PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Helix propensities calculations for amino acids in alanine based peptides using Jarzynski's equality

Ignacia Echeverria and L. Mario Amzel*

Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205

ABSTRACT

Jarzynski's equality (Jarzynski, Phys Rev E 1997; 56:5018 and Jarzynski, Phys Rev Lett 1997; 78:2690) relates equilibrium free energy differences between two states A and B to the work done when the system is driven repeatedly and irreversibly from an equilibrium state A to equilibrium state B. We present calculations of helix propensities using a novel procedure based on this equality. In particular, a work probability distribution is built based on combinations of multi-step trajectories that give representative work distributions without requiring computing an unreasonable large number of trajectories between states. A small number of trajectories (15) are used to construct a distribution that contains millions of work values. This distribution is used to calculate ΔG_{AB} using Jarsynski's equation. To apply and test this method, we used as a model system a dodeca-alanine helix, analyzing its extension using mechanical force. This helix was used as the basis of a host guest system in which two of the 12 residues are substituted by some other amino acid (as the guests). The differences between the unfolding free energies of the substituted peptides and the all-alanine peptide provided values for $\Delta\Delta G$ that can be interpreted as the helix propensities of each amino acid. Results show good correlation with the experimental measurements of Baldwin and coworkers (Chakrabartty *et al.*, Protein Sci 1994; 3:843–852).

Proteins 2010; 78:1302–1310.

Key words: helix propensities; free energy calculations; protein folding; Jarzynski's equality.

INTRODUCTION

The α-helix is one of the principal motifs in native proteins. Its highly geometric structure and fast folding makes it an excellent model system to study folding energetics. Although α-helices are stabilized by i to i + 4 backbone hydrogen bonds, different amino acids have different effects on the helix stability. This difference in stability is referred to as the helix propensity of the amino acid. From an entropic point of view, helix formation restricts the configurational freedom of both the backbone and the side chain rotations. Alanine, with its small side chain, is a strong helix former, leucine and arginine are helix-indifferent, and all other amino acids have a destabilizing effect. Hence, it has been proposed that the loss of configurational entropy accounts for the thermodynamic cost responsible for observed helix-formation propensities.[1,2] Thermodynamic helix-formation propensities have been studied experimentally and measured as differences in the Gibbs free energy of unfolding using host-guest schemes. In these systems an alanine-based peptide is used as the "host" and point substitutions with another amino acid residue X are introduced as "guests".

To calculate helix-formation propensities using computational methods we studied the process of extending a 12-residue α-helix using a mechanical force. The ΔG for the process was estimated for a reference 12-alanine peptide (Ala_{12}) and for peptides in which two of the 12 residues were substituted by the same nonalanine amino acid X referred to as the guest. We refer to this system as the $Ala_{10}X_2$ host-guest peptide (sequence $Ala_3XAla_4XAla_3$). Additionally, the difference between the free energies of extending the $Ala_{10}X_2$ and Ala_{12} helices, $\Delta\Delta G$, was calculated. The value of $\Delta\Delta G$, expressed on a per guest residue basis ($\Delta\Delta G/2$), can be interpreted as a measurement of the helix stabilization propensity of each amino acid type.

The calculation of free energy differences ΔG_x plays a fundamental role in the study of molecular processes such as protein folding, ligand binding and solubility. Several computational methods, including free energy calculations,[3–5] thermodynamic integrations,[6–10] and free energy

Abbreviations: CFT, Crooks fluctuation theorem; MD, molecular dynamics; MSTC, multi-step trajectory combination.

Grant sponsor: Department of Energy; Grant number: #DE-FG02-04ER25626
*Correspondence to: L. Mario Amzel, Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine, Baltimore, MD 21205. E-mail: mamzel@jhmi.edu.
Received 15 June 2009; Revised 13 October 2009; Accepted 20 October 2009
Published online 30 October 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22649

perturbation[11–13] have been used before to estimate free energy differences between equilibrium states. However, they require that all steps of the transition be carried out at or close to equilibrium, making the calculations difficult to perform. Furthermore, calculations using these methods suggest that comparisons between states or processes with different numbers of atoms are intrinsically inaccurate. Comparing the unfolding of an Ala_{12} helix with that of helices of other sequences falls into this category.

It has been shown that for atomic systems that involve large fluctuations the work necessary to move the system between two equilibrium states A and B will vary from one repetition to another. As a consequence of the second law of thermodynamics, the average work of a large number of repetitions carried out irreversibly is greater than the thermodynamic change in free energy between states A and B ($\langle W \rangle \geq \Delta G$, where $\langle \ldots \rangle$ denotes average). Recently, Jarzynski has shown that for a system driven between two equilibrium states A and B by doing work.

$$\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta G},$$

(where $\beta = 1/kT$)[14,15] even if during the process the system is far from equilibrium. This remarkable equation provides the basis for estimating free energy differences from data obtained for an irreversible process, provided that enough repetitions are carried out. Since 1997, when Jarzynski first presented this result, it has been rederived in many frameworks[16–18] and tested experimentally.[19–21] In the first reported experimental work, Liphardt *et al.*[19] performed single molecule experiments to mechanically unfold a single RNA molecule. Application of Jarzynski's equality allowed them to recover ΔG from a set of irreversible trajectories.

This equation also leads to new approaches for estimating free energies using molecular dynamics (MD) simulations to construct the work distribution $P(W)$. MD simulations have already been used to test this relation for simple systems,[22] as well as for some more complex systems.[23–26] However, for a system driven irreversibly between two equilibrium states A and B by a force, the computational time needed to access a representative work distribution may be prohibitively high, especially if the two states are well-separated energetically.[3]

In the present work, helix-formation propensities for different amino acids were calculated using a novel procedure based on Jarzynski's equality: Multi-step trajectory combination (MSTC). In this method, path ensembles are constructed and used to obtain representative work distributions for process in which the system is driven out of equilibrium, without requiring an unreasonably large number of trajectory simulations. The general scheme consists of dividing the trajectory between the initial and final states into m steps, equilibrating the system at the end of each step. After generating n trajectories, new paths are generated by combining steps of different trajectories, taking advantage of the fact that, because of equilibration, different trajectories at the end of each step share microstates. In this procedure, if we generate n trajectories of m steps each, they can be combined to produce n^m trajectories. This means that if one computes, for example, 15 full trajectories, each divided into 10 steps, equilibrating at the end of each step, one can obtain a distribution of work values corresponding to 15^{10} ($\sim 6 \times 10^{11}$) full trajectories. Even if equilibration after each step increases the computational time by a factor of 10, the gain in computational efficiency will still be close to a factor of 10^9. Computation of ΔG over the n^m trajectories is equivalent to using Jarzynski's equation to compute ΔG_m, the value of the free energy of each of the m steps averaged over the n trajectories, and adding them to obtain the total ΔG value (see appendix).

METHODOLOGY

Unfolding helices by force

All simulations were performed using the program CHARMM.[27] The complete system used was a 12-alanine helix solvated in a box of TIP3 water molecules of dimensions 49.67 × 27.94 × 27.94 Å (~1250 TIP3 molecules). The total number of atoms in the system, including hydrogens, was 3856; periodic boundary conditions were used in all calculations. Simulations were performed in a constant number of atoms, volume and energy (NVE) ensemble. The initial system was pre-equilibrated at a temperature of 300 K.

Extension of the helices was carried out by applying a force along the line connecting the Cα of the C-terminus and the Cα of the (acetylated) N-terminus. The C-terminal Cα was fixed and the N-terminal Cα was attached to a moving target point through a soft harmonic constraint with $k = 5$ kcal/(mol Å^2). To stretch the helix, the target point was pulled in steps of 0.05 Å. After each step, the system was allowed to equilibrate for 100 ps. Hereafter, these are referred to as small steps. When 20 small steps were completed, equivalent to a 1 Å extension, the system was equilibrated for a longer time (1 ns). This will be referred to as a long step (Figure 1). The time necessary to reach equilibrium at the end of the long steps was determined empirically. Different times were tested monitoring the temperature and the energy until these values became stable and no structural stress was present in the structure. The equilibration time was then set to twice the time necessary for stability. During these equilibrations, the average properties of the system (macrostate) fluctuate about the same values in the different trajectories. Thus, even if the time is not long enough for the peptides to sample some of the same conformations in all trajectories, the states accessed in the different tra-

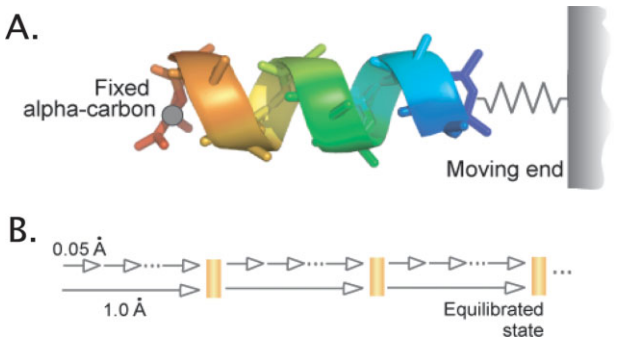

Figure 1

Unfolding the α-helix by pulling. (**A**) The Cα of the C-terminus is fixed while the Cα of the N-terminus is harmonically constrained while being pulled in sequential steps. (**B**) The peptide is unfolded by stretching it in steps of 1 Å, divided in 20 small steps of 0.05 Å each. After each extension of 1 Å the system is allowed to equilibrate for 1 ns.

jectories do belong to the same ensemble (macrostate). The total number of long steps performed (15 steps) is equivalent to stretching the molecule 15 Å. Each 15 Å extension constitutes a trajectory. (Subdivision of trajectories has been used before for sampling rare events in the context of boundary value formation for barrier transitions,[28] but this study is not relevant to the method presented here.)

For each trajectory the force exerted over the system was calculated as $F = -k(r_{CA} - r_{ref})$, where r_{CA} is the position (x, y, z) of the Cα of the N-terminus at each step and r_{ref} corresponds to the position of the target point where the harmonic constraint is set. The variable r_{ref} increases by 0.05 Å in each small step corresponding to a stretching velocity of 0.05 Å/ps. The work done on the system for each long step is given by $W = \int_{x_{CA,N}}^{x_{CA,N+1}} F \cdot dx_{CA}$, were F is the force projected onto the pulling axis (x-axis). Force and work behavior for various trajectories are shown in Figure 2. The calculated force shows the fluctuating nature of the system. Work values show great variability from trajectory to trajectory. Both positive and negative values of the work are observed. The positive portions of the curves correspond to events in which energy is input to the system to drive the unfolding process. Negative work values result when the helix is already unstable and pushes against the "spring" to continue unfolding.

MD calculations were used to compute 15 work values for each trajectory: one value for each of the 15 long (1 Å) steps. For each peptide, 15 trajectories were simulated. Microstates at the end of each long step for different trajectories belong to the same ensemble, after ensuring the proper equilibration. This equivalence can be used to build new trajectories by combining long steps of different trajectories at equilibration points. Individual ordered 1 Å long steps were randomly selected from the 15 trajectories until a new trajectory was built. By this process (MSTC), an ensemble of 10^7 trajectories were generated and used to construct the work probability distribution $P(W)$ (see Fig. 2).

Although test runs indicated that peptides are unfolded when stretching them by 15 Å (total N-terminal to C-terminal distance of 30.5 Å), the number of long steps needed to unfold the peptide could be smaller than 15. If this were the case, some regions of the peptide could become overly stretched. To determine the minimum number of long steps needed to unfold the helix, the hydrogen-bonded donor–acceptor distances and the internal energy of the peptide were analyzed. For all trajectories of the alanine peptide, the distances between backbone H-bond partners were calculated at every frame of the simulation and the average value was computed every 0.5 Å of extension. Figure 3 shows distances from the donor (N) to the acceptor (O) of the backbone between residues i, $i + 4$ for a typical trajectory. Initially, all backbone hydrogen bonds are formed. As one end is pulled these distances increase and hydrogen bonds are broken. After pulling 5 Å, averaging over all trajectories, 53% of all H-bond are still formed. At 10 Å, <10% of the H-bonds are formed; therefore this distance was used to calculate free energies. The results also show that between 10 and 11 Å the peptide internal energy and the peptide-solvent energy decrease significantly while the total energy, as expected, remains constant (see Fig. 3). These changes mean that at around 10 Å the unfolded helix becomes the most stable species, further supporting the choice of this extension for free-energy estimations.

The method outlined here (MSTC) was used to calculate the free energy change of unfolding by force of different 12-residue helices. Calculations were done for an all-alanine helix Ala_{12}, and for host-guest systems $Ala_{10}X_2$, in which the alanine-based helix is the host and two positions are substituted with another amino acid residue. Substitutions were made at the fourth and ninth positions to avoid side-chain side-chain interactions. Substitutions to 17 of the other 19 naturally occurring amino acids (proline and glycine were not modeled) were made. For each peptide, 15 trajectories were simulated and used to calculate $P(W)$ and ΔG of the transition. The values obtained for the different peptides can be used to calculate $\Delta\Delta G = \Delta G_{Ala_{10}X_2} - \Delta G_{Ala_{12}}$, a measure of the helix stabilization propensity of each amino acid vis-a-vis alanine.

The importance of sampling and of the number of trajectories needed

Jarzynski's equality is valid when averaging is carried out over a large number of repetitions of a process that takes the system from a certain equilibrium state A to another equilibrium state B. However, many problems arise when attempting to sample enough trials to provide

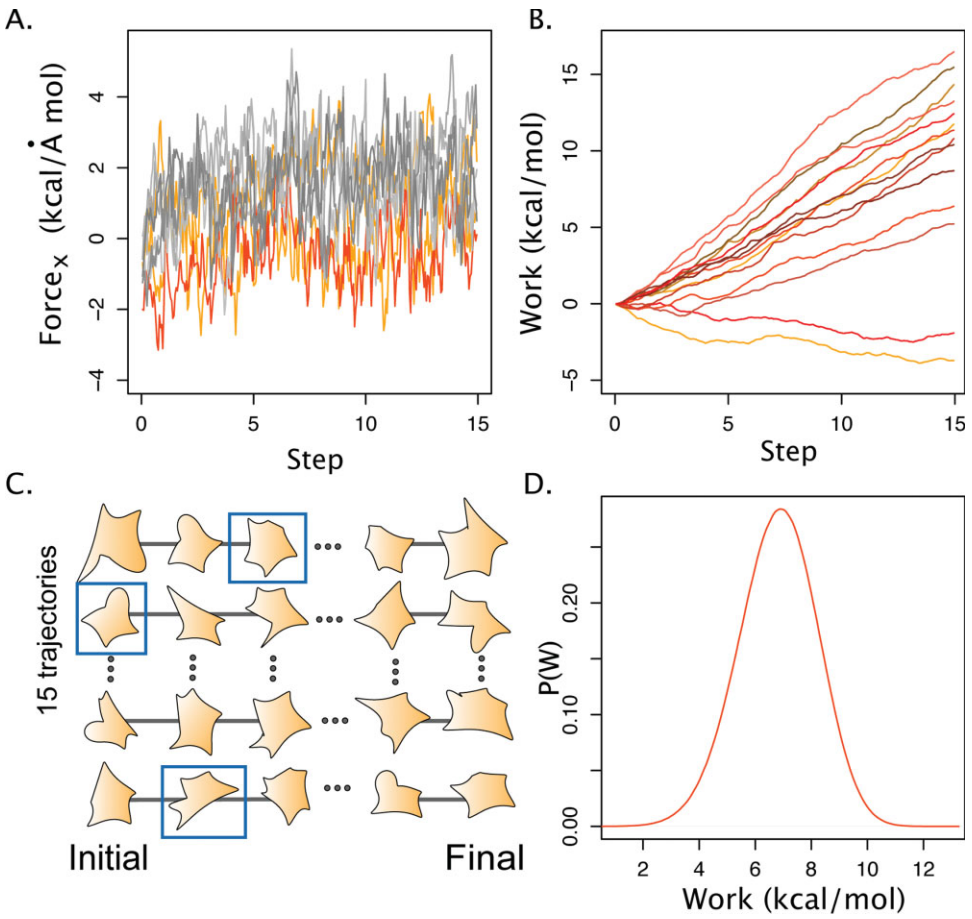

Figure 2

(**A**) Force values for different trajectories during the pulling process. (**B**) Work values calculated at every small step for different trajectories. (**C**) Path selection scheme. Every equilibrated state is shown in orange. We start with 15 trajectories each composed of 15 long steps. (**D**) Probability work distribution $P(W)$ built using the MSTC method up to 10^7 trajectories.

a representative work distribution. Of special interest is the ability to sample the low end (or tail) of the work distribution.[29–31] We know that $\langle W \rangle \geq \Delta G$, which is equivalent to saying that in each trajectory the difference $W - \Delta G$ corresponds to the dissipated work W_{diss}.[29,32] In a series of nonequilibrium experiments, the work is different from one irreversible trajectory to another. If we assume that the system is initially in an equilibrium state, the relation $\langle W \rangle \geq \Delta G$ holds on average, but work of multiple trajectories will have a certain distribution. In a generic distribution, for any given trajectory it is possible that $W_{\text{diss}} < 0$. Trajectories in this class are usually called "violating trajectories" or "transient violations of the second law."[20] In these situations, the work done on the system can be very small or work can even be done by the system. To recover ΔG, Jarzynski's equality enhances, through the exponential averaging, the weight of this low energy tail of the distribution.

The ability to sample all values of the work is of fundamental importance when building $P(W)$. If we think in terms of typical realizations as those in which the work is near the average value, and dominant realizations as those that will be heavily weighted in the exponential averaging $<e^{-\beta W}>$, it is necessary to ensure that enough dominant realizations are sampled.[29,33] In the context of the method presented in this work, two variables should be taken into account: (1) the number of original trajectories used for the combination process and (2) the number of combinations needed for convergence. We found that for this system 15 computed trajectories combined to give 10^7 total work values converge to a work distribution that does not change by adding more trajectories (see Fig. 4).

Once $P(W)$ is generated, work values are averaged using Jarzynski's equality. For this, direct averaging is used considering the 10^7 work values calculated. A similar method is to consider the distribution itself, that is, $\langle e^{-\beta W} \rangle = \int P(W) e^{-\beta W} dW$. Results of the two evaluations are comparable. After averaging, the free energy is obtained as $\Delta G = -\beta^{-1} \text{Ln}(\langle e^{-\beta W} \rangle)$.

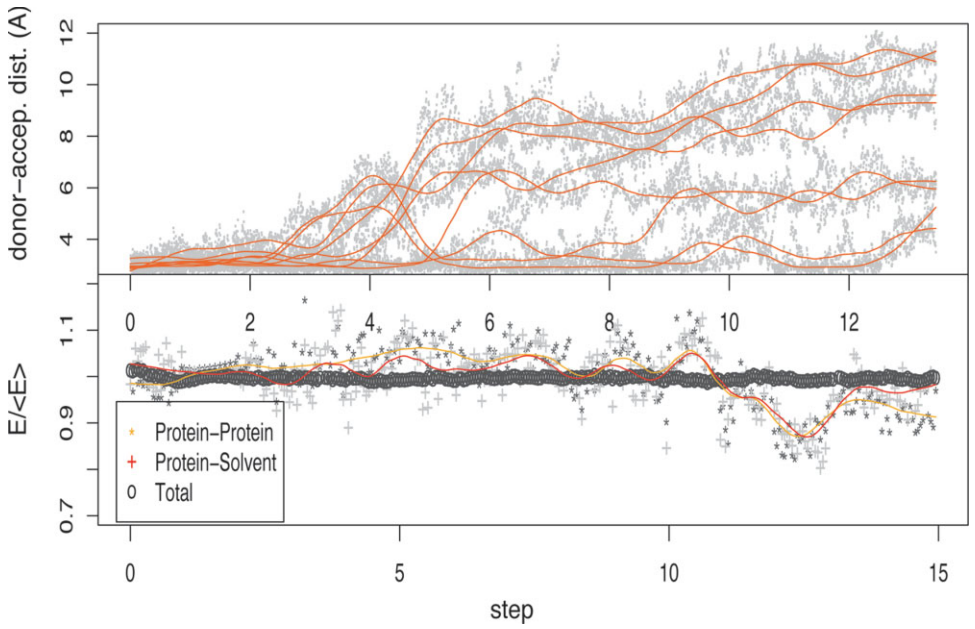

Figure 3

Top: distance between backbone donor and acceptor for all pairs that form hydrogen bond in the helical state. As one end is pulled hydrogen bonds are broken until the unfold state is reached. Bottom: energy normalized to the average energy for the pulling process. Protein–protein energy, protein–solvent interaction, and total energy are shown.

The free energy of extending the helix can also be calculated using the expression $\Delta G = \beta^{-1} \sum_{i_s=1}^{m} \mathrm{Ln} \frac{1}{n} \sum_{i_t}^{n} e^{-\beta W_{i_t,i_s}}$ (for derivation see Appendix), where $m = 10$ and $n = 15$. Values calculated this way are equivalent to those calculated for all possible n^m trajectories. As described above, we proved empirically that to obtain values of

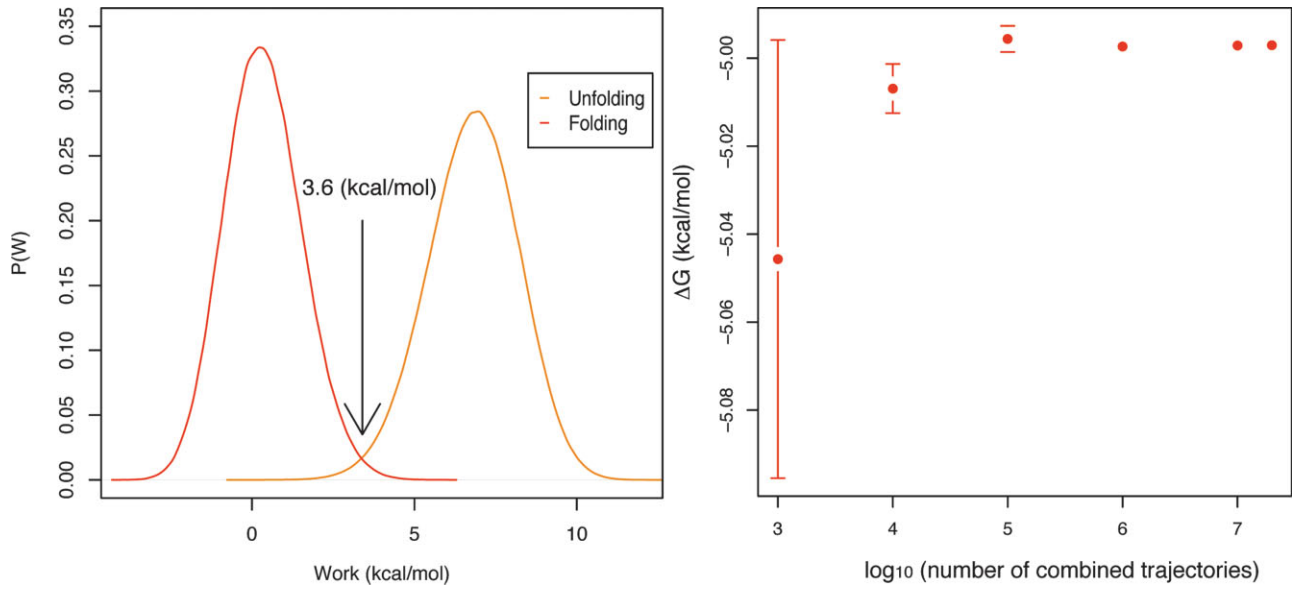

Figure 4

Left: work distribution functions for the forward (orange, unfolding) and reverse (red, folding) simulations of the Ala_{12} peptide. The value of ΔG, corresponding to the intersection between the two distributions is 3.6 kcal/mol. Right: ΔG calculated using different number of final trajectories after combination. A total of 10^7 new trajectories assures a converged value.

similar quality it would be necessary to compute at least 10^7 full trajectories.

Free energy estimations using Jarzynski's equality have associated systematic errors. These systematic errors are referred as the bias (difference between the Jarzynski estimations of the free energy and the true value of the free energy). Equations for estimating this bias have been presented by Gore *et al.*[29] for cases in which perturbations of the system are small, both for small and large number of trajectories. The bias corrections considered are described as a function of the estimated average dissipated work $\hat{\overline{W}}_{\mathrm{dis}}$. As a first approximation they propose the use of $\hat{\overline{W}}_{\mathrm{dis}} = \langle W \rangle - \Delta G_J$, where ΔG_J is the Jarzynski estimate of ΔG. Then if the estimated bias is $\hat{\mathrm{B}}$, the free energy can be corrected as $\Delta G_B = \langle \Delta G \rangle_J - \hat{B}_J$. When perturbations are near equilibrium, in the large N limit, the bias is $B(N) = (e^{2\beta \overline{W}_{\mathrm{dis}}} - 1)/2\beta N$. For small N this is no longer a good estimation and the bias is underestimated. For the calculations we presented for alanine peptides the average dissipated work is $\overline{W}_{\mathrm{dis}} \sim 1.8$ kcal/mol, then the numerator of the bias expression is ~430, which is $\ll N(10^7)$, therefore we are in the large N regime. In this regime, the bias in the estimation of ΔG is small (<0.002 kcal/mol).

Table I
Free Energy Differences with Respect to Alanine Peptide

	$\Delta\Delta G_{\mathrm{exp}}$	ΔG_{sim}	$\Delta\Delta G_{\mathrm{sim}}$	$\Delta G_{\mathrm{sim,ent}}$	$\Delta\Delta G_{\mathrm{sim,ent}}$
ALA	0.00	−5.002	0.000	−5.553	0.000
ARG	0.21	−4.024	0.489	−4.925	0.628
ASN	0.88	−2.920	1.041	−4.372	1.181
ASP	0.88	−3.515	0.743	−4.670	0.883
CYS	0.82	−2.788	1.107	−4.307	1.246
GLN	0.57	−4.307	0.348	−5.066	0.487
GLU	0.68	−3.947	0.527	−4.886	0.667
HIS	0.78	−4.003	0.499	−4.914	0.639
ILE	0.70	−4.118	0.442	−4.729	0.824
LEU	0.28	−4.670	0.166	−5.247	0.305
LYS	0.36	−4.239	0.381	−5.032	0.521
MET	0.51	−3.864	0.569	−4.844	0.708
PHE	0.93	−2.742	1.130	−4.283	1.270
SER	0.78	−3.827	0.587	−4.826	0.727
THR	1.32	−1.776	1.613	−3.801	1.752
TRP	0.78	−1.400	1.801	−3.370	2.183
TYR	0.60	−4.577	0.212	−5.201	0.352
VAL	1.05	−2.214	1.394	−3.777	1.776

Residue name corresponds to the guest residue in every chain. $\Delta\Delta G_{\mathrm{exp}}$: free energy difference reproduced from Chakrabartty et al.[37] for experimental measurements of helix propensities with respect to alanine. ΔG_{sim}: free energy calculated by simulations when the N-terminus of the peptide has been pulled 10 Å and both ends remained fixed during every equilibration. $\Delta\Delta G_{\mathrm{sim}}$: free energy difference respect to alanine per residue. $\Delta G_{\mathrm{sim,ent}}$: free energy obtained by simulation and corrected assuming a lost of entropy. Corrections were done assuming that from the 12 residues four are able to sample conformational space. $\Delta\Delta G_{\mathrm{sim,ent}}$: free energy difference with respect to alanine after entropic corrections.

Reverse simulations and the fluctuation theorem

The Crooks fluctuation theorem (CFT) relates typical and dominant realizations from forward (F) unfolding and reverse (R) folding simulations[16,34–36] using the ratio $P_F(W)/P_R(-W) = e^{\beta(W-\Delta G)}$. It is clear that the work value at which the two work distributions intersect corresponds to the value of the thermodynamic ΔG. We used CFT to test the validity of our work distributions. For this, reverse simulations were performed in the same stepwise fashion used for the forward realizations. Reverse simulations started at the final structures obtained during every long step from the forward realization. Fifteen trajectories were produced for the folding reaction. Equilibration times for forward and reverse simulations were equivalent.

With both sets of simulations it is possible to estimate the size of the ensemble needed to obtain a dominant realization. The number of trajectories needed for convergence of the free energy values have been estimated by Jarzynski[33] by considering the different probabilities of typical and dominant realizations. For the forward experiment the number of realizations needed is $N_F \sim e^{-\beta\langle W_R^d \rangle}$; analogously for the reverse experiment $N_R \sim e^{\beta\langle W_F^d \rangle}$.[33] The subscripts F and R stand for forward and reverse and d indicates dissipated work. Using the average values of W^d, we estimated that the number of realizations for the forward process was $\sim 10^5$, while for reverse simulations was $\sim 10^7$. Both distributions are shown in Figure 4. In this example, the free energy of the unfolding process using the CFT is 3.6 kcal/mol. If the free energy is calculated using only the forward realizations and Jarzynski's equality, a value of 5.0 kcal/mol is obtained. The similarity of these values suggests that the overall sampling of the distribution gives enough information about the dominant realizations.

RESULTS AND DISCUSSION

Free energy differences calculated with the simulations presented here were compared with experimental results of helix propensities obtained by Baldwin and coworkers[37] who also used alanine-based peptides containing guest amino acids with no side-chain side-chain interactions. The results are presented as difference in the free energies of helix formation and per residue ($-\Delta\Delta G_{\mathrm{extension}}/2$ given the two mutations per peptide). Values from the simulations show strong correlation with experimental values (Table I and Fig. 5) with scale slope (regression coefficient) of 1.1 and R^2 of 0.86. This high correlation shows that the method presented here allows the use of Jarzynski's equality for estimating differences in the free energy of unfolding in realistic computational times.

Entropic considerations

To make a direct comparison between the experimental results of Baldwin and coworkers[37] and the computed

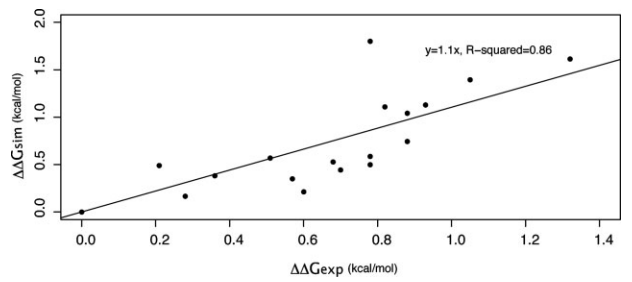

Figure 5

Correlation of helix stabilization propensity between experimental values reported by Baldwin and coworkers[37] and values calculated by simulations after entropic corrections.

values reported here, contributions of the entropy also need to be considered. In the experiments of Baldwin and coworkers,[37] the final states are unfolded peptides, while for the MD simulations presented here in the final state, the two ends of the peptides are positionally constrained. Initially, dihedral angles are in the helical region of the Ramachandran plot and backbone hydrogen bonds are intact. As the end-to-end distance is increased, hydrogen bonds are broken and dihedral angles can sample other regions of configuration space. In the extreme case of an end-to-end distance equivalent to that of the fully extended peptide length, dihedral angles are again fixed. For intermediate states, where all hydrogen bonds are broken but the peptide is still not fully extended (such as after a 10 Å extension), there are still constraints on some dihedral angles. For each peptide, only a small number of individual residues can adopt nonextended angles, while others are constrained to a certain subset of extended configurations.

Conformational entropy contributions to unfolding have been calculated in the past using structural parameterizations,[2,38] molecular mechanics dynamics,[39–41] and Monte Carlo simulations.[42,43] In the calculations performed here, even in the initial helical state the side chains are exposed, so for each amino acid only two contributions to conformational entropy need to be considered: (1) $\Delta S_{ex\rightarrow u,sc,}$ the entropy gained by side chain atoms when the backbone unfolds, and (2) ΔS_{bb}, the entropy gained by the backbone when it unfolds. The first contribution is highly similar, if not identical, between the experiments and the simulations, and no correction is necessary.

For the backbone entropy the results of the MD simulations include the contribution of going from a helix to a conformation with most of the H-bonds broken but with the ends fixed. To compare the MD results with the experimental results of Baldwin and coworkers,[37] it is necessary to estimate the change in entropy of taking the final state of the MD simulations partially unfolded with the ends fixed to a fully unfolded state. One way to accomplish this correction is to subtract from the estimated ΔG, the contribution of the term $\Delta S_{\text{helix->extended}}$ and to add back $\Delta S_{\text{helix->unfolded}}$.

Every peptide is formed by 10 hosts alanine and two guest residues. If j (with $j \leq 12$) is the number of residues within the extended peptide with the ends fixed that can sample multiple conformations, the backbone conformational entropy is given by:

$$\Delta S_{\text{helix}\rightarrow\text{extended}} = j \cdot \left(\frac{10}{12}\Delta S_{\text{Ala}} + \frac{2}{12}\Delta S_{\text{res}}\right) + R\ln\binom{12}{j}.$$

For ΔS_{Ala} and ΔS_{res}, we used the exposed to unfolded entropies from D'Aquino *et al.*[41] (Table I). Note that the fully unfolded state corresponds to $j = 12$ (i.e., no residue is constrained by the fixed points), and therefore the sought after value of the entropy difference is given by:

$$\Delta S_{\text{extended}\rightarrow\text{unfolded}} = (12 - j)\left(\frac{10}{12}\Delta S_{\text{Ala}} + \frac{2}{12}\Delta S_{\text{res}}\right) - R\ln\binom{12}{j}.$$

To further investigate how well configurational space is sampled during the extension process and determine the value of j, the correlation among consecutive dihedral angles was calculated. For every 1 Å step, dihedral angles ϕ and φ were calculated and the correlation between samples of 15 consecutive small steps computed. To avoid unwanted sampling effects, the sample size was chosen so that frames used in the estimation of the correlation did not include the ends of the long steps. High correlation between two frames can be interpreted as both frames having the same conformation in terms of dihedral angles, while low correlation indicates a change in conformation. A correlation threshold of 0.2 was used as an indication of a conformational change between frames. On average, a correlation lower than 0.2 corresponds to a standard deviation in 10° for ϕ or φ. When considering these values and looking over all trajectories, on average four amino acids per peptide are free to sample all angles when the helix has been pulled 10 Å. Thus, $j = 4$, was used to calculate the entropy of the backbone when the peptide is unfolded with the ends fixed 10 Å apart, and $j = 12$ was used to describe the completely unfolded peptide with free ends. In both cases all side chains are free to sample conformational space (Table I). These values, which correspond to the difference in entropy between the final state in the simulation and the final state in the solution experiments, were used to correct the computed free energies (Table II). Backbone entropy corrections are similar for all peptides independent of the host, with slightly lower values for β-branched residues (Ile, Thr, and Val). Also, helix pro-

Table II
Conformational Entropies for Amino Acids

	$\Delta S_{ex\rightarrow u,bb}$	$T\Delta S_{ex\rightarrow u,4}$
ALA	4.1	6.10
ARG	3.4	5.82
ASN	3.4	5.82
ASP	3.4	5.82
CYS	3.4	5.82
GLN	3.4	5.82
GLU	3.4	5.82
HIS	3.4	5.82
ILE	2.18	5.34
LEU	3.4	5.82
LYS	3.4	5.82
MET	3.4	5.82
PHE	3.4	5.82
SER	3.4	5.82
THR	3.4	5.82
TRP	2.18	5.34
TYR	3.4	5.82
VAL	2.18	5.34

ΔS_{bb} values were reproduced from D'Aquino et al.[41] Values are per amino acids. $T\Delta S_{ex\rightarrow u}$ corresponds to the calculated entropy difference for the 12 amino acid peptides. Residue name corresponds to the guest when is assumed that four residues are free to sample the nonextended conformational space, measured in kcal/mol.

pensities ($\Delta\Delta G$) after this correction show the same high correlation with the experimental results observed for the not corrected values, with a scale slope (regression coefficient) of 1.33 and $R^2 = 0.88$. Therefore, although the entropy corrections are important for ΔG estimations, they show no significant effect on free energy differences ($\Delta\Delta G$).

Free energies of helix unfolding were also computed by Park *et al.*[23] In this study, Jarzynski's equality was used to calculate a potential of mean force, but because these calculations were done in vacuum, comparison with experimental data was not possible. In the absence of solvent, the mechanism by which stabilizing hydrogen bonds are broken is purely mechanical, while in aqueous solvent peptide hydrogen bonds are bridged by water molecules as an intermediate state to bond breaking.[44] As a consequence Park estimates give a large overestimation of free energy values compared to the results presented in the present work.

SUMMARY AND CONCLUSIONS

We have presented estimations of helix propensities computed using MD simulations. In particular, we calculated differences in free energies of unfolding for alanine-based host-guest peptides. These differences correlate very well with the helix stabilization propensities obtained experimentally.

These calculations were performed with a robust method (MSTC) that allows calculation of free energy differences using Jarzynski's equality. The method presented here is sensitive enough to estimate free energy differences that are highly correlated with experimental data. Additionally, a simple calculation is presented to correct for entropic effects that arise because the ends of the peptide are fixed in the end state of the calculations.

The method can be widely applied to different systems and can be used to estimate free energy differences even if the number of atoms is different between the processes. The main strength of the method is that the system may be driven out of equilibrium during the process as long as it is equilibrated at the end of each step. By dividing the transition into steps between equilibrated states, microstates at the end of each step are members of the same ensemble given that they are equilibrated with identical external conditions (i.e., energy, spring constant, and position of the fixed point). A combinatorial process is used to generate a representative work distribution from a small number of computed trajectories. This combination of processes between equivalent microstates gives rise to a work distribution for the transition, which is used to estimate the free energy. To estimate how irreversible the process was and to test the effectiveness of the sampling used, a simulation of the reversed process was also performed. The values of ΔG obtained using Jarzynski's equality and CFT were highly similar, suggesting that the method presented here provides a realistic work distribution.

It should be noted that the method is general enough to be applied to other processes such as estimation of differences in binding affinity between different ligands binding to the same protein, binding of a ligand to wild type and mutant proteins, unfolding of RNAs with individual substitutions, unfolding of proteins with a small number of mutations, and other similar processes.

REFERENCES

1. Creamer TP, Rose GD. Alpha-helix-forming propensities in peptides and proteins. Proteins 1994;19:85–97.
2. Luque I, Mayorga OL, Freire E. Structure-based thermodynamic scale of alpha-helix propensities in amino acids. Biochemistry 1996; 35:13681–13688.
3. Zuckerman DM, Woolf TB. Theory of a systematic computational error in free energy differences. Phys Rev Lett 2002;89:180602–180606.
4. Shirts MR, Pande VS. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. J Chem Phys 2005;122:144107–144123.
5. Lechner W, Oberhofer H, Dellago C, Geissler PL. Equilibrium free energies from fast-switching trajectories with large time steps. J Chem Phys 2006;124:044113–044125.
6. Brooks CL, III, Karplus M. Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. J Mol Biol 1989;208:159–181.
7. Straatsma TP, McCammon JA. Computational alchemy. Annu Rev Phys Chem 1992;43:407–435.
8. Tobias DJ, Brooks CL, III. Thermodynamics and mechanism of alpha helix initiation in alanine and valine peptides. Biochemistry 1991;30:6059–6070.
9. Shea JE, Brooks CL, III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499–535.

10. Hummer G. Fast-growth thermodynamic integration: error and efficiency analysis. J Chem Phys 2001;114:7330–7337.
11. Pearlman DA, Kollman PA. A new method for carrying out free energy perturbation calculations: dynamically modified windows. J Chem Phys 1989;90:2460–2470.
12. Chipot C, Kollman PA, Pearlman DA. Alternative approaches to potential of mean force calculations: free energy perturbation versus thermodynamic integration. Case study of some representative nonpolar interactions J Comput Chem 1996;17:1112–1131.
13. Lu N, Kofke DA, Woolf TB. Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. J Comput Chem 2004;25:28–39.
14. Jarzynski C. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. Phys Rev E 1997;56:5018–5035.
15. Jarzynski C. Nonequilibrium equality for free energy differences. Phys Rev Lett 1997;78:2690–2693.
16. Crooks GE. Path-ensemble averages in systems driven far from equilibrium. Phys Rev E 2000;61:2361–2366.
17. Evans DJ. A non-equilibrium free energy theorem for deterministic systems. Mol Phys 2003;101:1551–1554.
18. Chernyak V, Chertkov M, Jarzynski C. Dynamical generalization of nonequilibrium work relation. Phys Rev E Stat Nonlin Soft Matter Phys 2005;71 (2 Part 2):025102–025106.
19. Liphardt J, Onoa B, Smith SB, Tinoco IJ, Bustamante C. Reversible unfolding of single RNA molecules by mechanical force. Science 2001;292:733–737.
20. Wang GM, Sevick EM, Mittag E, Searles DJ, Evans DJ. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. Phys Rev Lett 2002;89: 050601–050605.
21. Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I, Jr, Bustamante C. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature 2005;437:231–234.
22. Atilgan E, Sun SX. Equilibrium free energy estimates based on nonequilibrium work relations and extended dynamics. J Chem Phys 2004;121:10392–10400.
23. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys 2003;119:3559–3566.
24. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 2004;120: 5946–5961.
25. Oberhofer H, Dellago C, Boresch S. Single molecule pulling with large time steps. Phys Rev E Stat Nonlin Soft Matter Phys 2007;75 (6 Part 1):061106–061113.
26. Bastug T, Chen PC, Patra SM, Kuyucak S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. J Chem Phys 2008;128:155104–155113.
27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
28. Bai D, Elber R. Calculation of point-to-point short-time and rare trajectories with boundary value formulation. J Chem Theory Comput 2006;2:484–494.
29. Gore J, Ritort F, Bustamante C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. Proc Natl Acad Sci USA 2003;100:12564–12569.
30. Imparato A, Peliti L. Work-probability distribution in systems driven out of equilibrium. Phys Rev E Stat Nonlin Soft Matter Phys 2005;72 (4 Part 2):046114–046126.
31. Calderon CP, Chelli R. Approximating nonequilibrium processes using a collection of surrogate diffusion models. J Chem Phys 2008;128:145103–145113.
32. Vaikuntanathan S, Jarzynski C. Escorted free energy simulations: improving convergence by reducing dissipation. Phys Rev Lett 2008; 100:190601–190605.
33. Jarzynski C. Rare events and the convergence of exponentially averaged work values. Phys Rev E Stat Nonlin Soft Matter Phys 2006; 73:046105–046110.
34. Crooks GE. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. J Stat Phys 1998;90:1481–1487.
35. Crooks GE. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1999;60:2721–2726.
36. Ritort F. Work and heat fluctuations in two-state systems: a trajectory thermodynamics formalism. J Stat Mech Theory Exp 2004;2004:10016–10049.
37. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci 1994;3:843–852.
38. Pickett SD, Sternberg MJ. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
39. Doig AJ, Sternberg MJ. Side-chain conformational entropy in protein folding. Protein Sci 1995;4:2247–2251.
40. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins 1994;20:68–84.
41. D'Aquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. Proteins 1996;25:143–156.
42. Creamer TP. Side-chain conformational entropy in protein unfolded states. Proteins 2000;40:443–450.
43. Cheluvaraja S, Meirovitch H. Calculation of the entropy and free energy from Monte Carlo simulations of a peptide stretched by an external force. J Phys Chem B 2005;109:21963–21970.
44. Pabon G, Amzel LM. Mechanism of titin unfolding by force: insight from quasi-equilibrium molecular dynamics calculations. Biophys J 2006;91:467–472.

APPENDIX

The work performed during one simulation trajectory can be written as $W_t = \sum_{i_s=1}^{m} W_{i_s}$, where W_{i_s} is the work done in the i_s step. Then, the work done during one long step in a set of simulations is given by W_{i_s,i_t} being i_s the segment index and i_t the trajectory index. Using this nomenclature, Jarzynski's equality can be rewritten summing over trajectories for same steps. This is,

$$\begin{aligned} e^{-\beta\Delta G} &= \left\langle e^{-\beta W} \right\rangle \\ &= \left(\frac{1}{n}\sum_{i_t=1}^{n} e^{-\beta W_{i_t,1}}\right)\left(\frac{1}{n}\sum_{i_t=1}^{n} e^{-\beta W_{i_t,2}}\right)\cdots\left(\frac{1}{n}\sum_{i_t=1}^{n} e^{-\beta W_{i_t,m}}\right) \\ &= \frac{1}{n^m}\prod_{i_s}^{m}\sum_{i_t=1}^{n} e^{-\beta W_{i_t,i_s}} \end{aligned}$$

where steps are considered independent. Then, the free energy of the transition in terms of steps is given by:

$$\Delta G = -\beta^{-1}\sum_{i_s=1}^{m} \ln\frac{1}{n}\sum_{i_t}^{n} e^{-\beta W_{i_t,i_s}}$$

Where we used the fact that the logarithm of a product is the sum of the logarithms.