Free Energy Simulations: Thermodynamic Reversibility and Variability

Shobana S.,*,† Benoît Roux,‡,§ and Olaf S. Andersen†

Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, New York 10021; Departments of Physics and Chemistry, Université de Montréal, C. P. 6128, succursale Centre-Ville, Montréal (Québec), Canada H3C 3J7; and Department of Biochemistry and Structural Biology, Weill Medical College of Cornell University, New York, New York 10021

Received: November 29, 1999; In Final Form: March 21, 2000

The energetic consequences of amino acid substitutions were investigated using molecular dynamics free energy simulations (MD/FES). A focus of the present study is how one can treat some of the intrinsic problems associated with the use of dummy atoms, which are introduced in standard MD/FES procedures working with a constant number of particles. Specifically, we show how one can introduce dummy atoms, which retain all the covalent interactions, in a "hybrid residue", in such a way that the influence of the bonded interactions with dummy atoms do not influence the final free energy change. The simulations thus can be done using a transformation protocol in which all covalent bond contributions are maintained invariant throughout the calculations; only the nonbonded interactions are varied. That is, the number of atoms is maintained constant by introducing dummy atoms, which are covalently linked to the protein in question, but which have no nonbonded interactions at one or the other of the two end point reference states. The potential energy function describing the transformation was constructed such that all internal energy terms are invariant with respect to the thermodynamic coupling parameter λ ($0 \leq \lambda \leq 1$). This simulation procedure therefore has similarities with both the "single topology" and "dual topology" methods, which have been used in other investigations. This transformation and MD/FES strategy was evaluated using three different systems, in which an Ala is transformed into a Val (the italicized residues were transformed): formyl-*Ala*-ethanolamine; formyl-$(Ala)_3$-Ala-*Ala*-Ala-$(Ala)_3$-ethanolamine (as an α-helix); and formyl-$(Ala)_3$-Val-*Ala*-Val-$(Ala)_3$-ethanolamine (as an α-helix). The calculations examined how the free energy difference associated with an Ala-to-Val transformation depends on the choice of transformation path, the equilibration time, and the number windows used in the calculation. The results show that MD/FES provides for stable energy estimates for any intermediate state, as defined by the coupling parameter λ, including the end point reference states: $\lambda = 0$ and $\lambda = 1$. With an appropriate choice of sigmoid transformation path, free energy changes with minimal variability and good reversibility were obtained for all three systems using MD/FES lasting longer than 2.4 ns.

Introduction

The emergence of molecular systems that (at least in principle) can fold in different ways[1,2] provides for precise measurements of the role of amino acid substitutions in terms of the free energy difference (ΔG) between the two (or more) well-defined conformations. The available energetic window in structure−function studies is typically about ±4 kcal/mol, denoting a 1000-fold change (up or down) in function.[3] The interpretation of such measurements, in terms of dissecting the overall ΔG into its microscopic components, will be facilitated by computational studies of the free energy of the different conformations. The molecular dynamics free energy simulation method (MD/FES), which is based on an analysis of thermodynamic cycles,[4−6] remains the approach of choice for exploring the energetic consequences of amino acid substitutions on protein structure and function using detailed atomic models.

The foundation of the method is that the free energy is a state function, such that the sum of the free energy changes for a series of reversible transitions in a closed path is zero. This makes it possible to examine the consequences of chemical modifications on the relative free energies of a molecule in different conformational states. MD/FES can be used to calculate free energy difference of an amino acid substitution that changes molecule *i* into *j* ($\Delta G_{i\rightarrow j}$) for any given conformation, or the free energy difference between two well-defined conformations a and b ($\Delta G^{a\rightarrow b}$) for any given molecule. This concept is illustrated schematically in Figure 1.

The two vertical transitions denote conformational transitions of molecule *i* and molecule *j* between conformations a and b, as described by the free energy differences $\Delta G_i^{a\rightarrow b}$ and $\Delta G_j^{a\rightarrow b}$, which are experimentally accessible. The two horizontal transitions denote nonchemical processes in which residue *i* is mutated into residue *j* in the conformation a and conformation b. The associated free energy differences $\Delta G_{i\rightarrow j}^{a}$ and $\Delta G_{i\rightarrow j}^{b}$ are not accessible experimentally, but can be estimated by computational alchemy.[7,8] To evaluate the free energy differences of

* Address correspondence to: Shobana S., Ph.D., Department of Physiology and Biophysics, Weill Medical College of Cornell University, 1300 York Avenue, Rm C-505, New York, NY 10021-4805. Tel: (212) 746-6351. Fax: (212) 746-8690. E-mail: sshobana@mail.med.cornell.edu.

† Department of Physiology and Biophysics, Weill Medical College of Cornell University.

‡ Université de Montréal.

§ Department of Biochemistry and Structural Biology, Weill Medical College of Cornell University.

10.1021/jp994193s CCC: $19.00
Published on Web 05/05/2000

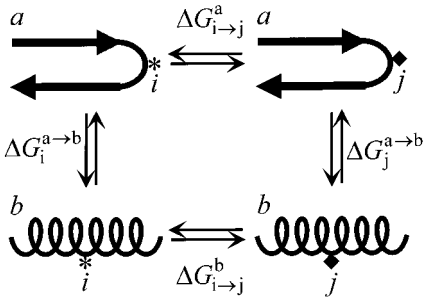

Figure 1. Thermodynamic cycle. "a" and "b" denote two different (folded) conformations of the molecule of interest. "*i*" and "*j*" denote two different amino acid sequence variants. That is, a_i and a_j denote states when *i* and *j* fold into conformation a; b_i and b_j denote the corresponding states when *i* and *j* are folded in conformation b.

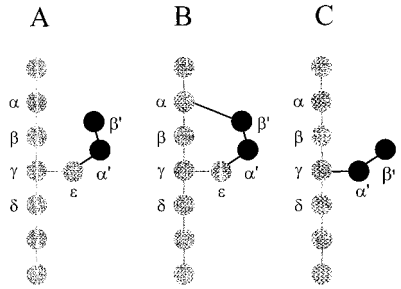

Figure 2. Connectivities in mutated proteins. The connectivities that retain the statistical equivalence in eq 9 and those that do not are shown in the left most (A) and two right (B and C) panels, respectively. The gray and black circles are the atoms in the native and the mutated residues, respectively. (A) Equivalence of the gray region is preserved when the black atoms are turned off in the final state. To maintain the equivalence, only one dihedral can be included in the final state when the black atoms are invisible, but all dihedrals can be included when the black atoms are interacting. (B and C) Equivalence of the gray region is not preserved when the black atoms are turned off in the final state. It would be preserved if, and only if, one bond angle and one dihedral torsion angle are retained in the final state.

one of the alchemical steps, $\Delta G^{a}_{i\rightarrow j}$ or $\Delta G^{b}_{i\rightarrow j}$, one can use the coupling parameter approach,[9] in which the transformation from *i* to *j* is described using a potential energy function *U* that is a monotonic function of a coupling parameter λ ($0 \leq \lambda \leq 1$) such that $U(0) = U_i$ and $U(1) = U_j$.

A methodological difficulty in such simulations is how one approaches the two end points, $\lambda \rightarrow 0$ and $\lambda \rightarrow 1$, which can give rise to major problems. First, a difficulty is encountered in MD/FES involving the materialization (insertion) of particles into a dense liquid using a coupling parameter λ, due to the very nature of the strong repulsive core potential of the particles. Linear (or nonlinear) scaling of the potential energy and thermodynamic integration with a coupling parameter may lead to numerical instabilities as small values of λ resulting from the singularity due to the presence of an infinitely small and infinitely repulsive particle in a dense liquid.[10] In fact, such instability is present even as a single particle is being inserted into the solvent. Effective ways to deal with this specific problem have been proposed.[11−13]

A second, different, difficulty arises from the topological connectivities of the molecular species, which can affect the results as the mutated chemically bonded groups appear and disappear at the end points, $\lambda \rightarrow 0$ and $\lambda \rightarrow 1$. The unphysical metamorphosis in the side chain structure that occurs during the alchemical calculation will usually be associated with a change in the number of interacting atoms, which is traditionally implemented in MD/FES using the so-called "single" or "dual" topology techniques.[14−19] Both techniques aim at handling the

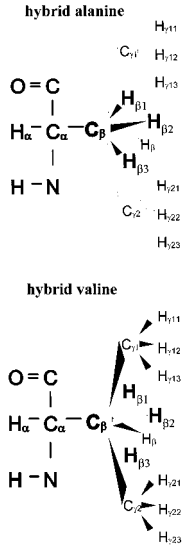

Figure 3. Schematic representation of (a) hybrid alanine and (b) hybrid valine. For clarity, the bonds are not drawn to scale.

unavoidable fact that, while the residue *i* and *j* may not have the same number of atoms, the number of particles cannot actually change during the MD/FES. This means that dummy, noninteracting atoms must be introduced to maintain the number of atoms constant in the simulations. In the single topology method, the atoms of residue *i* are mapped onto those of residue *j* as much as possible in order to minimize the number of dummy atoms,[17] whereas the side chains often are duplicated entirely in the dual topology method.[18]

Further, in the single topology method, the disappearing atoms are changed into dummy noninteracting atoms, while the covalent bonding energy terms are kept intact during the transformation process. Such dummy particles with no nonbonded interactions, but chemically bonded to the rest of the molecular system, however, can lead to nonphysical contributions to free energy differences, which require corrections for meaningful interpretations.[15] This difficulty is circumvented in the dual topology method, by turning off all the interactions (nonbonded and covalent) between the molecular system and the disappearing atoms. Thus, the end point of the calculation results in an "ideal gas" molecule reference state.[19] One problem with such a reference state is that, as the bonding force constants are reduced to very small values, the corresponding molecular fragments begin to wander extensively during the simulation. The fragment can never explore the complete accessible space of a true ideal gas fragment, however, and the free energy of the ideal gas atom end state cannot be calculated accurately.[19−21] The single topology method does not suffer from this problem, as the covalent energy terms are kept in the end-state (which causes other difficulties, see above).

Previously, the single and dual topology methods have been compared by Pearlman[15] who found that the two methods provide comparable estimates of the free energy changes associated with the transformation. Nevertheless, even though single and dual topology techniques have been used to study different molecular systems by various groups,[17,18,24,25] several fundamental and technical issues need to be resolved in order to make MD/FES a robust and reliable tool. It is in this context important that the distinction between the two techniques is, in part, semantic because both introduce dummy atoms at the end points and need thereafter to handle their interactions with the rest of the molecular system and it remains a challenge to determine the energies at the end points.[12] Not surprisingly,

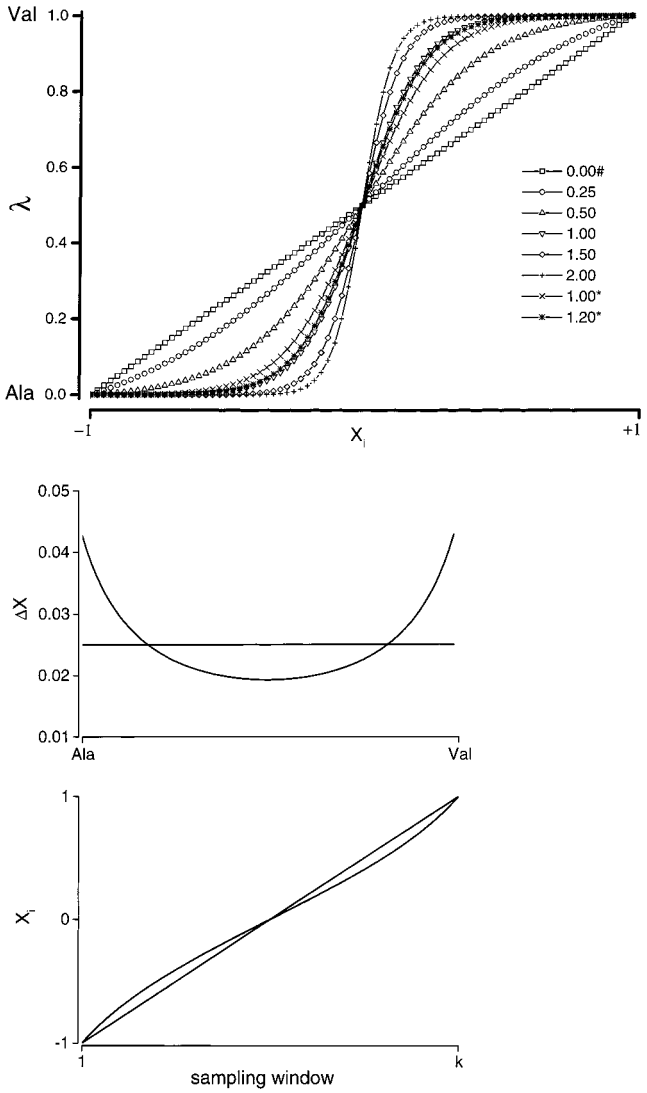

Figure 4. (A, upper panel) λ progression curves generated with different n values (cf. eqs 10 and 11). The curve labeled $n = 0.00\#$ is a straight line, the n^* curves are those with variable X increments. The simulation time at each sampling point is 12 ps (4 ps of equilibration time and 8 ps of production time). (B, lower panels) ΔX increment curves for the two strategies employed: (upper) constant (- - -) and variable (—) increments; (lower) X_i profile as a function of the sampling window for both constant (- - -) and variable (—) increments.

therefore, it is possible to devise intermediate schemes which retain some features of both techniques. Recent applications of MD/FES, based on the dual topology technique, have kept all the covalent bonding terms intact during the alchemical transformation to solve some of the problems associated with the ideal gas reference state.[26] Nonetheless, uncertainties remain about the interpretation of the calculations and the appropriate corrections that are needed for meaningful results. A critical assessment of MD/FES and alchemical transformation is necessary at the present time.

The focus of the present paper concerns the intrinsic problems associated with the treatment of the chemically bonded dummy atoms which must be introduced in standard MD/FES. Our aim is to clarify the interpretation of the role of dummy atoms in alchemical MD/FES and propose a practical and robust computational scheme for MD/FES. In particular, it is shown how to setup the dummy atoms in a "hybrid residue" that retains all the bonded interactions, in such a way that the influence of the bonded interactions with dummy atoms does not influence the final $\Delta\Delta G$. We find that the choice of methodology and simulation protocols result in different uncertainties in the numerical estimates of free energies, which will affect the reliability of the calculated conformational equilibria. The reliability of the MD/FES was evaluated by monitoring the reversibility and variability of the ΔG estimates obtained from the simulations.

An important motivation for the study was the discovery of some remarkable features exhibited by a series of analogues of the peptide antibiotic gramicidin A, gA, whose sequence is

Formyl-L-Val- - -Gly-L-Ala-D-Leu-L-Ala-D-
Val-L-Val-D-Val-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-
D-Leu-L-Trp-ethanolamine

gA usually forms only one type of channel, which can be determined (and identified) using electrophysiological methods.[27−29] A series of gA analogues, whose C-terminal sequence has been changed from (L-Trp-D-Leu)$_3$-L-Trp, to (L-Leu-D-Trp)$_3$-L-Leu, however, can form several different well-defined channel types.[2,29,30] These analogues are called gLWs, and the relative concentrations of the different channel types formed by gLWs in a bilayer can be determined using single-channel (single-molecule) methods. This allows for the determination of the free energy difference between the different well-defined conformations of a miniprotein. The absolute value of these free energy differences is ≤ 4 kcal/mol with an experimental uncertainty <0.3 kcal/mol.[2,30] Such measurements are indicative of the accuracy that is required for MD/FES calculations to be truly useful.

We first analyze the theoretical foundation for handling dummy atoms. Then, we examine the impact of several factors on the free energies $\Delta G_{i\rightarrow j}$ and $\Delta G_{j\rightarrow i}$ using three different model systems in which we transform an Ala into a Val: first, a single Ala that is blocked by formyl group (f) at the NH2-terminus and ethanolamine (eam) at the COOH-terminus; f-*Ala*-eam; second, a right-handed α-helical peptide f-(Ala)$_3$-Ala-*Ala*-Ala-(Ala)$_3$-eam; and third, a right-handed α-helical peptide f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam. The italicized Ala residues were the ones mutated. Formyl and ethanolamine were chosen as the blocking groups for the test peptides as they are the amino- and carboxy-terminal blocking residues in the linear gramicidins.[31] The two different α-helices were chosen so that it was possible to examine the Ala → Val transformation in the context of different chain interactions.

Theory

Alchemical Transformation of Protein Residues. We consider two different molecules that differ only by one amino acid (i or j) at some position m. Both can exist in conformation a and conformation b. The experimentally measurable probabilities of finding the system in conformation a or b, given that the amino acid at position n is type i or type j, are denoted by P_i^a, P_j^a, P_i^b, and P_j^b. By normalization, $P_i^a + P_i^b = 1$, and $P_j^a + P_j^b = 1$. For each molecular species, the free energy differences are given by probability ratios:

$$\Delta G_i^{a\rightarrow b} = -k_B T \ln\{P_i^b/P_i^a\}$$

$$\Delta G_i^{a\rightarrow b} = -k_B T \ln\{P_j^b/P_j^a\}$$

We wish to express the free energy difference in a manner that is computationally efficient for molecular dynamics simulations.

Let U_i and U_j denote the microscopic potential energy of the protein with amino acid i or j at position n. The relative free energy difference, $\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b} = [\Delta G_j^{a\rightarrow b} - \Delta G_i^{a\rightarrow b}]$ can then be expressed as

$$\exp(-\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b}/k_BT) = \left[\frac{\int_b dX_r\, dX_j\, e^{-U_j(X_r,X_j)/k_BT}}{\int_a dX_r\, dX_j\, e^{-U_j(X_r,X_j)/k_BT}}\right]\left[\frac{\int_b dX_r\, dX_i\, e^{-U_i(X_r,X_i)/k_BT}}{\int_a dX_r\, dX_i\, e^{-U_i(X_r,X_i)/k_BT}}\right]^{-1} \quad (1)$$

where X_i and X_j denote the coordinates of the n_i and n_j atoms, respectively, in residue type i and type j at position m, and X_r denotes the coordinates of the n_r remaining atoms in the molecule. The subscripts on the integrals indicate that the multidimensional integral is to be taken only over the conformational subspace relevant for conformations a or b, respectively.

The total number of atoms in the final, actual molecular systems, i.e., $n_r + n_i$ and $n_r + n_j$, will usually be different. Practically, however, the number of atoms cannot actually change in the course of a molecular dynamics trajectory free energy perturbation calculation[4] and all the degrees of freedom in the system $X \equiv \{X_r, X_i, X_j\}$, which include all the atoms in the residues at position m (i or j), must be included throughout the alchemical transformation. Introducing the thermodynamic coupling parameter λ[9] ($0 \le \lambda \le 1$), the microscopic potential energy of the intermediate states in the transformation $U(\lambda;X)$ can be expressed as

$$U(\lambda;X) = (1-\lambda)U_i(X_r,X_i) + \lambda U_j(X_r,X_j) \quad (2)$$

The free energy difference in eq 1 thus can be expressed as

$$\exp(-\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b}/k_BT) = \left[\frac{\int_b dX\, e^{-U(1;X)/k_BT}}{\int_a dX\, e^{-U(1;X)/k_BT}}\right]\left[\frac{\int_b dX\, e^{-U(0;X)/k_BT}}{\int_a dX\, e^{-U(0;X)/k_BT}}\right]^{-1} = \left[\frac{\int_b dX\, e^{-U(1;X)/k_BT}}{\int_b dX\, e^{-U(0;X)/k_BT}}\right]\left[\frac{\int_a dX\, e^{-U(1;X)/k_BT}}{\int_a dX\, e^{-U(0;X)/k_BT}}\right]^{-1} \quad (3)$$

or

$$\exp(-\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b}/k_BT) = \exp(-[\Delta G_{i\rightarrow j}^{a} - \Delta G_{i\rightarrow j}^{b}]/k_BT) \quad (4)$$

The $\Delta G_{i\rightarrow j}^{a}$ and $\Delta G_{i\rightarrow j}^{b}$ values can be calculated from MD/FES using free energy perturbation (FEP)[32,33] methods:

$$e^{-\Delta G_{i\rightarrow j}/k_BT} = \left[\frac{\int dX\, e^{-U(1;X)/k_BT}}{\int dX\, e^{-U(0;X)/k_BT}}\right] = \langle e^{-\Delta U/k_BT}\rangle_{(0)} \quad (5)$$

where $\Delta U = U(1;X) - U(0;X)$ and the average is performed relative to the energy $U(0)$, over the conformational subspace relevant for conformation a or b. Alternatively, one can use thermodynamic integration (TI):[34-36]

$$\Delta G_{i\rightarrow j} = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda}\right\rangle_{(\lambda)} \quad (6)$$

where the subscript on the brackets indicates that the average is calculated for a particular, but varying, value of λ.

Equation 4 forms the basis for the standard formulation used to calculate free energy differences in thermodynamical cycles involving alchemical transformations in biomolecular systems.[18,20,21] Although formally correct, there are important problems with this formulation because the energy $U(\lambda,X)$ usually is constructed such that all internal energy terms (bonds, angles, and dihedrals) are "turned off" at the end points of the alchemical transformation (when $\lambda = 0$ or $\lambda = 1$). The free energies at the end points therefore are defined with respect to a reference system consisting of the desired residue (i or j) plus an ideal gas of free noninteracting particles formed by the other residue (j or i). To clarify the origin of the problem, we write explicitly all the contributions to the potential energy $U(\lambda)$ in eq 2

$$\begin{aligned} U(\lambda) = {} & [U_{rr}^{bonds} + U_{rr}^{angles} + U_{rr}^{dihed} + U_{rr}^{nbond}] + \\ & (1-\lambda)[U_{ii}^{bonds} + U_{ii}^{angles} + U_{ii}^{dihed} + U_{ii}^{nbond} + U_{ir}^{bonds} + \\ & U_{ir}^{angles} + U_{ir}^{dihed} + U_{ir}^{nbond}] + \lambda[U_{jj}^{bonds} + U_{jj}^{angles} + U_{jj}^{dihed} + \\ & U_{jj}^{nbond} + U_{jr}^{bonds} + U_{jr}^{angles} + U_{jr}^{dihed} + U_{jr}^{nbond}] \end{aligned} \quad (7)$$

The calculated free energy includes a constant energy offset, which is unbounded, being proportional to the logarithm of the volume V of the simulation system that is available for the ideal gas particles: $\int dX_i = V^{n_i}$ or $\int dX_j = V^{n_j}$. In practical applications, the expression for $U(\lambda)$, as defined by eq 4, therefore may lead to large statistical errors and convergence problems. A better choice is to keep the internal energy terms invariant with respect to the thermodynamic coupling parameter

$$\begin{aligned} U^*(\lambda) = {} & [U_{rr}^{bonds} + U_{rr}^{angles} + U_{rr}^{dihed} + U_{rr}^{nbond}] + [U_{ii}^{bonds} + \\ & U_{ii}^{angles} + U_{ii}^{dihed} + (1-\lambda)U_{ir}^{nbond} + U_{ir}^{bonds} + U_{ir}^{angles} + \\ & U_{ir}^{dihed} + (1-\lambda)U_{ir}^{nbond}] + [U_{jj}^{bonds} + U_{jj}^{angles} + U_{jj}^{dihed} + \\ & \lambda U_{jj}^{nbond} + U_{jr}^{bonds} + U_{jr}^{angles} + U_{jr}^{dihed} + \lambda U_{jr}^{nbond}] \end{aligned} \quad (8)$$

In this case, the potential energy at the two end points corresponds to that of a reference system consisting of the desired residue in the presence of a fully bonded residue with no nonbonded interactions (as opposed to an ideal gas of free particles). In contrast to $U(\lambda)$ in eq 7, however, $U^*(\lambda)$ in eq 8 does not correspond to the potential energy of a true physical system. Nonetheless, the expression for the total free energy difference, $\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b}$, can be determined using $U^*(\lambda)$ in a manner that is completely equivalent to the estimate obtained using eq 7.

$$\exp(-\Delta\Delta G_{i\rightarrow j}^{a\rightarrow b}/k_BT) = \left[\frac{\int_b dX\, e^{-U^*(1;X)/k_BT}}{\int_a dX\, e^{-U^*(1;X)/k_BT}}\right]\left[\frac{\int_b dX\, e^{-U^*(0;X)/k_BT}}{\int_a dX\, e^{-U^*(0;X)/k_BT}}\right]^{-1} \quad (9)$$

This is possible because the bias that is introduced by keeping the internal energy terms will cancel out exactly, as long as the influence of the dummy atoms in the amino acids i and j can be factored out from the multidimensional integrals. When that is the case

$$\int_a dX_j\, e^{-U^*(0;X_r,X_i,X_j)/k_BT} = \int_b dX_j\, e^{-U^*(0;X_r,X_i,X_j)/k_BT} \equiv e^{-U_i(0;X_r,X_i)/k_BT} C_j \quad (10)$$

and

$$\int_a dX_i\, e^{-U^*(1;X_r,X_i,X_j)/k_BT} = \int_b dX_i\, e^{-U^*(1;X_r,X_i,X_j)/k_BT} \equiv e^{-U_j(0;X_r,X_j)/k_BT} C_i \quad (11)$$

where C_i and C_j are constants related to the partition functions

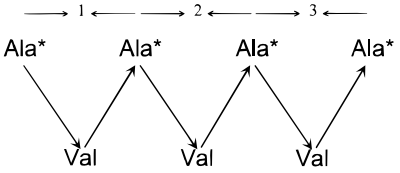

Figure 5. Three transformation cycles. The system is energy minimized at the beginning of every transformation cycle, which is denoted by *.

for the "invisible" noninteracting atoms of the residues *i* (when $\lambda = 1$) and *j* (when $\lambda = 0$), respectively. A closed-form analytical expression for the constants C_i and C_j is not required for eqs 10 and 11 to hold; but a critical issue in these calculations is how to treat the two end points, as they may lead to numerical divergence.[12]

It is possible to choose the coupling of the dummy atoms to the real atoms in such a way that it cancels out exactly from the calculated free energy differences. With such a choice, the resulting free energy difference is not really affected by the coupling to the noninteracting dummy atoms in the mutated residues at the end point. This is possible if, and only if, the statistical distribution function of the remaining real atoms with the noninteracting but covalently bonded dummy atoms is

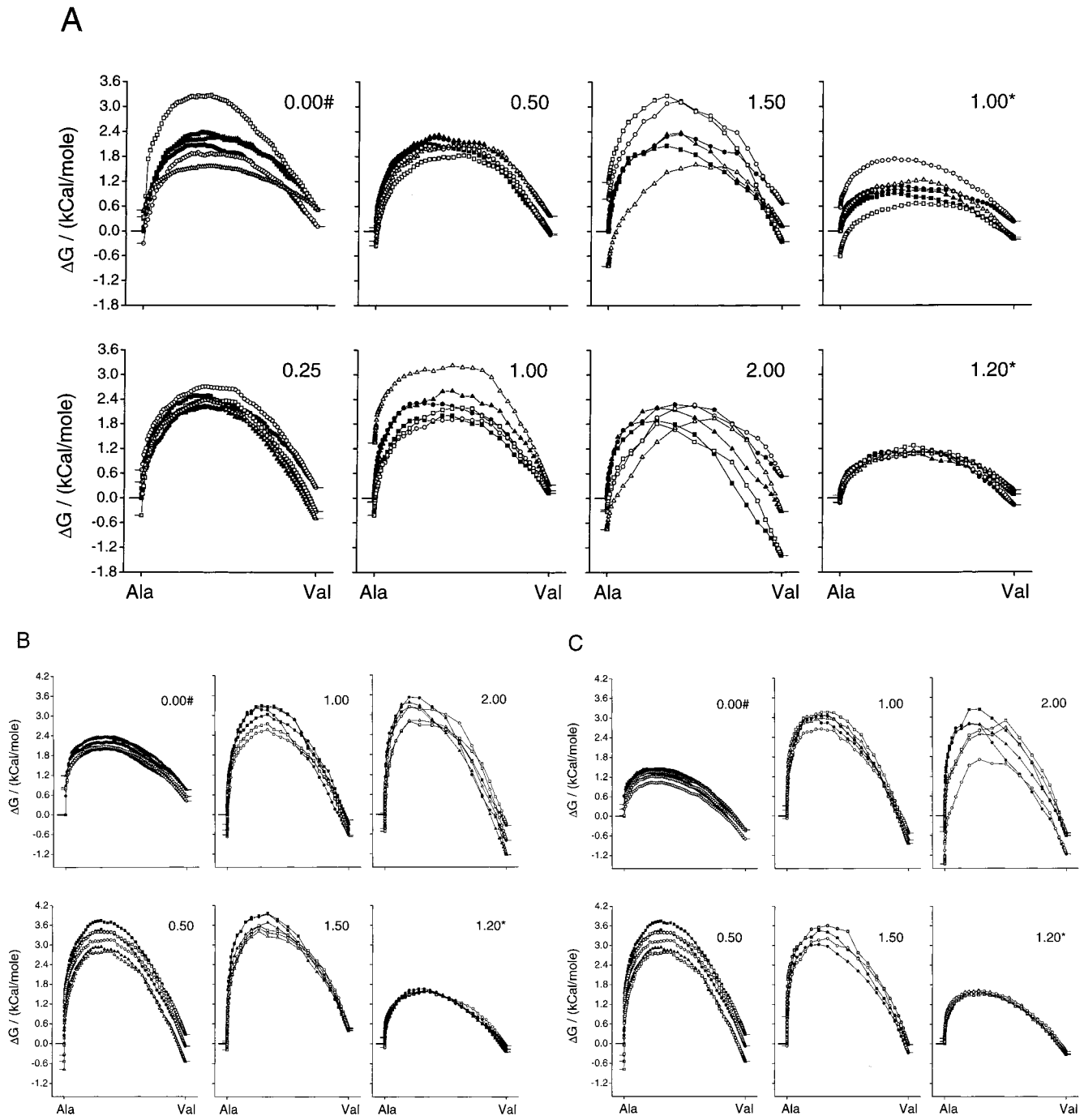

Figure 6. Energy profiles ($G(\lambda)$ vs λ) for Ala ↔ Val transformations in (A) f-*Ala*-eam, (B) f-$(Ala)_3$-Ala-*Ala*-Ala-$(Ala)_3$-eam, and (C) f-$(Ala)_3$-Val-*Ala*-Val-$(Ala)_3$-eam obtained using different λ progression curves. The numbers at the right-hand corner of each graph denote the "*n*" value used to generate the λ progression. The three transformation cycles are circles (first), squares (second), and triangles (third), respectively. The filled and open symbols denote the forward and backward runs. The starting points for the three cycles are marked by long bold dashes and the end points of the energy curves are marked by small dashes.

TABLE 1: Effect of the Shape of the λ Progression[a]

n	$\overline{\Delta G}_{Ala\rightarrow Val}$	sd (±)	hysteresis	sd (±)
A. Results of f-(Ala)-eam Simulations				
0.00#	+0.276	0.050	0.389	0.135
0.25	−0.308	0.342	0.505	0.172
0.50	+0.161	0.310	0.161	0.114
1.00	+0.073	0.364	0.358	0.139
1.50	−0.011	0.739	0.938	0.219
2.00	−0.172	0.968	0.451	0.265
1.00*	−0.048	0.176	0.397	0.335
1.20*	+0.053	0.178	0.097	0.028
B. Results of f-$(Ala)_3$-Ala-Ala-Ala-$(Ala)_3$-eam Simulations				
0.00#	+0.097	0.537	0.956	0.228
0.50	+0.164	0.447	0.545	0.217
1.00	−0.213	0.352	0.493	0.279
1.50	+0.457	0.073	0.094	0.083
2.00	−0.643	0.487	0.327	0.252
1.20*	−0.201	0.150	0.169	0.050
C. Results of f-$(Ala)_3$-Val-Ala-Val-$(Ala)_3$-eam Simulations				
0.00#	−0.635	0.169	0.239	0.167
0.50	+0.164	0.447	0.545	0.217
1.00	−0.770	0.173	0.187	0.116
1.50	−0.368	0.378	0.566	0.432
2.00	−0.390	0.507	0.765	0.613
1.20*	−0.350	0.084	0.134	0.056

[a] Energies in kcal/mol. Results based on three cycles of forward and backward simulations.

TABLE 2: Effect of Equilibration Time[a]

	$\overline{\Delta G}_{Ala\rightarrow Val}$	sd (±)	hysteresis	sd (±)
A. Results of f-(Ala)-eam simulations				
80 windows				
2 ps	−0.145	0.142	0.306	0.156
4 ps	−0.233	0.074	0.095	0.085
6 ps	−0.148	0.135	0.401	0.186
8 ps	+0.030	0.254	0.257	0.131
B. Results of f-$(Ala)_3$-Ala-Ala-Ala-$(Ala)_3$-eam simulations				
80 windows				
2 ps	−0.134	0.120	0.243	0.207
4 ps	−0.174	0.155	0.269	0.106
6 ps	−0.060	0.125	0.368	0.073
8 ps	−0.045	0.101	0.222	0.193
C. Results of f-$(Ala)_3$-Val-Ala-Val-$(Ala)_3$-eam simulations				
80 windows				
2 ps	−0.263	0.094	0.143	0.176
4 ps	−0.285	0.062	0.144	0.111
6 ps	−0.618	0.278	0.046	0.043
8 ps	−0.423	0.105	0.210	0.063

[a] Energies in kcal/mol. Results based on three cycles of forward and backward simulations.

equivalent to that of the true physical system at the end points. Even so, however, the free energy differences calculated on one branch of a thermodynamic cycle with the $U(\lambda)$ energy are not equal to the free energy differences calculated with the $U^*(\lambda)$ energy function. It is the difference over a closed loop of the cycle that is invariant. If the free energy difference using $U(\lambda)$ energy function is ΔG and the free energy difference using $U^*(\lambda)$ energy function is ΔG^*, the invariance actually means that $\Delta\Delta G = \Delta\Delta G^*$.

The equivalence of the statistical distribution functions, and of eqs 10 and 11, holds as long as the coupling between the dummy atoms and the rest of the system satisfies certain conditions. First, there can be only one bond between the dummy atoms of a mutated residue and the real atoms in the rest of the system, because multiple bonds would add spurious coupling between the real atoms. Second, to avoid spurious coupling between the dummy atoms and the rest of the system, there cannot be multiple bond angles and dihedral torsion angles between the dummy atoms of a transformed residue and more than two real atoms in the rest of the system. Figure 2 shows examples of correct connectivities between the dummy, invisible atoms (black circles) and the real atoms (gray circles), which preserve the equivalence of the distribution functions of the real atoms, as well as examples of incorrect connectivities that do not preserve this equivalence.

Figure 2A has the correct construction for the transformed residue if only one of the two dihedrals [α', ϵ, γ, β] and [α', ϵ, γ, δ] is included in the calculation. The presence of a single dihedral torsion term between the dummy atoms (black) and real atoms (gray) does not bias the intramolecular distribution function of the real atoms. If both dihedrals were included in the calculation, it would introduce a spurious coupling on the real atoms. In Figure 2B, the connectivity between α and β' is inadmissible because β' (black) introduces spurious coupling between α and ϵ (gray). The distribution function of the real atoms at the end point would not be perturbed only if either α' or β' is not bonded. In Figure 2C, inclusion of both the angles (β, γ, α') and (δ, γ, α') or the two dihedrals (δ, γ, α', β') and (β, γ, α', β') will introduce spurious coupling between the real atoms β, γ, and δ, even when the nonbonded interactions of the dummy atoms in the mutated residue (black) are turned off. If only one bond angle and one dihedral angle is included, it will preserve the equivalence of the distribution function of the real atoms at the end points. It is essential that such spurious couplings between the noninteracting dummy atoms and the real atoms are not included into the expression for U^* if one is to preserve the equivalence of the distribution function of the real atoms at the end points.

Computational Protocols

f-Ala-eam, f-$(Ala)_3$-Ala-Ala-Ala-$(Ala)_3$-eam and f-$(Ala)_3$-Val-Ala-Val-$(Ala)_3$-eam were built using the sequence builder of quanta96 (Molecular Simulations Inc.) and were minimized energetically until the norm of the gradient was less than 0.001 kcal/mol/Å. The energy-minimized peptides were used for the MD/FES alchemical calculations. The simulations were performed in vacuum using the PERT module of the program CHARMM[36] academic version 24b2 using the PARAM23 all-atom force field.[37] Langevin dynamics was used and the temperature of the external bath was kept at 300 K. A time step of 2 fs was used. Electrostatic interactions were calculated on groupwise basis with effectively no truncation (the nonbonded cutoff distance was set to 65 Å). The calculations were done on a INDIGO-II SGI workstation with a 200 MHz processor. In single-user mode, the CPU times for 10 ps of simulation were 40 s for f-Ala-eam and 5.5 min for the α-helices.

To provide for a smooth mutation from Ala to Val and back, a "hybrid alanine" and a "hybrid valine" were included as PATCH residues in the CHARMM topology file. A summary of the PATCH is given in Table 5 in the Appendix. The two hybrid amino acids are related as follows: the hybrid alanine is an Ala with an "invisible" Val and the hybrid valine is a Val with an "invisible" Ala (Figure 3).

Invisibility was achieved by setting the radii and charges of the noninteracting side chain atoms to zero. It is thus possible to "turn on" or "turn off" the Ala or the Val as the coupling parameter λ varies between 0 (Ala) and 1 (Val). The advantage of this construction is that the internal energy terms describing the valence force field (bond lengths, bond angles, and torsion angles) of the residues are kept during the alchemical transfor-

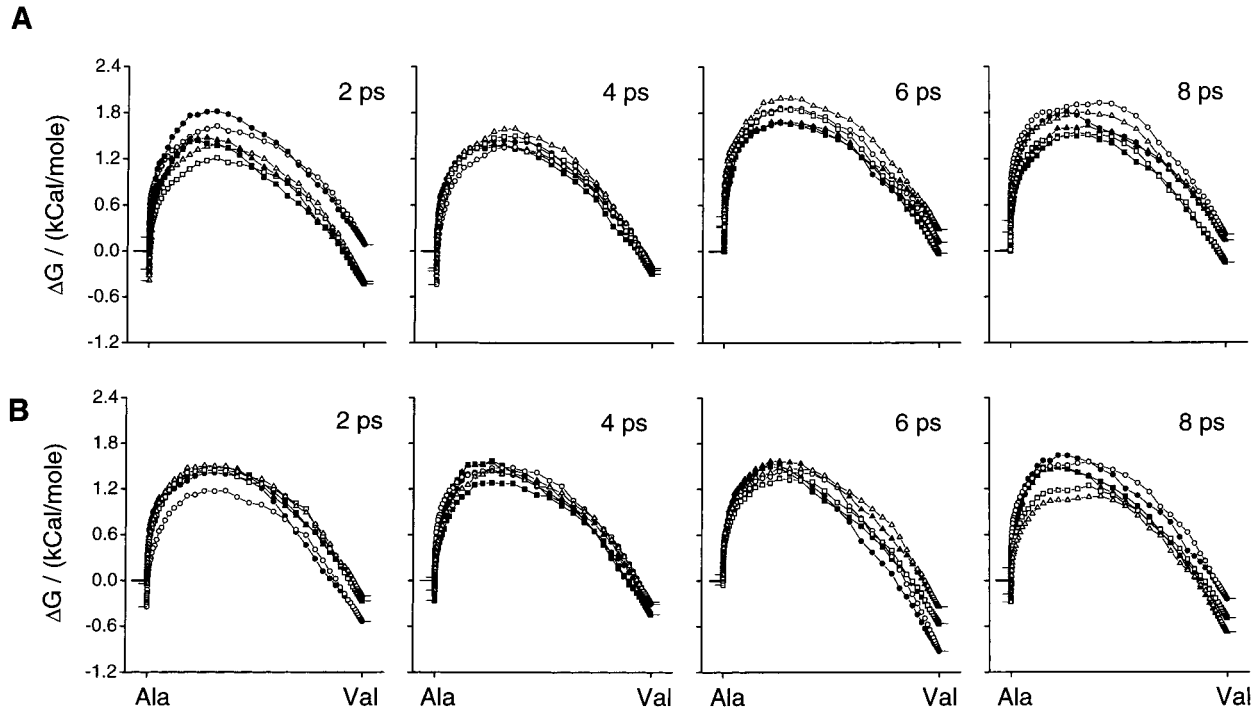

Figure 7. Energy profiles for Ala ↔ Val transformations in (A) f-$(Ala)_3$-Ala-*Ala*-Ala-$(Ala)_3$-eam and (B) f-$(Ala)_3$-Val-*Ala*-Val-$(Ala)_3$-eam using different equilibration times, which are denoted by the number in the right-hand corner of the four panels (symbols as in Figure 6). The production run is 8 ps in all simulations.

TABLE 3: Effect of Number of Windows (and Simulation Time)[a]

no. of windows	$\overline{\Delta G}_{Ala\rightarrow Val}$	sd (±)	hysteresis	sd (±)
A. Results of f-(Ala)-eam Simulations				
100	+0.254	0.046	0.341	0.132
200	−0.011	0.056	0.274	0.116
300	−0.142	0.085	0.119	0.071
400	−0.161	0.050	0.060	0.030
B. Results of f-$(Ala)_3$-Ala-Ala-Ala-$(Ala)_3$-eam Simulations				
100	−0.068	0.085	0.137	0.104
200	−0.174	0.067	0.171	0.090
300	−0.201	0.057	0.335	0.107
400	−0.391	0.163	0.379	0.099
200*	−0.223	0.144	0.176	0.072
200**	−0.223	0.272	0.113	0.076
C. Results of f-$(Ala)_3$-Val-Ala-Val-$(Ala)_3$-eam Simulations				
100	−0.452	0.068	0.234	0.125
200	−0.183	0.033	0.374	0.033
300	−0.145	0.007	0.187	0.065
400	−0.273	0.007	0.093	0.072
200*	−0.072	0.064	0.041	0.028
200**	−0.213	0.075	0.135	0.140

[a] Energies in kcal/mol. Results based on three cycles of forward and backward simulations.

mation. That is, only the noncovalent terms change as Ala "swells" into a Val or Val "shrinks" to an Ala. The valence force field of Ala is not erased as it is transformed to Val, only the nonbonded energies are erased; as described in the "Theory" section, and the turn off and turn on procedure provides for a gradual increase or decrease of the electrostatic and nonbonded interactions of the swelling and shrinking atoms.

It is generally believed that a nonlinear λ progression, with small window lengths near $\lambda = 0$ and $\lambda = 1$, is advantageous when doing alchemical calculations.[12,38] A sigmoid curve provides these general features.

To parametrize the shape of the λ progression, we used the following procedure to generate the progression curves (Figure 4A):

$$\lambda_i = \frac{a_i - a_1}{a_k - a_1} \tag{12}$$

where i varies from 1 to k, such that $k - 1$ is the number of windows in the calculation, and

$$a_i = \frac{(e^{x_i})^n}{1 + (e^{x_i})^n} \tag{13}$$

where x_i varies between -1 to $+1$ with the exponent n defining the shape of the λ progression.

Two strategies were used to calculate the x increments: first, a constant increment with

$$\Delta x_i = \frac{2}{k-1}$$

and second, a variable increment with

$$\Delta x_i = \frac{-\log\left\{\frac{(2+s_i)}{(2-s_i)}\right\}}{\log\left\{\frac{0.5}{3.5}\right\}}$$

where $-1.5 \le s_i \le 1.5$. Figure 4B shows the different ways in which x_i progresses from -1 to $+1$. In λ progressions with variable increment, the sampling points are spaced more equally over the central portion of the trajectory, as compared to the constant increment curve (Figure 4A).

In order to evaluate whether a given MD/FES procedure is robust, it is necessary to do redundant calculations using a range of parameter choices. In order to evaluate the reversibility and variability of the $\Delta G_{i\rightarrow j}$ estimates, all transformation cycles

(forward + backward) were performed three times. The final structure of the forward simulation was used as the starting structure for the backward simulation. The structure obtained at the end of each transformation cycle was energy minimized before the beginning of the next cycle (Figure 5). The free energy for the transformation and the hysteresis are determined as

$$\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}} = (\Delta G_{\text{Ala}\rightarrow\text{Val}} - \Delta G_{\text{Val}\rightarrow\text{Ala}})/2 \qquad (14)$$

$$\text{hysteresis} = |\Delta G_{\text{Ala}\rightarrow\text{Val}} + \Delta G_{\text{Ala}\rightarrow\text{Val}}| \qquad (15)$$

Results and Discussion

In the present study, many of the problems concerning the end points are overcome by building hybrid residues for the two end states. The amino acid residues are modified in such a way (Figure 3) that during the transformation process the disappearing atoms are mutated to dummy atoms, as in the dual topology method; however, only the noncovalent forces are changed during the mutation, as in the single topology method. Also, the end states: "hybrid Ala" and "hybrid Val" are nonphysical residues, but the free energies could be determined at both end states, $\lambda = 0$ and $\lambda = 1$. Although the calculated energies do not correspond to the energies of the "real" end point molecules (see above), the energy difference is physically meaningful. A similar approach, which uses dual topology in single topology framework, has been recently tested.[19,23]

The effects of changing the three identified parameters (the shape of the λ progression, the simulation time per window, and the number of sampling windows) on $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and hysteresis were examined first in f-Ala-eam simulations. The peptide f-Ala-eam was used as the test system for two reasons: first, f-Ala-eam is a highly flexible molecule and obtaining a reversible simulation for a peptide with no rigid secondary structure should serve as a good test for the robustness of the method; and second, f-Ala-eam is small such that the simulations are less computer intensive which makes it possible to explore a wide variety of parameter combinations. The effects of the shape of the λ progression were examined using the eight different curves (Figure 4A) with 80 windows. The simulations were done for 12 ps (4 ps of equilibration and 8 ps of production), and the energy trajectories for the three consecutive transformation cycles in each set of simulations are plotted in Figure 6A (see also ref 15) where the aggregate energy at the end of each window is plotted as a function of the window number. This plot provides information about not only the end points but the whole transformation cycle and their reversibility, which becomes helpful when evaluating the various transformation strategies. Because the λ windows are narrow, the energies obtained from thermodynamic perturbation (TP) and thermodynamic integration (TI) are indistinguishable; only the TP energies are plotted in the figure. For the backward runs, the first point at ($\lambda = 1$) is made to coincide with the last point of the forward run; hence, the end point difference at ($\lambda = 0$) gives a direct readout of the hysteresis in the transformation cycle. The envelope defined by the trajectories provides a visual measure of the statistical variability of the free energies. The shape of the energy trajectories varies with the particular shapes of the sigmoidal λ progression not only where the windows are narrow, close to the two end points of $\lambda = 0$ and $\lambda = 1$, but also in the steep phase of the sigmoidal curve. The plots also illustrate well the difficulties associated with determination of the energies close to the end points, as the slope of ΔG vs λ relation is steepest in the vicinity of $\lambda = 0$ and $\lambda = 1$. For example, the difference between $G(0)$ and $G(0.05)$ can be >1 kcal/mol. This large variation in G illustrates the numerical difficulties associated with the appearance/disappearance of the "real" atoms in the calculation (see also ref 12).

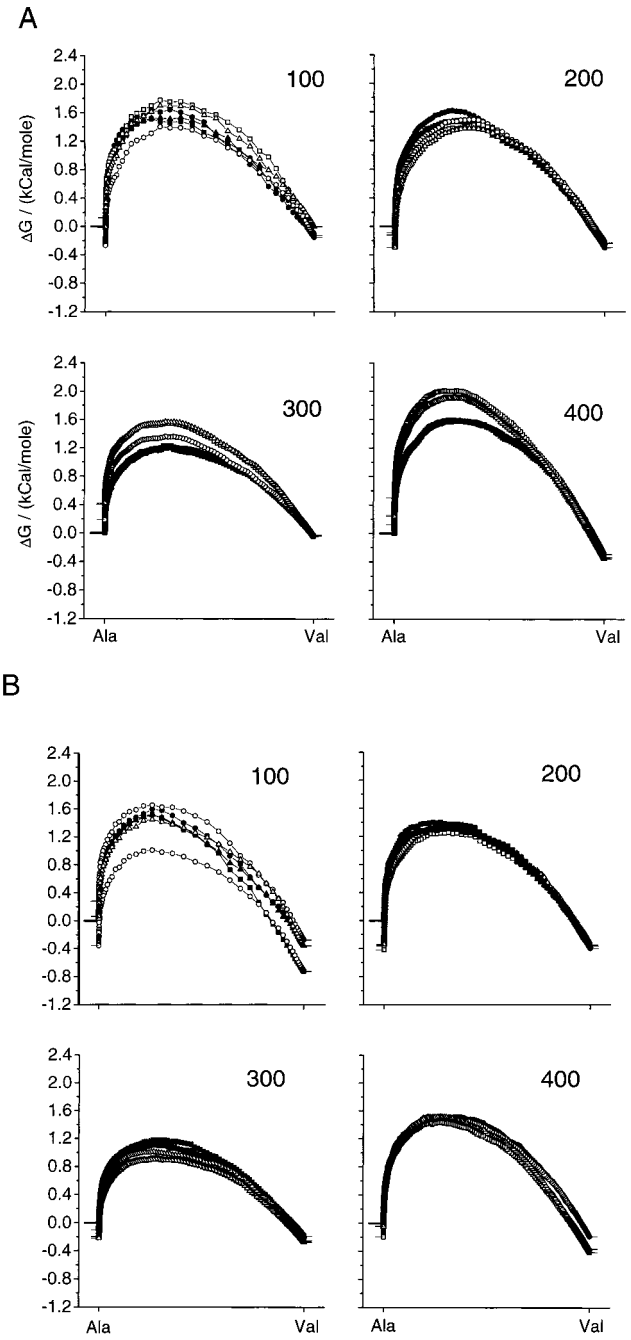

Figure 8. Energy profiles of (A) f-$(Ala)_3$-Ala-*Ala*-Ala-$(Ala)_3$-eam and (B) f-$(Ala)_3$-Val-*Ala*-Val-$(Ala)_3$-eam simulations using different number of windows, which are denoted by the number in the right-hand corner of the four panels. The equilibration time is 4 ps and the production time is 8 ps in all simulations (symbols as in Figure 6).

On the basis of visual inspection of Figure 6A, the λ progressions with variable increments (1.20*) have the least spread in the energy trajectories. Quantitative measures are provided in Table 1A, which summarizes the results for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and hysteresis in the f-Ala-eam simulations.

In Table 1A, the sd for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ for the λ progressions with variable step size, $n = 1.00$* and 1.20*, are less than the values obtained with all the fixed interval sigmoidal curves. The linear progression, $n = 0.00$#, has an sd $\leq$ 0.10 kcal/mol; but

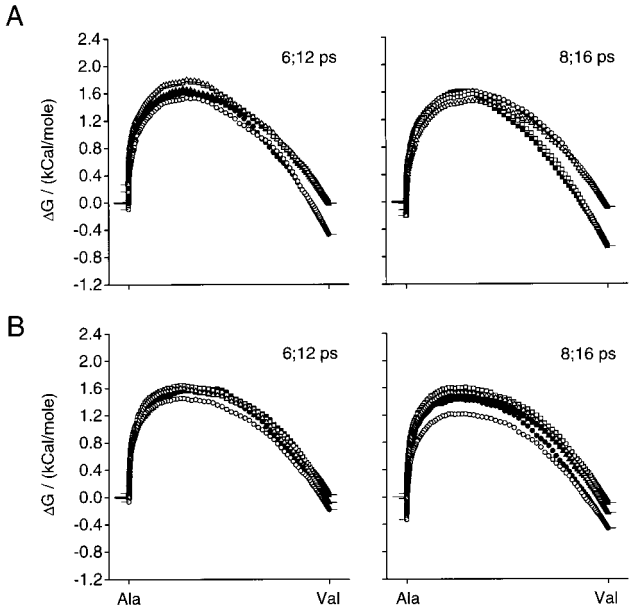

Figure 9. Energy profiles of (A) f-(Ala)$_3$-Ala-*Ala*-Ala-(Ala)$_3$-eam and (B) f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam simulations with 200 windows. The first number in the right-hand corner of the four panels denotes the equilibration time and the second number denotes the production run time in all (symbols as in Figure 6).

the hysteresis is large, >0.35 kcal/mol. The λ progression denoted as $n = 1.20$* produces the lowest hysteresis and a low sd, compared to other progression curves. This can be also seen from Figure 6A, where the deviations in the end point energies at $\lambda = 0$ and $\lambda = 1$ are minimal when using the $n = 1.20$* λ progression and the spread of the energy trajectories is small. Overall, when examining the results in Figure 6A and Table 1A, $n = 1.20$* gives the most satisfactory results in terms of both reversibility and statistical variability.

The patterns of variation in the sd for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and the hysteresis, however, are such that it is difficult to conclude whether the $n = 1.20$* progression indeed is the optimal one. To examine the generality of the conclusion obtained from the f-Ala-eam studies, simulations were performed with the right-handed α-helices f-(Ala)$_3$-Ala-Ala-Ala-(Ala)$_3$-eam and f-(Ala)$_3$-Val-Ala-Val-(Ala)$_3$-eam with $n = 0.00$#, 0.50, 1.00, 1.50, 2.00, and 1.20*. The energy trajectories are plotted in Figure 6, B and C, and summarized in Table 1, B and C.

Again, the initial energy of the backward simulation (at $\lambda = 1$) is set equal to the final energy of the forward simulation so that a direct visual readout of hysteresis can be obtained from the start and end point energies of the forward and backward runs, respectively. On the basis of visual inspection of Figure 6, B and C, the λ progression with variable increment (1.20*) again displays the least spread in energy trajectories compared to other curves. From Table 1, B and C, the sd for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and hysteresis are least for both α-helices when the $n = 1.20$* λ progression is used. The results confirm that the $n = 1.20$* progression provides thermodynamically reversible energy estimates, which display little variability.

The second parameter to be investigated was the equilibration period required for the simulations. Again, the f-Ala-eam peptide was used for the initial simulations, which were done using the $n = 1.2$* progression curve. The simulations tested had 2, 4, 6, or 8 ps of equilibration periods, with the production run held constant at 8 ps per window. Calculations were done using 80 windows; the results are summarized in Table 2A. The 4 ps equilibration time gives the best estimates with the least sd of

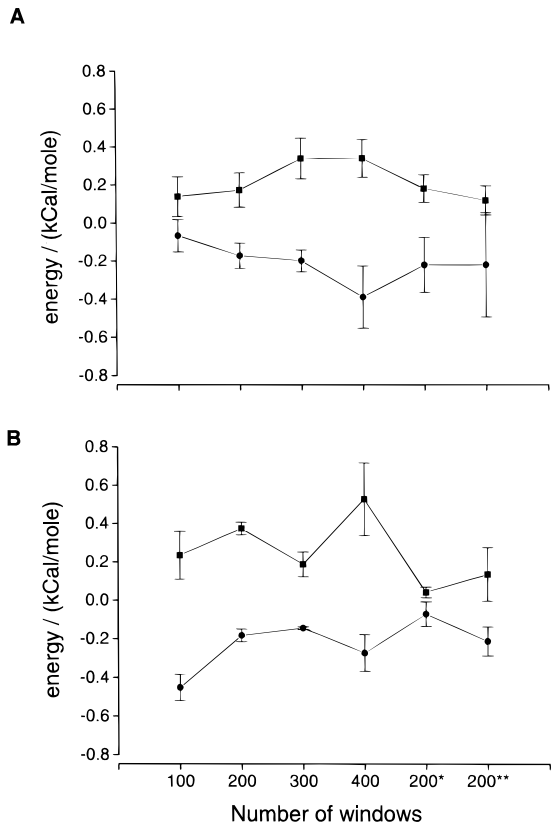

Figure 10. Plot of $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ (closed circles) and mean hysteresis (closed squares) with their sd (±) for (A) f-(Ala)$_3$-Ala-*Ala*-Ala-(Ala)$_3$-eam and (B) f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam simulations shown in Figures 8A and 9A and 8B and 9B, respectively. The equilibration time is 4 ps and the production time is 8 ps in all simulations (symbols as in Figure 6).

0.074 kcal/mol in $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and the least hysteresis of 0.095 kcal/mol; however, there is no obvious trend in the results. On the basis of these results, an equilibration time of 4 ps provides for satisfactory simulations.

To evaluate the generality of this result, we again did simulations on the two α-helices using 2, 4, 6, and 8 ps of equilibration time and a constant production run time of 8 ps. Figure 7 shows the energy profiles, and Table 2B,C summarizes the quantitative estimates of $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and hysteresis for the various simulations.

Figure 7 shows that the "spread" of trajectories is least for the 4 ps as compared to 2, 6, and 8 ps, again suggesting that a 4 ps equilibration period should allow for satisfactory ΔG estimates. But, from the sd for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and the hysteresis results in Table 2, B and C, we conclude that either of the two, 4 and 6 ps, equilibration times should be acceptable with no obvious advantage of one over the other. Given the narrow envelope of the energy trajectories, equilibration of 4 ps, however, was chosen for the further calculations.

The importance of the number of windows between $\lambda = 0$ and $\lambda = 1$ was investigated using f-Ala-eam and 100, 200, 300, and 400 windows with the $n = 1.2$* progression curve and an equilibration period of 4 ps and a production period of 8 ps per window. The results are summarized in Table 3A.

The sd for $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ for 100, 200, 300, and 400 windows are ≤0.1 kcal/mol. As the number of windows increases, the sd of hysteresis decreases and the hysteresis decreases as the number of windows increases, suggesting that long simulations with narrow windows are advantageous.

TABLE 4: RMS Deviations for the Backbone Atoms (N, C$^{\square}$, C, O) in the Helix Simulations

	f-(Ala)$_3$-Ala-Ala-Ala-(Ala)$_3$-eam		*f*-(Ala)$_3$-Val-Ala-Val-(Ala)$_3$-eam	
path description	1 ↔ 2	1 ↔ 3	1 ↔ 2	1 ↔ 3
100 windows	0.149	0.123	0.014	0.802
	0.021	0.008	0.027	0.106
200 windows	0.242	0.242	0.455	0.401
	0.143	0.521	0.019	0.023
300 windows	0.488	0.498	0.673	0.468
	0.026	0.102	0.006	0.008
400 windows	0.145	0.145	0.433	0.333
	0.044	0.314	0.368	0.388
200* windows	0.689	0.306	0.187	0.004
	0.428	0.307	0.188	0.005
200** windows	0.303	0.652	0.388	0.279
	0.310	0.651	0.355	0.230

[a] The structures after the second and third transformation cycles are compared with the structure after the first cycle. For each set of simulations, the first and second row denote the forward and backward results, respectively. Effect of number of windows (and simulation time).

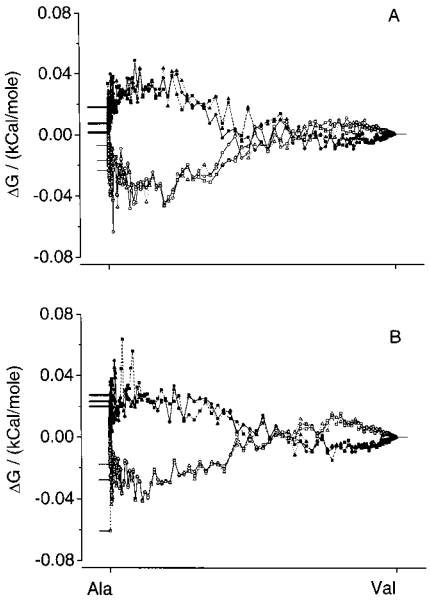

Figure 11. Nonbonded energy component obtained from TI calculations for (A) f-(Ala)$_3$-Ala-*Ala*-Ala-(Ala)$_3$-eam and (B) f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam simulations with 200 windows where the equilibration time is 4 ps and the production time is 8 ps. Symbols as in Figure 6. For clarity's sake, the first, second, and third free energy cycles are denoted as solid, dashed, and dotted lines.

Simulations were again performed with the two α-helices. Figure 8A,B shows the resulting energy trajectories for the f-(Ala)$_3$-Ala-Ala-Ala-(Ala)$_3$-eam and f-(Ala)$_3$-Val-Ala-Val-(Ala)$_3$-eam helices, respectively. $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and hysteresis results from the two helix simulations are summarized in Table 3B,C.

Simulations with 200 and 300 windows provide comparable $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ for both α-helices (Figure 9; Table 3B,C). In the f-(Ala)$_3$-Ala-Ala-Ala-(Ala)$_3$-eam simulations, the hysteresis increases as the number of windows increases beyond 200, and the simulation with 400 windows has the maximum hysteresis. In the f-(Ala)$_3$-Val-Ala-Val-(Ala)$_3$-eam simulations, however, the hysteresis decreases as the number of windows increases. In both the f-(Ala)$_3$-Ala-Ala-Ala-(Ala)$_3$-eam and f-(Ala)$_3$-Val-Ala-Val-(Ala)$_3$-eam simulations, there is no significant change in $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$, as the number of windows is increased from 200 to 300; but there is a rather large change in $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$, (0.15 to 0.24 kcal/mol) when the number of windows is increased from 300 to 400 and, somewhat surprisingly, there is no improvement in the hysteresis. Overall, there is significant "improvement"

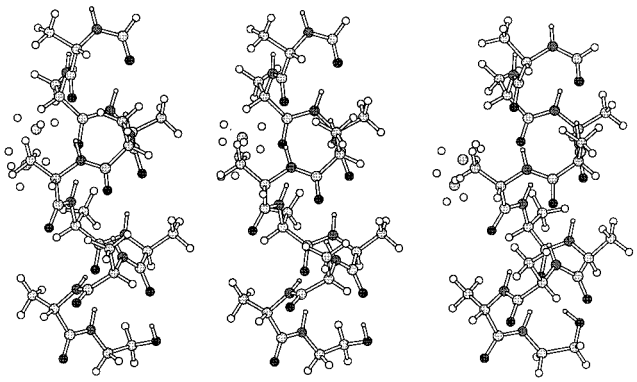

Figure 12. f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam structures obtained from simulations with 200 windows. The final structures after the first, second, and thirrd transformation cycle are plotted in the left, middle, and right structures. The dummy atoms of the hybrid residue ("hybrid Ala") are those that are not covalently linked to the helix.

in $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ for all three (f-*Ala*-eam; f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam; and f-(Ala)$_3$-Val-*Ala*-Val-(Ala)$_3$-eam) simulations, as the number of windows increases from 100 to 200 but there is no obvious advantage in going beyond 200 windows (Table 3A−C).

To investigate why simulations with >200 windows do not provide better results than the ones with 200 windows, simulations were done with 200 windows but increased equilibration and production run time using both helices. First, 6 ps of equilibration and 12 ps of production time per window (total time = 3.6 ns); and second, 8 ps of equilibration and 16 ps of production time per window (total time = 4.8 ns). Since the total time periods for these simulations are equal to the simulations with 300 and 400 windows, they are denoted as 200* and 200**, respectively. The energy trajectories are shown in Figure 9.

Comparing the spread of trajectories in Figures 8A,B and 9, the envelope is less dispersed in the 300 and 400 as compared to the 200* and 200** window simulations, in both the helices. Figure 10 summarizes the results on $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ and the hysteresis.

On the basis of these results, we conclude that simulations with 200 windows and (4 + 8) ps per window provide for thermodynamically reversible transformations with little variability.

To evaluate how the "dual topology within a single topology" framework works, we monitored the valence energies as a function of λ. Inspection of the output files (results not shown) shows that the valence energies vary between ±0.03 kcal/mol.

TABLE 5[a]

	hybrid alanine		hybrid valine	
atoms	type	charge	type	charge
CB	CT3	−0.27	CT1	−0.09
HB1	HA	+0.09	HA	+0.00 no LJ
HB2	HA	+0.09	HA	+0.00 no LJ
HB3	HA	+0.09	HA	+0.00 no LJ
HB	HA	+0.00 no LJ	HA	+0.09
CG1	CT3	+0.00 no LJ	CT3	−0.27
HG11	HA	+0.00 no LJ	HA	+0.09
HG12	HA	+0.00 no LJ	HA	+0.09
HG13	HA	+0.00 no LJ	HA	+0.09
CG2	CT3	+0.00 no LJ	CT3	−0.27
HG21	HA	+0.00 no LJ	HA	+0.09
HG22	HA	+0.00 no LJ	HA	+0.09
HG23	HA	+0.00 no LJ	HA	+0.09

angles		angles		angles	
hybrid alanine	hybrid valine	hybrid alanine	hybrid valine	hybrid alanine	hybrid valine
CA−CB−HB1	CA−CB−HB1	CA−CB−HB	CA−CB−HB	HB−CB−CG1	HB−CB−CG1
CA−CB−HB2	CA−CB−HB2	CB−CG1−HG11	CB−CG1−HG11	HB−CB−CG2	HB−CB−CG2
CA−CB−HB3	CA−CB−HB3	CB−CG1−HG12	CB−CG1−HG12	HG11−CG1−HG12	HG11−CG1−HG12
HB1−CB−HB2	HB1−CB−HB2	CB−CG1−HG13	CB−CG1−HG13	HG11−CG1−HG13	HG11−CG1−HG13
HB1−CB−HB3	HB1−CB−HB3	CG1−CB−CG2	CG1−CB−CG2	HG12−CG1−HG13	HG12−CG1−HG13
HB2−CB−HB3	HB2−CB−HB3	CB−CG2−HG21	CB−CG2−HG21	HG21−CG2−HG22	HG21−CG2−HG22
CA−CB−CG1	CA−CB−CG1	CB−CG2−HG22	CB−CG2−HG22	HG21−CG2−HG23	HG21−CG2−HG23
CA−CB−CG2	CA−CB−CG2	CB−CG2−HG23	CB−CG2−HG23	HG22−CG2−HG23	HG22−CG2−HG23

dihedrals		dihedrals		dihedrals	
hybrid alanine	hybrid valine	hybrid alanine	hybrid valine	hybrid alanine	hybrid valine
N−CA−CB−HB1	(N−CA−CB−HB1)[b]		HA−CA−CB−HB	HB−CB−CG1−HG11	HB−CB−CG1−HG11
N−CA−CB−HB2			HA−CA−CB−CG1	HB−CB−CG1−HG12	HB−CB−CG1−HG12
N−CA−CB−HB3			HA−CA−CB−CG2	HB−CB−CG1−HG13	HB−CB−CG1−HG13
C−CA−CB−HB1			HB−CB−CA−C	HB−CB−CG2−HG21	HB−CB−CG2−HG21
C−CA−CB−HB2			CG1−CB−CA−C	HB−CB−CG2−HG22	HB−CB−CG2−HG22
C−CA−CB−HB3			CG2−CB−CA−C	HB−CB−CG2−HG23	HB−CB−CG2−HG23
HA−CA−CB−HB1		CA−CB−CG1−HG11	CA−CB−CG1−HG11	CG1−CB−CG2−HG21	CG1−CB−CG2−HG21
HA−CA−CB−HB2		CA−CB−CG1−HG12	CA−CB−CG1−HG12	CG1−CB−CG2−HG22	CG1−CB−CG2−HG22
HA−CA−CB−HB3		CA−CB−CG1−HG13	CA−CB−CG1−HG13	CG1−CB−CG2−HG23	CG1−CB−CG2−HG23
	N−CA−CB−HB	CA−CB−CG2−HG21	CA−CB−CG2−HG21	HG11−CG1−CB−CG2	HG11−CG1−CB−CG2
(N−CA−CB−CG1)[b]	N−CA−CB−CG1	CA−CB−CG2−HG22	CA−CB−CG2−HG22	HG12−CG1−CB−CG2	HG12−CG1−CB−CG2
	N−CA−CB−CG2	CA−CB−CG2−HG23	CA−CB−CG2−HG23	HG13−CG1−CB−CG2	HG13−CG1−CB−CG2

[a] There are no nonbonded interactions between any of the atoms of the alanine (CB, HB1, HB2, HB3) and those of the valine (HB, CG1, HG11, HG12, HG13, CG2, HG21, HG22, HG23) in the two hybrid residues. [b] In principle, one dihedral angle belonging to the invisible part of the residue can be kept to act on the dummy atoms. For example, one dihedral from the alanine(left column) (e.g., N−CA−CB−HB1) could be kept for the hybrid valine (right), and one dihedral from the valine (right column) (e.g., N−CA−CB−CG1) could be kept for the hybrid alanine (left column). In the present simulations, no dihedrals were kept for the dummy atoms for the sake of simplicity.

This variation is attributed to including all residues (the residue that is transformed and the rest of the molecule) of the molecular system in the perturbation calculations. Earlier studies[15] found that van der Waals interactions vary significantly in MD/FES simulations, to the point where they may cause divergence. Figure 11 shows the van der Waals energy trajectories for the $n = 1.20$* profile (with 200 points) for the two α-helices. The energies are bounded and do not vary dramatically. The variation is less around the hybrid valine as compared to hybrid alanine, but the energies are constrained between ±0.06 kcal/mol. The variations in nonbonded energies (Figure 11, A and B) are small compared to the changes in the total energy in the three cycles of the simulations (Figure 8, A and B). This behavior is noteworthy, which also shows that the "dual topology within a single topology" framework works well for alchemical transformations.

Similar to the $\overline{\Delta G}_{\text{Ala}\rightarrow\text{Val}}$ energy plots, the three transformation cycles are plotted. The results are normalized differently, however. Nonbonded energy at the first point of the backward run ($\lambda = 1$) is made to coincide with the last point of the forward run; hence the end point difference at ($\lambda = 0$) will give the direct read-out in nonbonded energy changes in the Ala conformations within the three transformation cycles.

Finally, an important issue to be concerned about is the stability of the secondary structures. This was evaluated by calculations the root-mean-square deviations (rmsd) of the various helical structures. Table 4 summarizes the backbone rmsd between the structures after first and second transformation cycle and between the first and third transformation cycle for the two helices.

When the structures were visualized using quanta97, those that had rmsd > 0.5 Å showed backbone deviations only at the terminal formyl and ethanolamine residues, whereas the rest of the regions remained helical (Figure 12).

This suggests again that the transformation procedure is robust and retains the secondary structure during the course of the simulation.

Summary

We have shown that stable and meaningful ΔG averages can be performed with MD/FES at any value of the coupling

parameter λ, including the end points $\lambda = 0$ and $\lambda = 1$. The present approach is an intermediate between the single and dual topology techniques. Dummy atoms with no nonbonded interactions are present at the end point reference states; some of their covalent bonded interactions with the real atoms of the rest of the system are kept. The detailed shape of λ progression used in the alchemical transmutation allows for free energy estimates at $\lambda = 0$ and $\lambda = 1$. This approach, coupled with a nonlinear λ progression curve, leads to robust ΔG determinations, which avoid the singularities that tend to appear at the two end points.[12] The interpretation of the calculated ΔG's is clear and unambiguous within this framework. Although the influence of the covalent terms is nonzero on any single branch of a thermodynamic cycle, these effects cancel out exactly in the closed thermodynamic cycle. That is, although the ΔG from $U(\lambda)$ is not equal to the ΔG^* from $U^*(\lambda)$, the difference cancels out exactly and $\Delta\Delta G = \Delta\Delta G^*$. The present approach hence focuses mainly on the intrinsic problems associated with the treatment of dummy atoms in standard MD simulations. It also, provides insight into how to setup these dummy atoms in a hybrid residue, retaining all the bonded interactions in such a way that the bonded interactions with dummy atoms do not influence the final $\Delta\Delta G$. It should be noted that ΔG vs λ plots provide for a convenient examination of the results of alchemical calculations which may also prove useful in other settings.

We conclude that MD/FES methods are robust for estimating free energies differences. Notwithstanding the general stability of the estimates, some conditions must be met in order to obtain reliable free energy differences: first, an appropriate sigmoidal λ progression; second, an adequate equilibration period, where 4 ps appears to be sufficient; and third, simulation/transformation times ≥ 2.4 ns should be used. This latter conclusion is in concordance with that of Pearlman[22] who found that ~ 1 ns was necessary to achieve best converged simulation for the zero-sum model of ethane $\leftrightarrow$ ethane in water. It still raises the question, however, whether slower conformational changes will be important when comparing the computational results to experimental determinants of the free energy differences.

Finally, we have performed the MD/FES only in vacuum, though the real effects are measured in condensed-phase environment. This choice allowed us to focus on the main issues of MD/FES, which concern the treatment of the reference states and the covalent energy terms at the end points. Nonbonded interactions with the environment, such as a lipid bilayer, can be introduced using standard methods.

Acknowledgment. This work was supported by a NIH grant GM21342 (O.S.A.), a Norman and Rosita Winston Postdoctoral Fellowship (S.S.), and a grant from the Medical Research Council of Canada (B.R.).

Appendix

The modified residues hybrid alanine and hybrid valine are generated by including the force-field parameters listed in Table 5. The atoms that have no nonbonded or Lennard-Jones (LJ) interactions in the hybrid residues are indicated as "no LJ".

References and Notes

(1) Dalal, S.; Balasubramanian, S.; Regan, L. *Nature Struct. Biol.* **1997**, *4*, 548−552.
(2) Shobana, S.; Saberwal, G.; Andersen, O. S.; Greathouse, D. V.; Koeppe, R. E. *Biophys. J.* **1998**, *74*, A387.
(3) Durkin, J. T.; Providence, L. L.; Koeppe, R. E., II; Andersen, O. S. *Biophys. J.* **1992**, *62*, 145−159.
(4) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: London, 1987.
(5) Mezei, M.; Beveridge, D. L. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1−23.
(6) van Gunsteren, W. F. In *Computer Simulation in Biomolecular Systems. Theoretical and experimental applications;* van Gunsteren, W. F., Ed.; ESCOM Science Publishers B. V.: Amsterdam, 1989; pp 27−59.
(7) Yun, R. H.; Hermans, J. *Prot. Eng.* **1991**, *4*, 761−6.
(8) Wang, L.; O'Connell, T.; Tropsha, A.; Hermans, J. *J. Mol. Biol.* **1996**, *262*, 283−293.
(9) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300−313.
(10) Simonson, T. *Mol. Phys.* **1993**, *80*, 441.
(11) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(12) Beutler, T. C.; Mark. A. E.; Vanschaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(13) Pomes, R.; Eisenmesser, E.; Post, C. B.; Roux, B. *J. Chem. Phys.* **1999**, *111*, 3387−3395.
(14) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395−2417.
(15) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487−1493.
(16) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103−118.
(17) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 119−136.
(18) Pearlman, D. A. *J. Chem. Phys.* **1993**, *98*, 8946−8957.
(19) van Gunsteren, W. F.; Berendsen, H. J. C. GROningen MOlecular Simulation (GROMOS) Library Manual; Laboratory of Physical Chemistry, University of Groningen; Groningen 1987.
(20) Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1991**, *94*, 4532−4545.
(21) Prevost, M.; Wodak, S. J.; Tidor, B. K. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10880−4.
(22) Elofsson, A.; Nilsson, L. *Mol. Simul.* **1993**, *10*, 255−276.
(23) Eriksson, M. A. L.; Nilsson, L. *J. Mol. Biol.* **1995**, *253*, 453−472.
(24) Sun., Y. C.; Veenstra, D. L.; Kollman, P. A. *Prot. Eng.* **1996**, *9*, 273−281.
(25) Tidor, B.; Karplus, M. *Biochemistry* **1991**, *30*, 3217−3228.
(26) Axelson, P. H.; Li, D. *J. Comput. Chem.* **1998**, *19*, 1278−1283.
(27) Durkin, J. T.; Koeppe, R. E.; II; Andersen, O. S. *J. Mol. Biol.* **1990**, *211*, 221−234.
(28) Durkin, J. T.; Providence, L. L.; Koeppe, R. E.; II; Andersen, O. S. *J. Mol. Biol.* **1993**, *231*, 1102−1121.
(29) Andersen, O. S.; Saberwal, G.; Greathouse, D. V.; Koeppe, R. E., II *Ind. J. Biochem. Biophys.* **1996**, *33*, 331−342.
(30) Shobana, S. Unpublished work.
(31) Sarges, R.; Witkop, B. *J. Am. Chem. Soc.* **1965**, *87*, 2011−2019.
(32) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55−67.
(33) Jorgensen, W. L.; Ravimohan, C. J. *Chem. Phys.* **1985**, *83*, 3050−3054.
(34) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F. *Molecular Dynamics and Protein Structure*; Polycrystal Book Service: Chicago, IL, 1985.
(35) Mruzik, M. R.; Abraham, F. F.; Schreiber, D. E.; Pound, G. M. *J. Chem. Phys.* **1976**, *64*, 481−491.
(36) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. *J. Chem. Phys.* **1986**, *85*, 6720.
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(38) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuckzera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Workiewicz-Kuczera, J.; Yinb, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(39) Cross, A. J. *Chem. Phys. Lett.* **1986**, *128*, 198−202.