region ($p(F)$) and the conditional probabilities of finding the proline residue in the F region and having either a trans or cis form (denoted by $p(F|T)$ and $p(F|C)$, respectively). Since an ideal PPII proline residue is in a trans conformation and also in the F region, one can find the probability of each residue being in the PPII conformation as $p(\text{PPII}) = p(T) \times p(F|T)$. Similarly, the same holds true for a residue in the PPI configuration so that $p(\text{PPI}) = p(C) \times p(F|C)$. The average PPII (PPI) content of proline pentamer is estimated as 62(9)%. We note that, the last proline residue at the amidated end is mostly found to be in the α -helix region rather than F region (see Fig. 9b). This effect is also evident in the neighboring “next-to-the last” residue, particularly when it is in trans conformation.

6. Summary

In summary, we have briefly outlined three MD-based methods (ABMD, REMD, and SMD) useful for calculating biomolecular properties. Their workings, along with simulation details, was presented in terms of two examples drawn from sugar puckering and polyproline peptide folding. A practical implementation of these methods is currently to be found in the SANDER module of the AMBER versus10 (24) simulation package.

Acknowledgements

This research was supported by NSF via FRG-0804549 and 1021883, and DOE. We also thank the NC State HPC for extensive computational support.

References

1. Babin V, Roland C, Sagui C (2008) Adaptively biased molecular dynamics for free energy calculations. *J Chem Phys* 128:134101
2. Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. *Computing Science and statistics: the 23rd symposium on the interface*. Interface Foundation, Fairfax, pp 156–163
3. Izrailev S, Stepaniants B, Kosztin D, Lu H, Molnar F, Wriggers W, Schulten K (1998) *Steered molecular dynamics*. Computational molecular dynamics: challenges, methods, ideas. Springer, Berlin, Germany
4. Moradi M, Babin V, Roland C, Darden T, Sagui C (2009) Conformations and free energy landscapes of polyproline peptides. *Proc Natl Aca Sci USA* 106:20746
5. Moradi M, Babin V, Roland C, Sagui C (2010) A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers. *J Chem Phys* 133:125104
6. Moradi M, Lee JG, Babin V, Roland C, Sagui C (2010) Free energy and structure of polyproline peptides: an ab initio and classical molecular dynamics investigation. *Int J Quant Chem* 110:2865–2879
7. Frenkel D, Smit B (2002) *Understanding molecular simulation*. Academic Press, Orlando, USA
8. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys* 23:187–199
9. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63:1195–1198
10. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg J (1992) The weighted histogram analysis method for free energy calculation of biomolecules: I The Method. *J Comput Chem* 13:1011
11. van Duijneveldt JS, Frenkel D (1992) Computer simulation study of free energy barriers in crystal nucleation. The method. *J Chem Phys* 96:4655–4668
12. Lelièvre T, Rousset M, Stoltz G (2007) Computation of free energy profiles with parallel adaptive dynamics. *J Chem Phys* 126:134111
13. Bussi G, Laio A, Parrinello M (2006) Equilibrium free energies from nonequilibrium metadynamics. *Phys Rev Lett* 96:090601
14. Huber T, Torda AE, van Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular dynamics simulation. *J Comput Aided Mol Des* 8:695–708
15. Wang F, Landau DP (2001) *Phys Rev Lett* 115:2050–2053
16. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci* 99:12562–12566

17. Iannuzzi M, Laio A, Parrinello M (2003) Efficient exploration of reactive potential energy surfaces using car-parrinello molecular dynamics. *Phys Rev Lett* 90:238302–238311
18. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
19. De Boor C (1978) A practical guide to splines. Springer, New York
20. Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J Phys Chem* 110:3533–3539
21. Sugita Y, Kitao A, Okamoto Y (2000) *J Chem Phys* 113:6042
22. Bussi G, Gervasio FL, Laio A, Parrinello M (2006) Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc* 128:13435–13441
23. Piana S, Laio A (2007) A bias-exchange approach to protein folding. *J Phys Chem B* 111:4553–4559
24. Case D et al (2008) AMBER 10. University of California, San Francisco
25. Hooft RWW, van Eijck BP, Kroon J (1992) An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J Chem Phys* 97:6690–6694
26. Babin V, Roland C, Darden TA, Sagui C (2006) The free energy landscape of small peptides as obtained from metadynamics with umbrella sampling corrections. *J Chem Phys* 125:204909
27. Babin V, Sagui C (2010) Conformational free energies of methyl- α -L-iduronic and methyl- β -D-glucuronic acids in water. *J Chem Phys* 132:104108
28. Laghaei R, Mousseau N, Wei G (2010) Effect of the disulfide bond on the monomeric structure of human amylin studied by combined Hamiltonian and temperature replica exchange molecular dynamics simulations. *J Phys Chem B* 114:7071–7077
29. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78:2690–2693
30. Crooks GE (2000) Path-ensemble averages in systems driven far from equilibrium. *Phys Rev E* 61:2361–2366
31. Shirts MR, Bair E, Hooker G, Pande VS (2003) Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys Rev Lett* 91:140601
32. Kosztin I, Barz B, Janosi L (2006) Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J Chem Phys* 124:064106
33. Ensing B, Laio A, Parrinello M, Klein M (2005) A recipe in the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J Phys Chem B* 109:6676–6687
34. Rao VSR, Qasba PK, Balaji PV, Chandrasekaran R (1998) Conformation of carbohydrates. Harwood Academic Publishers, Amsterdam
35. Kirschner KN, Woods RJ (2001) Solvent interactions determine carbohydrate conformation. *Proc Natl Acad Sci* 98:10541–10545
36. Almond A, Sheehan JK (2003) Predicting the molecular shape of polysaccharides from dynamic interactions with water. *Glycobiology* 13:255–264
37. Kräutler V, Müller M, Hünenberger PH (2007) Conformation, dynamics, solvation and relative stabilities of selected [beta]-hexopyranoses in water: a molecular dynamics study with the gromos 45A4 force field. *Carbohydr Res* 342:2097–2124
38. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J Comp Chem* 29:622–655
39. Christen M, van Gunsteren WF (2008) On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: a review. *J Comp Chem* 29:157–166
40. Cremer D, Pople JA (1975) General definition of ring puckering coordinates. *J Am Chem Soc* 97:1354–1358
41. Jorgensen WL, Chandrasekhar J, Madura J, Klein ML (1983) Comparison of Simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935
42. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112:6127–6129
43. Hawkins GD, Cramer CJ, Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* 100:19824–19839
44. Hawkins GD, Cramer CJ, Truhlar DG (1995) Pairwise solute descreening of solute charges from a dielectric medium. *Chem Phys Lett* 246:122–129
45. Tsui V, Case D (2001) Theory and applications of the generalized Born solvation model in

- macromolecular simulations. *Biopolymers* (Nucl Acid Sci) 56:275–291
46. Brünger A, Brooks CL, Karplus M (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem Phys Lett* 105:495–500
 47. Darden TA, York DM, Pedersen LG (1993) Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
 48. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
 49. Brandts JF, Halvorson HR, Brennan M (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* 14:4953–4963
 50. Tanaka S, Scheraga HA (1977) Hypothesis about the mechanism of protein folding. *Macromolecules* 10:291–304
 51. Schmid FX, Mayr LM, Mücke M, Schönbrunner ER (1993) Prolyl isomerases: role in protein folding. *Adv Protein Chem* 44:25–66
 52. Houry WA, Scheraga HA (1996) Nature of the unfolded state of Ribonuclease A: effect of cis-trans X-Pro peptide bond isomerization. *Biochemistry* 35:11719–11733
 53. Wedemeyer WJ, Welker E, Scheraga HA (2002) Proline cis-trans isomerization and protein folding. *Biochemistry* 41:14637–14644
 54. Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerization. *J Mol Biol* 279:449–460
 55. Grathwohl C, Wuthrich K (1976) The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers* 15:2025–2041
 56. Grathwohl C, Wuthrich K (1976) Nmr studies of the molecular conformations in the linear oligopeptides H-(L-Ala)_n-L-Pro-OH. *Biopolymers* 15:2043–2057
 57. Steinberg IZ, Harrington WF, Berger A, Sela M, Katchalski E (1960) The configurational changes of poly-L-proline in solution. *J Am Chem Soc* 82:5263–5279
 58. Gornick F, Mandelkern L, Diorio AF, Roberts DE (1964) Evidence for a cooperative intramolecular transition in poly-L-proline. *J Am Chem Soc* 86:2549–2555
 59. Mandelkern L (1967) Poly-L-proline. In: Fasman GD (ed.) *Poly- α -amino acids*. Marcel Dekker, New York
 60. Strassmair H, Engel J, Zundel G (1969) Binding of alcohols to the peptide CO-group of poly-L-proline in the I and II conformation. I. Demonstration of the binding by infrared spectroscopy and optical rotatory dispersion. *Biopolymers* 8:237–246
 61. Tanaka S, Scheraga HA (1975) Theory of the cooperative transition between two ordered conformations of poly(L-proline). II. Molecular theory in the absence of solvent. *Macromolecules* 8:504–516
 62. Tanaka S, Scheraga HA (1975) Theory of the cooperative transition between two ordered conformations of poly(L-proline). III. Molecular theory in the presence of solvent. *Macromolecules* 8:516–521
 63. Kofron JL, Kuzmic P, Kishore V, Colon-Bonilla E, Rich DH (1991) Determination of kinetic constants for peptidyl prolyl cis-trans isomerases by an improved spectrophotometric assay. *Biochemistry* 30:6127–6134
 64. Kakinoki S, Hirano Y, Oka M (2005) On the stability of polyproline-I and II structures of proline oligopeptides. *Poly Bull* 53:109–115
 65. Traub W, Shmueli U (1963) Structure of poly-L-proline I. *Nature* 198:1165–1166
 66. Cowan PM, McGavin S (1955) Structure of poly-L-proline. *Nature* 176:501–503
 67. Mutter M, Wöhr T, Gioria S, Keller M (1999) Pseudo-prolines: induction of cis/trans-conformational interconversion by decreased transition state barriers. *Biopolymers* 51:121–128
 68. Stryer L, Haugland RP (1967) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc Natl Acad Sci USA* 58:719–726
 69. Jacob J, Baker B, Bryant GR, Cafiso DS (1999) Distance estimates from paramagnetic enhancements of nuclear relaxation in linear and flexible model peptides. *Biophys J* 77:1086–1092
 70. Watkins LP, Chang H, Yang H (2006) Quantitative single-molecule conformational distributions: a case study with poly-(L-proline). *J Phys Chem A* 110:5191–5203
 71. Sahoo H, Roccatano D, Hennig A, Nau WM (2007) A spectroscopic ruler for short polyprolines. *J Am Chem Soc* 129:9762–9772
 72. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the spectroscopic ruler revisited with single-molecule fluorescence. *Proc Natl Acad Sci USA* 102:9754–9759

73. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A, Eaton WA (2007) Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proc Natl Acad Sci USA* 104:18964–18969
74. Doose S, Neuweiler H, Barsch H, Sauer M (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. *Proc Natl Acad Sci USA* 104:17400–17405
75. Dolgih E, Ortiz W, Kim S, Krueger BP, Krause JL, Roitberg AE (2009) Theoretical studies of short polyproline systems: recalibration of molecular ruler. *J Phys Chem A* 113:4639–4646
76. Grathwohl C, Wuthrich K (1981) Nmr studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* 20:2623–2633
77. Phillips WD (1955) Restricted rotation in amides as evidenced by nuclear magnetic resonance. *J Phys Chem* 23:1363–1264
78. Pauling L (1960) The nature of the chemical bond, 3rd edn. Cornell Univ. Press, Ithaca, NY
79. Tonelli AE (1973) Estimate of the barriers hindering rotation about the C.alpha.-C' bond between the cis' and trans' conformations in an isolated L-proline residue. *J Am Chem Soc* 95:5946–5948
80. McDonald DC, Still WC (1996) Molecular Mechanics parameters and conformational free energies of proline-containing peptides. *J Org Chem* 61:1385–1391
81. Venkatachalam CM, Price BJ, Krimm S (2004) A theoretical estimate of the energy barriers between stable conformations of the proline dimer. *Biopolymers* 14:1121–1132
82. Kang YK, Choi HY (2004) Cis-trans isomerization and puckering of proline residue. *Biophys Chem* 111:135–142
83. Kang YK, Jhon JS, Park HS (2006) Conformational preferences of proline oligopeptides. *J Phys Chem B* 110:17645–17655
84. Moradi M, Babin V, Sagui C, Roland C (2010) A statistical analysis of the PPII propensity of amino acid guests in proline-rich peptides. *Biophysical J* 100:1083
85. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* 104:3712–3720
86. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* 55:383–394
87. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–725
88. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J Am Chem Soc* 117:5179–5197
89. Lin LN, Brandts JF (1979) Role of cis-trans isomerism of the peptide bond in protease specificity. Kinetic studies on small proline-containing peptides and on polyproline. *Biochemistry* 18:5037–5042
90. Lin LN, Brandts JF (1980) Kinetic mechanism for conformational transitions between poly-L-prolines I and II: a study utilizing the cis-trans specificity of a proline-specific protease. *Biochemistry* 19:3055–3059
91. Torchia DA, Bovey FA (1971) a nuclear magnetic resonance study of poly(l-proline) in aqueous and aqueous salt solutions. *Macromolecules* 4:246–251
92. Zimmerman SS, Pottle MS, Némethy G, Scheraga HA (1977) Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. *Macromolecules* 10:1–9