BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Relative Free Enthalpies for Point Mutations in Two Proteins with Highly Similar Sequences but Different Folds

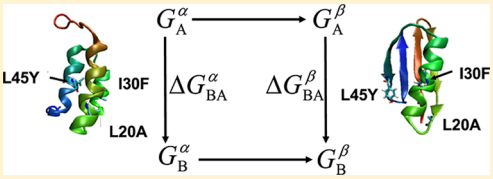
Niels Hansen,[†] Jane R. Allison,[‡] Florian H. Hodel,[†] and Wilfred F. van Gunsteren*[,†]

[†]Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, CH-8093 Zürich, Switzerland

[‡]Centre for Theoretical Chemistry and Physics, Institute of Natural and Mathematical Sciences, Massey University Albany, Auckland 0632, New Zealand

Supporting Information

ABSTRACT: Enveloping distribution sampling was used to calculate free-enthalpy changes associated with single amino acid mutations for a pair of proteins, G_A95 and G_B95, that show 95% sequence identity yet fold into topologically different structures. Of the L → A, I → F, and L → Y mutations at positions 20, 30, and 45, respectively, of the 56-residue sequence, the first and the last contribute the most to the free-enthalpy difference between the native and non-native sequence−structure combinations, in agreement with the experimental findings for this protein pair. The individual free-enthalpy changes are almost sequence-independent in the four-strand/one-helix structure, the stable form of G_B95, while in the three-helix bundle structure, the stable form of G_A95, an interplay between residues 20 and 45 is observed.



The three-dimensional structure a protein adopts in solution is a complex function of the amino acid sequence and the properties of the surrounding solvents,[1−6] making de novo structure prediction a very challenging problem. Although ongoing refinement of NMR structure prediction tools[7,8] will continuously enhance the automation of sequence-to-structure mapping and will push the size and complexity limit of these methods forward, failures in automatic fold recognition are reported occasionally,[9,10] showing that some fundamental principles underlying the way in which the protein sequence determines structure are not entirely understood.[11] The question of how much of the sequence is required to specify the final fold has recently been addressed experimentally by systematic modifications of two 56-residue proteins derived from the binding domains of *Streptococcus* protein G with increasing levels of sequence identity (SI). It was demonstrated that a single residue may act as a switch to change the conformation from a three-helix bundle (3α-structure) into a four-strand β-sheet with a single α-helix ($4\beta+\alpha$-structure).[12−15] Eighty-five percent of residues change their secondary structure, with only eight residues in the central α-helix plus two turn residues retaining the same conformation in both forms. Clearly, this system poses a challenge to automatic structure prediction methods. Two pairs of proteins in the series, G_A95 and G_B95, differing only in three amino acids were used as targets in the CASP8 structure prediction competition.[16] G_A95 has amino acids Leu, Ile, and Leu at positions 20, 30, and 45, respectively, and adopts the 3α-structure, while G_B95 has Ala, Phe, and Tyr and adopts the $4\beta+\alpha$-structure. Only 4 out of 159 participants recognized the two different folds.[10] In particular, methods that rely on comparative modeling approaches failed for this set of proteins, while structure determination methods based on chemical shifts were shown to be more successful.[17] Because of the nonlocal nature of the physical processes which direct the folding of proteins[18−20] and the strong dependence of the backbone conformation on solvent affinities of amino acid side chains,[21] atomistic simulations in explicit solvent are promising tools to study these phenomena.[22−24] However, the direct sampling of protein folding by molecular dynamics (MD) is in general not feasible, except for small proteins (fewer than 50 residues).[25] The long MD runs necessary to study folding directly require special purpose hardware that is not widely available.[26,27] Moreover, for a reliable estimate of the free energy, or rather free enthalpy[28] or Gibbs energy,[29] of folding, many folding−unfolding events have to be sampled, which is far from being feasible. Another issue in this context arises from the short time scales usually used in force field parametrization. A force field that performs well on short time scales is not necessarily adequate for long time simulations[30] or accurate enough to predict the correct global free-enthalpy minimum.[31]

In the present work, we make use of the fact that in a relatively short MD simulation (10−20 ns) of protein G, no unfolding is observed even if the non-native sequence−structure combinations are simulated. This allows the use of the thermodynamic cycle shown in Figure 1, in which the vertical branches denote a change in the amino acid sequence within a given tertiary structure.

In a previous study, we investigated the two proteins G_A95 and G_B95 by a combination of structural and energetic analyses.[32] However, unequivocal support for the experimental findings by means of the simulation data was not obtained. In

Received: March 4, 2013
Revised: June 25, 2013
Published: June 26, 2013

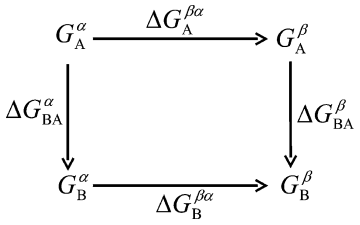


Figure 1. Thermodynamic cycle used for the calculation of relative free-enthalpy differences $\Delta\Delta G_{BA}^{\beta\alpha} = \Delta G_{BA}^{\beta} - \Delta G_{BA}^{\alpha}$, where A and B denote two different amino acid sequences, while α and β denote two different folds. The vertical branches of the cycle can be computed using enveloping distribution sampling.

the present work, we study in detail the changes in free enthalpy associated with single amino acid mutations in the two folds. For this purpose we use the enveloping distribution sampling (EDS) methodology.[33−36] It belongs to the class of expanded ensemble methods that find widespread use in Monte Carlo[37−45] and molecular dynamics[34,46−52] simulations in cases where non-overlapping phase spaces need to be sampled. EDS has been applied successfully to a number of model systems[33−36] and, after some refinement of the parameter determination procedure, to more complex alchemical perturbations such as those that occur in the calculation of the relative binding free enthalpies of phenylethanolamine N-methyltransferase (PNMT) inhibitors[53] and of netropsin to various base pair combinations in the DNA minor groove.[54] It was also shown to be useful for the calculation of relative free enthalpies between different conformations of helices.[55] However, as EDS is a relatively recently proposed method, its potential has not yet been fully exploited. To evaluate the performance of EDS as an alternative to thermodynamic integration (TI),[56] a question of particular relevance is the allowed degree of non-overlap of the important phase space regions of the EDS end states. By application of the method to different systems, its limits can be explored. The second aim of the present work is a further step toward this goal.

THEORY

The EDS scheme is a special form of the umbrella sampling approach[57] that focuses the sampling on the relevant phase space of a system presenting N_H different states n using a reference Hamiltonian H_R defined as[33,38,44,49]

$$H_R(\mathbf{r}, \mathbf{p}) = -\frac{1}{\beta s} \ln \sum_{n=1}^{N_H} \exp\{-\beta s[H_n(\mathbf{r}, \mathbf{p}) - E_n^R]\} \quad (1)$$

where $\mathbf{r}$ and $\mathbf{p}$ denote the $3N$-dimensional vectors of spatial coordinates and conjugate momenta, respectively. N is the number of atoms in the system, and $\beta = 1/k_BT$, where k_B is the Boltzmann constant and T the absolute temperature. The smoothness parameter s $(0 < s \leq 1)$ controls the barriers determining the rate of transitions between the different states n. The energy offsets E_n^R of the different states serve to bring their energies to the same level for even sampling of the states n. These parameters have to be determined iteratively prior to the EDS production simulation.[35,53,54] The free-enthalpy difference between a pair of states (n and m) is calculated based on this production simulation using

$$G_n - G_m = -\frac{1}{\beta} \ln \frac{\langle \exp\{-\beta[H_n(\mathbf{r}, \mathbf{p}) - H_R(\mathbf{r}, \mathbf{p})]\}\rangle_R}{\langle \exp\{-\beta[H_m(\mathbf{r}, \mathbf{p}) - H_R(\mathbf{r}, \mathbf{p})]\}\rangle_R} \quad (2)$$

where $\langle ... \rangle_R$ denotes ensemble averaging over a simulation generated using the reference-state Hamiltonian H_R (eq 1).

In force fields typically used for molecular simulations, the potential energy function is independent of velocity. As a result, the kinetic contribution to the free enthalpy can be calculated analytically. For a two-state system, the potential energy term of the reference-state Hamiltonian then reads

$$V_R(\mathbf{r}; s, \Delta E_{BA}^R) = -\frac{1}{\beta s} \ln\{\exp[-\beta s(V_A(\mathbf{r}) - E_A^R)] + \exp[-\beta s(V_B(\mathbf{r}) - E_B^R)]\} \quad (3)$$

In the reference-state ensemble R, the distribution of the energy difference $\Delta V_{BA}(t) = V_B(t) - V_A(t)$ between the two end states can be calculated as[58,59]

$$\rho_R(\Delta V_{BA}) = \langle \delta[\Delta V_{BA} - (V_B(\mathbf{r}) - V_A(\mathbf{r}))]\rangle_R \quad (4)$$

where δ is the Dirac delta function (approximated in practice by a histogram−binning function). This distribution can be converted to predicted distributions, denoted by a tilde, corresponding to the end-state ensemble A or B via reweighting,[60,61]

$$\tilde{\rho}_n(\Delta V_{BA}) = \frac{\langle \delta[\Delta V_{BA} - (V_B(\mathbf{r}) - V_A(\mathbf{r}))] \exp\{-\beta[V_n(\mathbf{r}) - V_R(\mathbf{r})]\}\rangle_R}{\langle \exp\{-\beta[V_n(\mathbf{r}) - V_R(\mathbf{r})]\}\rangle_R} \quad (5)$$

with n = A or B, respectively. The distributions given by eqs 4 and 5 are used to gain insight into the quality of the sampling obtained in the EDS simulation.[54]

In a similar manner, distributions of the potential energies V_A and V_B sampled in the reference-state ensemble R, $\rho_R(V_n)$, can be converted to predicted distributions corresponding to end-state ensemble A or B, $\tilde{\rho}_n(V_n)$. These predicted distributions can be compared to actual distributions $\rho_n(V_n)$ based on independent simulations at the two end states and thus used for further assessment of the quality of the sampling obtained in the EDS simulation.[62]

COMPUTATIONAL DETAILS

Simulation Parameters. All MD simulations were carried out using a modified version of the GROMOS05 program package[63] and the 54A7 force field.[64,65] The simulations were performed under periodic boundary conditions based on cubic boxes. Initial coordinates for the two proteins in their native sequence and for the corresponding homology models solvated in 9162 (3α-structure) or 7797 ($4\beta+\alpha$-structure) simple point charge (SPC) water molecules[66] were taken from previous work.[32]

The equations of motion were integrated using the leapfrog algorithm[67] with a time step of 2 fs. Bond lengths and the bond angle of water molecules were constrained by applying the SHAKE algorithm[68] with a relative geometric tolerance of 10^{-4}. The center of mass motion of the computational box was removed every 2 ps. All simulations were performed at constant pressure and temperature. The temperature was maintained at 298 K by weak coupling to an external bath[69] with a relaxation time of 0.1 ps. Solute and solvent were coupled to separate heat baths. The pressure was calculated using a group-based virial and held constant at 1 atm using the weak coupling method with a relaxation time of 0.5 ps[69] and an isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$.

dx.doi.org/10.1021/bi400272q | Biochemistry 2013, 52, 4962−4970

The nonbonded van der Waals and electrostatic interactions were calculated using a twin-range cutoff scheme,[70] with short- and long-range cutoff distances set to 0.8 and 1.4 nm, respectively. The short-range interactions were calculated every time step using a group-based pairlist updated every fifth time step. The intermediate-range interactions were reevaluated at each pairlist update and assumed constant between updates. A reaction field correction[71,72] was applied to account for the mean effect of electrostatic interactions beyond the long-range cutoff distance, using a relative dielectric permittivity ε_{RF} of 61.[73] To be consistent with previous work on this system,[32] the reaction field self-term and excluded-atom-term contributions were not included in the calculation of the total electrostatic energy.[74]

Simulated Systems. End-state simulations, i.e., standard (non-EDS) MD simulations of the eight sequences considered for each of the two folds, were conducted for structural analysis and to determine the distributions of the solute−solute plus solute−solvent nonbonded energies. These distributions can be compared to those obtained from EDS reference-state simulations via reweighting, allowing an assessment of the quality of sampling in the EDS reference-state simulations. The end-state simulations were carried out for 10 ns after a 1 ns equilibration. Atom coordinates and energies were saved for analysis every 5 ps.

EDS reference-state simulations were performed separately for each of the two distinct protein folds to calculate free-enthalpy changes for each of the 12 alchemical transformations defined in Figure 2. These simulations, involving fixed

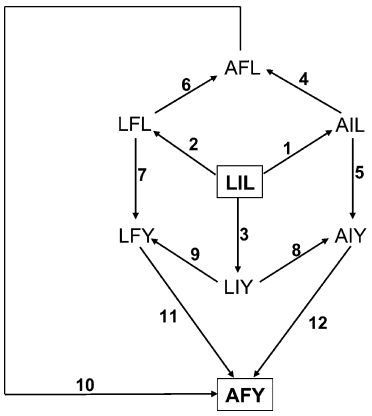


Figure 2. Single-mutation pathways from L20I30L45 to A20F30Y45.

smoothness parameter and energy offset values, were carried out for a period of 10−20 ns after a 1 ns equilibration. Atom coordinates and energies were saved for analysis every 2 ps. The EDS parameters were determined independently for each of the 24 calculations prior to production simulations using the parameter update procedure described previously.[54] The simulation time necessary to obtain converged parameters was 10−20 ns, depending on the perturbation. The optimized values are listed in Table S1 of the Supporting Information.

All end- and reference-state simulations were based on a common topology containing three hybrid building blocks for residues 20, 30, and 45 that contain all atoms to simulate Leu/Ala, Ile/Phe, and Leu/Tyr amino acids, respectively. The single-topology[75] representations of each hybrid building block are shown in Figure S1 of the Supporting Information. Atom types and partial charges are specified in Tables S2−S4 of the Supporting Information.

Trajectory Analysis. All analyses were carried out using the GROMOS++ suite of programs.[76] The trajectories of all end states were analyzed in terms of the atom-positional root-mean-square deviation (rmsd) from the starting configuration of each production simulation. The rmsd was calculated based on the C_α atoms. For the trajectories of all end- and reference-state simulations, secondary structure elements were assigned according to the rules defined by Kabsch and Sander.[77]

The EDS reference-state trajectories were analyzed in terms of the time series of the potential energy difference between state B and state A, $\Delta V_{BA}(t) = V_B(t) - V_A(t)$, the distribution of this potential energy difference in the reference-state ensemble $\rho_R(\Delta V_{BA})$, the potential energy difference distributions predicted for the end-state ensembles, $\tilde{\rho}_A(\Delta V_{BA})$ and $\tilde{\rho}_B(\Delta V_{BA})$, and the solute−solute plus solute−solvent potential energy distributions, $\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$, predicted for the end-state ensembles.

Distributions $\rho_A(V_A)$ and $\rho_B(V_B)$ of solute−solute plus solute−solvent nonbonded energies were calculated for all standard MD (non-EDS) end-state trajectories to compare them with corresponding reweighted end-state potential energy distributions $\tilde{\rho}_A(V_A)$ and $\tilde{\rho}_B(V_B)$ extracted from the EDS reference-state simulations. As the perturbations involve only the solute, solvent−solvent potential energy and entropic contributions exactly cancel out each other and can be omitted.[78−80]

RESULTS AND DISCUSSION

Structural Properties. As a simple estimate of the structural stability during the simulations, the atom-positional root-mean-square deviations of the C_α atoms of each sequence structure combination from the first structure of the trajectory were calculated for all end- and reference-state simulations and are displayed in Figures S2 and S3 of the Supporting Information. The nature of the structural changes corresponding to an increase in rmsd can be seen in the secondary structure analysis of the different sequence−structure combinations (see Figures S4 and S5 of the Supporting Information). In the end-state simulations, the 3α-structure exhibits high structural stability for all sequence combinations except the native sequence of the 4β+α-structure (AFY), for which the rmsd increases to ∼0.4 nm during 11 ns of simulation. This increase in rmsd is accompanied by a loss in the helical character of the first few residues of the first α-helix. For the 4β+α-structure the rmsd values are in most cases slightly higher than those for the 3α-structure, except for the native sequence which shows the lowest rmsd. One sequence (AFL) shows a slowly increasing rmsd that is accompanied by a loss in the helical character of the last few residues of the α-helix.

In the reference-state simulations, the 3α-structure exhibits high structural stability for all perturbations studied, whereas some cases (1, 8, and 12) of the 4β+α-structure show some loss in the helical character of the last few residues of the α-helix.

However, the general picture that arises from this analysis is that all sequence−structure combinations retained their initial secondary structure during the sampling period of the MD simulations, justifying the use of the thermodynamic cycle shown in Figure 1.

Free-Enthalpy Changes. Table 1 summarizes all point mutations studied and the corresponding changes in free enthalpy, ΔG^{α}_{BA} and ΔG^{β}_{BA}, respectively. The subscripts A and B

dx.doi.org/10.1021/bi400272q | Biochemistry 2013, 52, 4962−4970

Table 1. Free-Enthalpy Differences, ΔG^{α}_{BA} and ΔG^{β}_{BA}, and Their Differences, $\Delta\Delta G^{\beta\alpha}_{BA}$, upon Mutating Residues 20, 30, and 45[a]

		$\Delta G^{\alpha/\beta}_{BA}$				$\Delta\Delta G^{\beta\alpha}_{BA}$	
		3α-structure		4β+α-structure			
code	mutation (A → B)	direct	optimized	direct	optimized	direct	optimized
1	LIL → AIL	+7.7 ± 2.0	+6.5 ± 1.8	−11.7 ± 1.6	−9.0 ± 4.6	−19.4 ± 2.6	−15.5 ± 4.9
2	LIL → LFL	−18.3 ± 1.6	−17.2 ± 2.0	−18.7 ± 1.1	−21.0 ± 4.6	−0.4 ± 1.9	−3.9 ± 5.0
3	LIL → LIY	−77.0 ± 0.8	−76.9 ± 2.0	−92.2 ± 1.4	−92.6 ± 2.4	−15.2 ± 1.6	−15.6 ± 3.1
4	AIL → AFL	−16.6 ± 1.2	−17.0 ± 1.1	−23.6 ± 1.9	−21.6 ± 4.6	−7.0 ± 2.2	−4.6 ± 4.7
5	AIL → AIY	−84.3 ± 2.1	−85.1 ± 1.8	−92.6 ± 1.1	−91.9 ± 2.4	−8.3 ± 2.4	−6.8 ± 3.0
6	LFL → AFL	+7.2 ± 1.0	+6.7 ± 1.7	−7.5 ± 0.7	−9.5 ± 4.6	−14.7 ± 1.2	−16.2 ± 4.9
7	LFL → LFY	−77.4 ± 2.4	−75.8 ± 2.0	−90.7 ± 1.4	−91.0 ± 1.5	−13.3 ± 2.8	−15.2 ± 2.5
8	LIY → AIY	−3.1 ± 1.8	−1.7 ± 1.8	−7.3 ± 1.7	−8.4 ± 2.4	−4.2 ± 2.5	−6.7 ± 3.0
9	LIY → LFY	−14.7 ± 1.1	−16.0 ± 2.0	−20.2 ± 2.2	−19.5 ± 1.5	−5.5 ± 2.4	−3.4 ± 2.5
10	AFL → AFY	−82.2 ± 2.0	−83.1 ± 1.7	−92.2 ± 2.0	−92.2 ± 1.1	−10.0 ± 2.8	−9.1 ± 2.0
11	LFY → AFY	−1.0 ± 1.5	−0.7 ± 1.7	−11.2 ± 1.2	−10.8 ± 1.3	−10.2 ± 1.9	−10.1 ± 2.1
12	AIY → AFY	−15.6 ± 1.5	−15.0 ± 1.5	−21.4 ± 2.0	−21.8 ± 1.3	−5.8 ± 2.5	−6.8 ± 2.0

[a]All free enthalpies are reported in kilojoules per mole. The errors in $\Delta\Delta G^{\beta\alpha}_{BA}$ were estimated from the square root of the sum of the squares of the individual errors. The "direct" values are directly obtained from the energy trajectories from the EDS reference-state simulations using eq 2. The "optimized" values are statistically optimal estimates according to ref 81.

denote the two end states of the mutation while the superscripts α and β denote the 3α or 4β+α-structure, respectively. Figure 2 shows the single-mutation pathways resulting from the different amino acid substitutions. The numbers used to label single-mutation pathways in the text refer to this figure. The results for the perturbations in the 3α- and 4β+α-structures are discussed separately in the two following sections. For each mutation we report the free-enthalpy difference directly obtained from the EDS reference-state simulation using eq 2 and a statistically optimal estimate calculated from all free-enthalpy differences within a particular fold, including cycle closure (see Supporting Information for additional details).[81] If not otherwise stated, the values given in the following sections are those obtained directly from the EDS reference-state simulations because the data sets do not differ significantly.

Amino Acid Mutations in the 3α-Structure. The single amino acid mutations 1, 6, 8, and 11 involve the substitution L20 → A20 with all possible combinations for residues 30 and 45. If 45 = L, the change in free enthalpy is positive by approximately 7.5 kJ mol^{-1}, independent of whether 30 = I or 30 = F. However, if 45 = Y, the change in free enthalpy is slightly negative by approximately −2 kJ mol^{-1} (ΔG^{α}_{8} = −3.1 ± 1.8 kJ mol^{-1} for LIY → AIY and ΔG^{α}_{11} = −1 ± 1.5 kJ mol^{-1} for LFY → AFY), again independent of whether 30 = I or 30 = F. The weak effect amino acid 30 has on residues 20 and 45 is also seen from perturbations 2, 4, 9, and 12 that involve the perturbations I30 → F30 with all possible combinations for residues 20 and 45. All free-enthalpy changes are within −18.3 to −14.7 kJ mol^{-1}, virtually independent of the type of the other residues. Perturbations 3, 5, 7, and 10 involve the substitution L45 → Y45. The free-enthalpy changes for perturbations 5 and 10 are consistent with the results discussed so far in that they do not depend on the nature of residue 30. A comparison between perturbation 3 and 5 shows that the change in free enthalpy does depend on residue 20, which is consistent with the results discussed above. For perturbation 7, LFL → LFY, the calculated free-enthalpy difference of −77.4 ± 2.4 kJ mol^{-1} relies on rather poor sampling, as can be seen from the time series of the energy difference $\Delta V_{BA}(t) = V_B(t) - V_A(t)$, shown in Figure S6 of the Supporting Information. However, ΔG^{α}_{7} is very similar to ΔG^{α}_{3} (LIL → LIY) as expected due to the weak effect of the nature of residue 30.

The 12 studied perturbations allow for 6 different pathways to estimate the change in free enthalpy for the process LIL → AFY. The free-enthalpy changes along these pathways obtained from Table 2 are between −91 and −96 kJ mol^{-1}, with a mean value of −94 kJ mol^{-1}.

Table 2. Free-Enthalpy Differences for the Overall Mutation LIL → AFY Calculated along Different Pathways Involving Single Amino Acid Mutations in Each Step (See Figure 2)[a]

	$\Delta G^{\alpha/\beta}$	
path	3α-structure	4β+α-structure
1−4−10	−91.1 ± 3.1	−127.5 ± 3.2
1−5−12	−92.2 ± 3.3	−125.7 ± 2.8
2−6−10	−93.3 ± 2.7	−118.4 ± 2.4
2−7−11	−96.7 ± 3.3	−120.6 ± 2.1
3−8−12	−95.7 ± 2.5	−120.9 ± 3.0
3−9−11	−92.7 ± 2.0	−123.6 ± 2.8
average	−93.6 ± 2.8(3.1)[b]	−122.8 ± 2.7(5.1)[b]

[a]All free enthalpies are reported in kilojoules per mole. The errors in $\Delta G^{\alpha/\beta}$ were estimated from the square root of the sum of the squares of the individual errors. [b]Average error obtained using statistically optimal estimates for each path.

We also tried to obtain direct estimates of free-enthalpy changes for simultaneous perturbations of more than one amino acid. However, except for LIL → AFL, no pair of smoothness parameter and energy offset values could be identified that resulted in sufficient sampling of both end states in a single EDS reference-state simulation. For the perturbation LIL → AFL, the computed value of ΔG^{α} = −12.6 ± 1.5 kJ mol^{-1} is in agreement with the free enthalpies along the 2−6 and 1−4 paths.

For the single amino acid perturbations, the quality of sampling was evaluated by means of the energy difference distribution, the end-state energy distribution functions, and the time series of the energy difference $\Delta V_{BA}(t)$. Figure 3 shows the energy difference and end-state energy distributions for all perturbations studied in the 3α-structure. In all cases the

dx.doi.org/10.1021/bi400272q | Biochemistry 2013, 52, 4962−4970

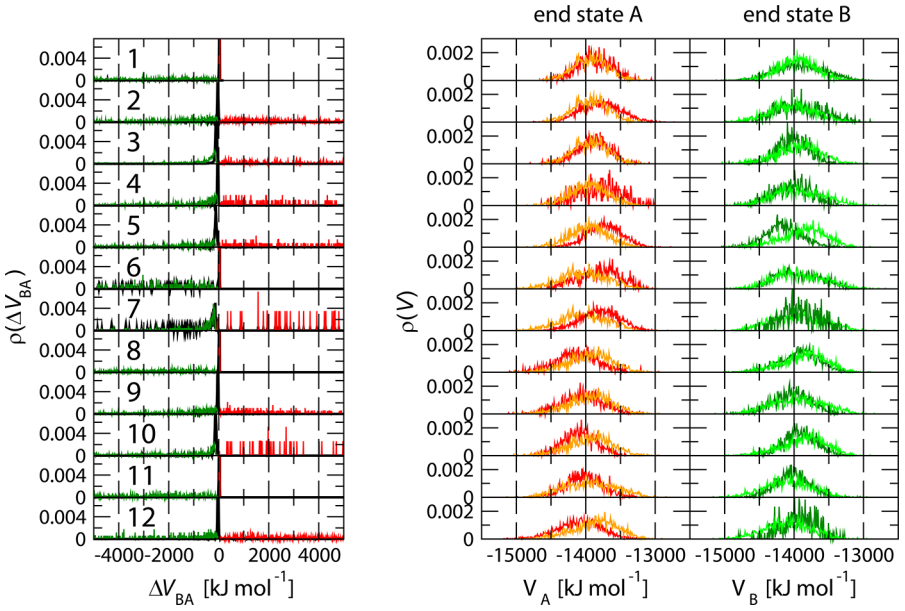


Figure 3. (Left panel) Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\tilde{\rho}_A(\Delta V_{BA})$ (red) and $\tilde{\rho}_B(\Delta V_{BA})$ (green), for each of the 12 perturbations in protein G as obtained from the EDS reference-state simulations of the 3α-structure. The 12 mutations from sequence A to B are defined in Table 1. The energy difference distributions of the end states were determined by reweighting. (Right panel) Nonbonded solute−solute plus solute−solvent energy distributions of the EDS end states obtained through reweighting (red and dark green) from the 12 EDS reference-state simulations and from independent MD simulations of the eight end states (orange and light green).

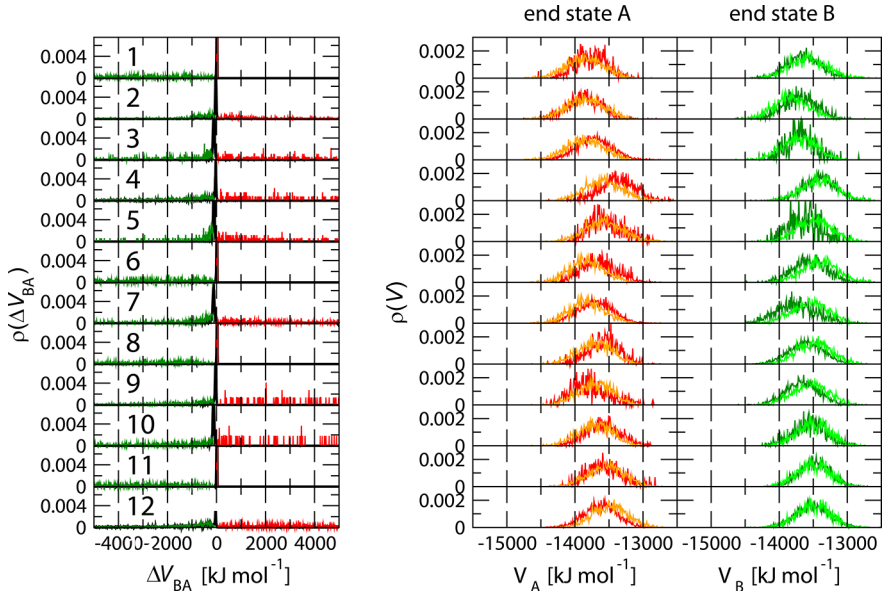


Figure 4. (Left panel) Energy difference distributions for the reference state, $\rho_R(\Delta V_{BA})$ (black), and the two end states, $\tilde{\rho}_A(\Delta V_{BA})$ (red) and $\tilde{\rho}_B(\Delta V_{BA})$ (green), for each of the 12 perturbations in protein G as obtained from the EDS reference-state simulations of the 4β+α-structure. The 12 mutations from sequence A to B are defined in Table 1. The energy difference distributions of the end states were determined by reweighting. (Right panel) Nonbonded solute−solute plus solute−solvent energy distributions of the EDS end states obtained through reweighting (red and dark green) from the 12 EDS reference-state simulations and from independent MD simulations of the eight end states (orange and light green).

reweighted energy difference distributions exhibit high-energy tails, showing that two distinct states are sampled in the EDS production simulations. This is also evident from the time series of ΔV_{BA}, shown in Figure S6 of the Supporting Information. However, the rather long residence times in each state, in some cases amounting to more than 3 ns, are unfavorable for an efficient use of EDS because it gives rise to a significant computational overhead when optimizing the EDS parameters.

The solute−solute plus solute−solvent nonbonded energy distributions obtained from the EDS reference-state simulations through reweighting all have the same shape as those obtained from independent MD simulations of the end states. In some cases (5, 6, 8, and 12) the distributions are not exactly on top of each other, showing the influence of conformational changes on the sampled nonbonded energies. The consequences for the obtained free-enthalpy changes are discussed in the next section.

dx.doi.org/10.1021/bi400272q | *Biochemistry* 2013, 52, 4962−4970

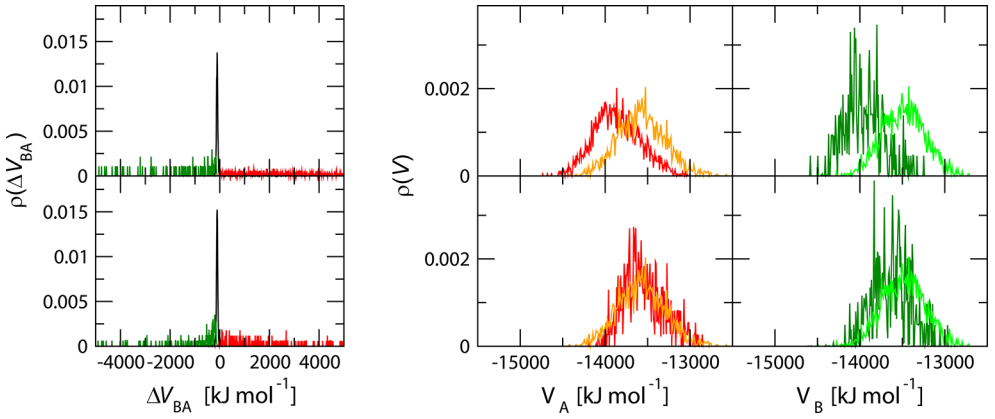


Figure 5. Illustration of the starting coordinate dependence of the energy distributions for the perturbation AIL → AIY in the 4β+α-structure. The top panels show the energy difference and end-state energy distributions obtained when starting the EDS reference-state simulation from the last configuration of the EDS parameter optimization simulation. The bottom panels show the distributions obtained when starting the EDS reference-state simulation from the same coordinates as used in the plain MD simulation of the AIY end state. The quantities and color codes are identical to those used in Figure 4.

The statistically optimal estimates for all free-enthalpy differences agree well with the direct results, with a maximum difference of 1.6 kJ mol^{-1}, further validating the reliability of the latter.

Amino Acid Mutations in the 4β+α-Structure. In the 4β+α-structure, the substitution L20 → A20 (1, 6, 8, and 11) was found to depend weakly on residue 30 in addition to a weak dependence on residue 45. If 30 = I and 45 = L, the change in free enthalpy is $\Delta G_1^\beta = -11.7 \pm 1.6$ kJ mol^{-1}, whereas $\Delta G_6^\beta = -7.5 \pm 0.7$ kJ mol^{-1} if 30 = F. If 45 = Y, the free-enthalpy change of L20 → A20 is $\Delta G_8^\beta = -7.3 \pm 1.7$ kJ mol^{-1} if 30 = I and $\Delta G_{11}^\beta = -11.2 \pm 1.2$ kJ mol^{-1} if 30 = F. Compared to mutations in the 3α-structure, the changes in free enthalpy for L20 → A20 differ less strongly along the possible sequence combinations in the 4β+α-structure. For the perturbations 2, 4, 9, and 12 that involve the substitution I30 → F30, the variation in free enthalpies is slightly enhanced compared to that of the mutations in the 3α-structure. If 20 = A, $\Delta G_4^\beta = -23.6 \pm 1.9$ kJ mol^{-1} for AIL → AFL and $\Delta G_{12}^\beta = -21.4 \pm 2.0$ kJ mol^{-1} for AIY → AFY. If 20 = L, $\Delta G_2^\beta = -18.7 \pm 1.1$ kJ mol^{-1} for LIL → LFL and $\Delta G_9^\beta = -20.2 \pm 2.2$ kJ mol^{-1} for LIY → LFY. The latter result relies on rather poor sampling as can be seen from the time series of the energy difference $\Delta V_{BA}(t) = V_B(t) - V_A(t)$ shown in Figure S6 of the Supporting Information. However, the free-enthalpy change along the 3−9−11 path is in good agreement with those along the other paths.

As was the case for the L20 → A20 mutation, the changes in free enthalpy associated with the perturbation L45 → Y45 in the 4β+α-structure show fewer differences among the possible combinations compared to those associated with mutations in the 3α-structure.

Therefore, we conclude that in the 4β+α structure, the effects of changes in the amino acids are more independent of one another than in the 3α-structure. The free-enthalpy changes along the six different pathways are in the range from −118 to −128 kJ mol^{-1} with a mean value of −123 kJ mol^{-1}.

Direct estimation of free-enthalpy changes for the simultaneous perturbation of more than one amino acid was not successful because no pair of smoothness parameter and energy offset values could be identified that resulted in sufficient sampling of both end states in a single EDS reference-state simulation.

For the single amino acid perturbations, the quality of sampling was evaluated by means of the energy difference distribution, the end-state energy distribution functions, and the time series of the energy difference $\Delta V_{BA}(t) = V_B(t) - V_A(t)$. Figure 4 shows the energy difference and end-state energy distributions for all perturbations studied in the 4β+α-structure. In all cases the reweighted energy difference distributions exhibit high-energy tails, showing that two distinct states are sampled in the EDS production simulations. This is also evident from the time series of ΔV_{BA} shown in Figure S6 of the Supporting Information. Again, the rather long residence times in each state are unfavorable for an efficient use of EDS. The result reported for perturbation 12 relied on the combined analysis of two trajectories sharing the same smoothness parameter value but slightly different energy offsets of −21.0 and −22.0 kJ mol^{-1} because the sampling turned out to be quite sensitive to the energy offset in this case.

The solute−solute plus solute−solvent nonbonded energy distributions obtained from the EDS reference-state simulations through reweighting all have the same shape as those obtained from independent MD simulations of the end states. In some cases (4, 5, 6, and 7) the distributions are not exactly on top of each other, showing that EDS reference-state and end-state simulations have sampled different conformations of the protein. For the most severe case (perturbation 5, AIL → AIY), we have estimated the effect on the free-enthalpy change by performing two simulations, the first starting from the final configuration of the parameter update simulation (the standard procedure used in this work), and the second from the same coordinates and velocities as used for the end-state simulation of the AIY sequence. The obtained end-state energy distributions, shown in Figure 5, differ significantly while the calculated free-enthalpy changes differ by only 4.3 kJ mol^{-1}.

The statistically optimal estimates for the free-enthalpy differences agree well with the direct results, with a maximum difference of 1.1 kJ mol^{-1}, except for those involved in the 1−4−6−2 cycle. This cycle shows a comparatively large hysteresis of 9.1 kJ mol^{-1}, giving rise to larger uncertainites for the involved free-enthalpy differences. The largest difference between direct calculation and optimized results (2.7 kJ mol^{-1}) occurs for perturbation 1, LIL → AIL.

dx.doi.org/10.1021/bi400272q | *Biochemistry* 2013, 52, 4962−4970

Relative Preference of Amino Acids between the 3α- and 4β+α-Structures. The computed relative free-enthalpy differences for all single amino acid mutations are also listed in Table 1. A negative sign indicates a preference of the state B sequence for the 4β+α-environment relative to the state A sequence. Perturbations 1, 3, 6, 7, 10 and 11, i.e., those involving the mutations L20 → A20 or L45 → Y45, all show negative $\Delta\Delta G$ values which are at least twice as large as the maximum uncertainty of 5 kJ mol^{-1} and thus show a clear preference for the 4β+α-fold. The six perturbations can be grouped into three pairs. For perturbations 1 and 3, state A represents the native sequence for the 3α-structure, leading to $\Delta\Delta G$ values with the largest magnitude. Perturbations 6 and 7 have the sequence LFL in state A which differs from the native sequence by the amino acid at position 30. The effect of this substitution is marginal, as expected, and the $\Delta\Delta G$ value is comparable to that of perturbations 1 and 3. Finally, for perturbations 10 and 11, two of the three residues that distinguish the native G_B95-sequence from the G_A95-sequence are already present and the effect of the last perturbation is smaller than that for the former cases, illustrating the importance of the context in which a mutation is performed.

For perturbations 2, 4, 5, 8, 9, and 12, no or only weak preferences for the 4β+α-structure are visible. Four cases of the latter set (2, 4, 9, and 12) involve the transformation I30 → F30, which is not affected by a change in the secondary structure of amino acid 30. Perturbation 5 (AIL → AIY) shows a preference for the 4β+α-environment that is only slightly smaller in magnitude than −10 kJ mol^{-1}. Perturbation 8 does not show a clear preference for the 4β+α-environment, although it involves the mutation L20A. The latter result is supported by experimental data that show that the variants are unfolded when 45 = Y and 30 = I, meaning that neither of the two structural environments is preferred in this case.[14] This shows again the importance of the context in which a mutation is performed.

The calculated relative free-enthalpy differences are compatible with the experimental findings that the amino acid mutations L20A and L45Y are responsible for fold switching.[14,82] Moreover, they also show reasonable quantitative agreement. For perturbation 7 (LFL → LFY), the calculations predict a relative preference of the LFY sequence for the 4β+α-fold by −13.3 ± 2.8 kJ mol^{-1}. It was found experimentally that the LFL sequence prefers to fold into the 3α-structure (>90% populated at 20 °C),[14] while the relative gain in stability of the 4β+α-structure introduced by the L → Y mutation at position 45 is enough to shift the equilibrium to this state almost completely (≈99%).[82] The corresponding relative free-enthalpy difference can be calculated as $\Delta\Delta G^{\beta\alpha}_{\mathrm{LFL}\rightarrow\mathrm{LFY}} = -RT(\ln 99/1 + \ln 90/10) = -16.6$ kJ mol^{-1}, in good agreement with the simulated result. For perturbation 11 (LFY → AFY), He et al.[82] report switching from a predominantly (≈95%) 3α-fold to a 4β+α-fold (>95% populated). The corresponding relative free-enthalpy difference is $\Delta\Delta G^{\beta\alpha}_{\mathrm{LFY}\rightarrow\mathrm{AFY}} = -RT(\ln 95/5 + \ln 95/5) = -14.4$ kJ mol^{-1}, while our simulated result is −10.2 ± 1.9 kJ mol^{-1}. Note that for this particular case the experimental result corresponds to a sequence with isoleucine at position 25, while the simulated sequence has threonine at this position.

CONCLUSION

The present study served two purposes. First, it complements earlier work on the same system focusing on structural and energetic analyses[32] with an extensive thermodynamic analysis in terms of free enthalpies. Second, the feasibility of calculating free-enthalpy differences for point mutations using the method of enveloping distribution sampling (EDS) was evaluated. With respect to the first purpose, the 12 calculated free-enthalpy changes for each fold showed that the three investigated residues, 20, 30, and 45, do not act independently but that the free-enthalpy change for the L20A mutation depends on the nature of residue 45 while the free-enthalpy change for the L45Y mutation depends on the nature of residue 20. These interdependencies in the 3α-structure are stronger than those in the 4β+α-structure. The changes in free enthalpy associated with the I30F mutation are essentially independent of other residue changes. The calculation results support the experimental findings that residues 20 and 45 are the main conformational switches for this protein pair.

With respect to the second purpose, we conclude that the investigated point mutations constitute a borderline case for the efficient use of the EDS methodology without any intermediate states between the two end states. The reasons are (i) the significant computational overhead caused by the EDS parameter optimization, a process which requires sufficient transitions between the two end states, and (ii) the strong sensitivity of the sampling quality on the EDS parameters, in particular the energy offset. A strong sensitivity requires a careful parameter optimization, which again increases the computational cost. For future studies, involving large perturbations, it might therefore be useful to include a limited number of intermediate states (one or two) between the end states to increase the phase space overlap and make the parameter optimization more efficient.

ASSOCIATED CONTENT

Supporting Information

EDS parameters, details on the topologies, plots showing the atom-positional rmsd and secondary structures during the MD simulations, plots showing the time series of the energy difference between the two end states during the EDS reference-state simulations, and details regarding the computation of statistically optimal free-enthalpy differences. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wfvgn@igc.phys.chem.ethz.ch. Phone: +41 44 632 5501. Fax: +41 44 632 1039.

Funding

This work was supported by the National Center of Competence in Research (NCCR) in Structural Biology, by Grant 200020-137827 from the Swiss National Science Foundation, and by Grant 228076 from the European Research Council.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Anfinsen, C. B. (1973) Principles that govern folding of protein chains. *Science 181*, 223−330.

(2) Kidera, A., Konishi, Y., Ooi, T., and Scheraga, H. A. (1985) Relation between sequence similarity and structural similarity in proteins. Role of important properties of amino acids. *J. Protein Sci. 4*, 265−297.

(3) Zwanzig, R., Szabo, A., and Bagchi, B. (1992) Levinthal's paradox. *Proc. Natl. Acad. Sci. U.S.A. 89*, 20−22.
(4) Ambroggio, X. I., and Kuhlmann, B. (2006) Design of protein conformational switches. *Curr. Opin. Struct. Biol. 16*, 525−530.
(5) Harano, Y. (2012) Application of hydration thermodynamics to the evaluation of protein structures and protein-ligand binding. *Entropy 14*, 1443−1468.
(6) Ben-Naim, A. (2013) *The Protein Folding Problem and Its Solutions*, World Scientific, New York.
(7) Rosato, A., Bagaria, A., Baker, D., Bardiaux, B., Cavalli, A., Doreleijers, J. F., Giachetti, A., Guerry, P., Güntert, P., Herrmann, T., Huang, Y. J., Jonker, H. R. A., Mao, B., Malliavin, T. E., Montelione, G. T., Nilges, M., Raman, S., van der Schot, G., Vranken, W. F., Vuister, G. W., and Bonvin, A. M. J. J. (2009) CASD-NMR: Critical assessment of automated structure determination by NMR. *Nat. Methods 6*, 625−626.
(8) Guerry, P., and Herrmann, T. (2011) Advances in automated NMR protein structure determination. *Q. Rev. Biophys. 44*, 257−309.
(9) Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., Jaskolski, M., and Baker, D. (2011) Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat. Struct. Mol. Biol. 18*, 1175−1177.
(10) Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S., and Xiao, X. (2011) Biomolecular modeling and simulation: A field coming of age. *Q. Rev. Biophys. 44*, 191−228.
(11) Zhou, Y., Duan, Y., Yang, Y., Faraggi, E., and Lei, H. (2011) Trends in template/fragment-free protein structure prediction. *Theor. Chem. Acc. 128*, 3−16.
(12) Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2007) The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl. Acad. Sci. U.S.A. 104*, 11963−11968.
(13) He, Y., Chen, Y., Alexander, P., Bryan, P. N., and Orban, J. (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. *Proc. Natl. Acad. Sci. U.S.A. 105*, 14412−14417.
(14) Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2009) A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. U.S.A. 106*, 21149−21154.
(15) Bryan, P. N., and Orban, J. (2013) Implications of protein fold switching. *Curr. Opin. Struct. Biol. 23*, 314−316.
(16) Horst, J., and Samudrala, R. (2009) Diversity of protein structures and difficulties in fold recognition: The curious case of protein G. *F1000 Biology Reports 1*, 69.
(17) Shen, Y., Bryan, P. N., He, Y., Orban, J., Baker, D., and Bax, A. (2010) De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Sci. 19*, 349−356.
(18) Bryan, P. N., and Orban, J. (2010) Proteins that switch folds. *Curr. Opin. Struct. Biol. 20*, 482−488.
(19) Rackovsky, S. (2011) Spectral analysis of a protein conformational switch. *Phys. Rev. Lett. 106*, 248101.
(20) Shao, Q., Zhu, W., and Gao, Y. Q. (2012) Robustness in protein folding revealed by thermodynamics calculations. *J. Phys. Chem. B 116*, 13848−13856.
(21) König, G., and Boresch, S. (2009) Hydration free energies of amino acids: Why side chain analog data are not enough. *J. Phys. Chem. B 113*, 8967−8974.
(22) van Gunsteren, W. F. (1988) The role of computer simulation techniques in protein engineering. *Protein Eng. 2*, 5−13.
(23) Levy, R. M., and Gallicchio, E. (1998) Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modelling of electrostatic effects. *Annu. Rev. Phys. Chem. 49*, 531−567.
(24) König, G., Bruckner, S., and Boresch, S. (2013) Absolute hydration free energies of blocked amino acids: Implications for protein solvation and stability. *Biophys. J. 104*, 453−462.
(25) Meinke, J. H., Mohanty, S., Nadler, W., Zimmermann, O., and Hansmann, U. H. E. (2009) Computer simulation of proteins: Thermodynamics and structure prediction. *Eur. Phys. J. D51*, 33−40.
(26) Schulz, R., Lindner, B., Petridis, L., and Smith, J. C. (2009) Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer. *J. Chem. Theory Comput. 5*, 2798−2808.
(27) Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science 330*, 341−346.
(28) IUPAP (1978) *Symbols, Units and Nomenclature in Physics*, Vol. *93A*, pp 1−60, Physica, Amsterdam.
(29) IUPAC (1988) *Quantities, Units and Symbols in Physical Chemistry*, Blackwell Scientific Publications, Oxford, U.K.
(30) Freddolino, P. L., Park, S., Roux, B., and Schulten, K. (2009) Force field bias in protein folding simulations. *Biophys. J. 96*, 3772−3780.
(31) Faver, J. C., Benson, M. L., He, X., Roberts, B. P., Wang, B., Marshall, M. S., Sherrill, C. D., and Merz, K. M. (2011) The energy computation paradox and ab initio protein folding. *PLoS One 6*, e18868.
(32) Allison, J. R., Bergeler, M., Hansen, N., and van Gunsteren, W. F. (2011) Current computer modeling cannot explain why two highly similar sequences fold into different structures. *Biochemistry 50*, 10965−10973.
(33) Christ, C. D., and van Gunsteren, W. F. (2007) Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *J. Chem. Phys. 126*, 184110.
(34) Christ, C. D., and van Gunsteren, W. F. (2008) Multiple free energies from a single simulation: Extending enveloping distribution sampling to nonoverlapping phase-space distributions. *J. Chem. Phys. 128*, 174112.
(35) Christ, C. D., and van Gunsteren, W. F. (2009) Simple, efficient, and reliable computation of multiple free energy differences from a single simulation: A reference Hamiltonian parameter update scheme for enveloping distribution sampling. *J. Chem. Theory Comput. 5*, 276−286.
(36) Christ, C. D., and van Gunsteren, W. F. (2009) Comparison of three enveloping distribution sampling Hamiltonians for the estimation of multiple free energy differences from a single simulation. *J. Comput. Chem. 30*, 1664−1679.
(37) Berg, B. A., and Neuhaus, T. (1992) Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett. 68*, 9−12.
(38) Han, K. K. (1992) A new Monte Carlo method for estimating free energy and chemical potential. *Phys. Lett. A 165*, 28−32.
(39) Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V., and Vorontsov-Velyaminov, P. N. (1992) New approach to Monte Carlo calculations of the free energy: Method of expanded ensembles. *J. Chem. Phys. 96*, 1776−1783.
(40) Marinari, E., and Parisi, G. (1992) Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett. 19*, 451−458.
(41) Escobedo, F. A., and de Pablo, J. J. (1995) Monte Carlo simulation of the chemical potential of polymers in an expanded ensemble. *J. Chem. Phys. 103*, 2703−2710.
(42) Smith, G. R., and Bruce, A. D. (1995) A study of the multicanonical Monte Carlo method. *J. Phys. A: Math. Gen. 28*, 6623−6643.
(43) Geyer, C. J., and Thompson, E. A. (1995) Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Stat. Assoc. 90*, 909−920.
(44) Han, K. K. (1996) Multiensemble sampling: An alternative efficient Monte Carlo technique. *Phys. Rev. E 54*, 6906−6910.
(45) Wu, D. (2010) An efficient umbrella potential for the accurate calculation of free energies by molecular simulation. *J. Chem. Phys. 133*, 044115.
(46) Kong, X., and Brooks, C. L., III (1996) λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys. 105*, 2414−2423.

(47) Mitsutake, A., Sugita, Y., and Okamoto, Y. (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers 60*, 96−123.
(48) Bitetti-Putzer, R., Yang, W., and Karplus, M. (2003) Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett. 377*, 633−641.
(49) Chen, Y. G., and Hummer, G. (2007) Slow conformational dynamics and unfolding of the calmodulin C-terminal domain. *J. Am. Chem. Soc. 129*, 2414−2415.
(50) Li, H., Fajer, M., and Yang, W. (2007) Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J. Chem. Phys. 126*, 024106.
(51) Paluch, A. S., Shah, J. K., and Maginn, E. J. (2011) Efficient solvation free energy calculations of amino acid analogs by expanded ensemble molecular dynamics simulation. *J. Chem. Theory Comput. 7*, 1394−1403.
(52) Okamoto, Y. (2012) Molecular simulations in generalised ensemble. *Mol. Simul. 38*, 1282−1296.
(53) Riniker, S., Christ, C. D., Hansen, N., Mark, A. E., Nair, P. C., and van Gunsteren, W. F. (2011) Comparison of enveloping distribution sampling and thermodynamic integration to calculate binding free energies of phenylethanolamine N-methyltransferase. *J. Chem. Phys. 135*, 024105.
(54) Hansen, N., Dolenc, J., Knecht, M., Riniker, S., and van Gunsteren, W. F. (2012) Assessment of enveloping distribution sampling to calculate relative free enthalpies of binding for eight netropsin-DNA duplex complexes in aqueous solution. *J. Comput. Chem. 33*, 640−651.
(55) Lin, Z., Liu, H., Riniker, S., and van Gunsteren, W. F. (2011) On the use of enveloping distribution sampling (EDS) to compute free enthalpy differences between different conformational states of molecules: Application to 3_{10}-, α-, and π-helices. *J. Chem. Theory Comput. 7*, 3884−3897.
(56) Kirkwood, J. G. (1935) Statistical mechanics of fluid mixtures. *J. Chem. Phys. 3*, 300−313.
(57) Torrie, G. M., and Valleau, J. P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys. 23*, 187−199.
(58) Jacucci, G., and Quirke, N. (1982) Free energy calculations for crystals. *Lect. Notes Phys. 166*, 38−57.
(59) Powles, J. G., Evans, W. A. B., and Quirke, N. (1982) Non-destructive molecular-dynamics simulation of the chemical potential of a fluid. *Mol. Phys. 46*, 1347−1370.
(60) Salsburg, Z. W., Jacobson, J. D., Fickett, W., and Wood, W. W. (1959) Application of the Monte Carlo method to the lattice-gas model. I. Two-dimensional triangular lattice. *J. Chem. Phys. 30*, 65−72.
(61) Ferrenberg, A. M., and Swendsen, R. H. (1988) New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett. 61*, 2635−2638.
(62) Wu, D., and Kofke, D. A. (2005) Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J. Chem. Phys. 123*, 054103.
(63) Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Kräutler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and van Gunsteren, W. F. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem. 26*, 1719−1751.
(64) Poger, D., van Gunsteren, W. F., and Mark, A. E. (2010) A new force field for simulating phosphatidylcholine bilayers. *J. Comput. Chem. 31*, 1117−1125.
(65) Schmid, N., Eichenberger, A., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and van Gunsteren, W. F. (2011) Definition and testing of the GROMOS force-field versions: 54A7 and 54B7. *Eur. Biophys. J. 40*, 843−856.
(66) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) in *Intermolecular Forces* (Pullmann, B., Ed.) pp 331−342, Reidel, Dordrecht, The Netherlands.
(67) Hockney, R. W. (1970) The potential calculation and some applications. *Methods Comput. Phys. 9*, 136−211.
(68) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys. 23*, 327−341.
(69) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684−3690.
(70) Berendsen, H. J. C. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 18−22, Polycrystal Book Service, Western Springs, IL.
(71) Barker, J. A., and Watts, R. O. (1973) Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys. 26*, 789−792.
(72) Tironi, I. G., Sperb, R., Smith, P. E., and van Gunsteren, W. F. (1995) A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys. 102*, 5451−5459.
(73) Heinz, T. N., van Gunsteren, W. F., and Hünenberger, P. H. (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *J. Chem. Phys. 115*, 1125−1136.
(74) van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich.
(75) Pearlman, D. A. (1994) A comparison of alternative approaches to free energy calculations. *J. Phys. Chem. 98*, 1487−1493.
(76) Eichenberger, A. P., Allison, J. R., Dolenc, J., Geerke, D. P., Horta, B. A. C., Meier, K., Oostenbrink, C., Schmid, N., Steiner, D., Wang, D., and van Gunsteren, W. F. (2011) The GROMOS++ software for the analysis of biomolecular simulation trajectories. *J. Chem. Theory Comput. 7*, 3379−3390.
(77) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577−2637.
(78) Yu, H.-A., and Karplus, M. (1988) A thermodynamic analysis of solvation. *J. Chem. Phys. 89*, 2366−2379.
(79) Peter, C., Oostenbrink, C., van Dorp, A., and van Gunsteren, W. F. (2004) Estimating entropies from molecular dynamics simulations. *J. Chem. Phys. 120*, 2652−2661.
(80) van Gunsteren, W. F., Geerke, D. P., Trzesniak, D., Oostenbrink, C., and van der Vegt, N. F. A. (2007) in *Protein Folding and Drug Design, Proceedings of the International School of Physics "Enrico Fermi", course CLXV* (Broglia, R. A., Serrano, L., and Tiana, G., Eds.) pp 177−191, IOS Press, Amsterdam.
(81) Wang, L., Deng, Y., Knight, J. L., Wu, Y., Kim, B., Sherman, W., Shelley, J. C., Lin, T., and Abel, R. (2013) Modeling local structural rearrangements using FEP/REST: Application to relative binding affinity predictions of CDK2 inhibitors. *J. Chem. Theory Comput. 9*, 1282−1293.
(82) He, Y., Chen, Y., Alexander, P., Bryan, P. N., and Orban, J. (2012) Mutational tipping points for switching protein folds and functions. *Structure 20*, 283−291.