J|A|C|S COMMUNICATIONS

Published on Web 05/11/2010

Prediction of Folding Equilibria of Differently Substituted Peptides Using One-Step Perturbation

Zhixiong Lin,[†,‡] Jörgen Kornfeld,[†] Markus Mächler,[†] and Wilfred F. van Gunsteren*[,†]

Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, 8093 Zürich, Switzerland, and School of Life Sciences and Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China

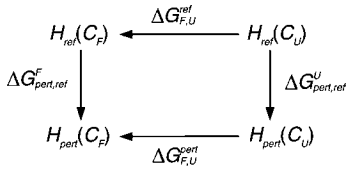
Received February 1, 2010; E-mail: wfvgn@igc.phys.chem.ethz.ch

Today, the folding equilibrium of short polypeptides can be sampled using molecular dynamics (MD) computer simulation over hundreds of nanoseconds.[1−4] For example, the folding equilibrium of a hepta-β-peptide with different side-chain substitutions was investigated by such long-time MD simulations, which showed that a slight change of the side-chain substitution did affect the folding equilibrium dramatically.[5] However, a systematic investigation of the influence of the side-chain composition and position at the backbone is computationally as well as experimentally too expensive because of the exponentially growing number of possible side-chain compositions and combinations along the peptide chain. Here, we show that application of the one-step perturbation technique[6−10] may solve this problem, at least computationally; that is, one can sample many folding equilibria of a polypeptide with different side-chain substitutions from just one single MD simulation.

Assume that we have available a conformational ensemble representing a peptide (un)folding equilibrium that was generated in one long MD simulation using a reference state Hamiltonian H_{ref}. If we now divide the conformational ensemble into two subensembles C_{F} and C_{U}, i.e., representing folded versus unfolded conformations, we can calculate the free enthalpy or Gibbs free energy difference between C_{F} and C_{U}, $\Delta G^{\mathrm{ref}}_{\mathrm{F,U}}$, using the expression

$$\Delta G^{\mathrm{ref}}_{\mathrm{F,U}} = -k_{\mathrm{B}}T \ln(N^{\mathrm{ref}}_{\mathrm{F}}/N^{\mathrm{ref}}_{\mathrm{U}}) \tag{1}$$

where the number of configurations belonging to subensembles C_{F} or C_{U} is denoted by N_{F} or N_{U}, respectively, and $k_{\mathrm{B}}T$ is the Boltzmann constant multiplied by the temperature. Using the thermodynamic cycle, we may calculate the corresponding free enthalpy $\Delta G^{\mathrm{pert}}_{\mathrm{F,U}}$ for



(2)

a perturbed state Hamiltonian H_{pert}, i.e., the peptide with a slightly different side-chain composition and position at the backbone. According to statistical mechanics, the free enthalpy change of a (sub)ensemble C due to changing the Hamiltonian from H_{ref} to H_{pert} is given by

$$\Delta G^{\mathrm{C}}_{\mathrm{pert,ref}} = -k_{\mathrm{B}}T \ln\langle \mathrm{e}^{-(H_{\mathrm{pert}}-H_{\mathrm{ref}})/k_{\mathrm{B}}T}\rangle_{\mathrm{ref,C}} \tag{3}$$

where the (sub)ensemble averaging is denoted by $\langle ... \rangle$ and is carried out over all configurations that were generated using H_{ref} and that belong to the conformational (sub)ensemble C. We note that formula (3) is exact, i.e., does not contain an approximation. The accuracy of the obtained value for ΔG will, however, depend on whether the conformational ensemble generated using H_{ref} that is used in the averaging comprises conformations that are relevant in the ensemble that belongs to H_{pert}. Using the thermodynamic cycle shown in eq 2, we have

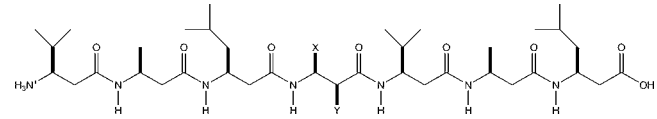


Figure 1. Chemical formulas of the hepta-β-peptides studied, where X,Y are soft-core atoms in the reference state and (X,Y) = (S,S) (CH_3,CH_3), (S,S) (CH_3,H), (S,S) (H,CH_3), (S,S) (H,H) in the perturbed states.

$$\Delta G^{\mathrm{pert}}_{\mathrm{F,U}} = \Delta G^{\mathrm{ref}}_{\mathrm{F,U}} + \Delta G^{\mathrm{F}}_{\mathrm{pert,ref}} - \Delta G^{\mathrm{U}}_{\mathrm{pert,ref}} \tag{4}$$

From a single equilibrium simulation using a possibly unphysical reference state Hamiltonian H_{ref}, we may thus derive the free enthalpy differences $\Delta G^{\mathrm{pert}}_{\mathrm{F,U}}$ for many different perturbed state Hamiltonians H_{pert}. We note that the ΔG obtained via eq 4 does not explicitly depend on the choice of H_{ref}, because its free enthalpy is subtracted out. Therefore, the choice of H_{ref} may be used to optimize the sampling of conformational space such that the ΔG values from eq 4 are as accurate as possible.

In this Communication, we report the results of using this one-step perturbation technique to predict the folding equilibrium of a hepta-β-peptide[11,12] with different side-chain substitutions (Figure 1), for which experimental NMR and CD data are available.[11] Four peptides with different combinations of CH_3 and H as X,Y atoms (Figure 1) were chosen to test this technique, because large conformational ensembles were available from a previous study[5] of these peptides that could serve to validate the one-step perturbation results. As reference state was chosen the peptide with soft-core[13] side-chain X,Y atoms. Soft-core interactions allow a spatial overlap between atoms, the X,Y solute atoms with other solute atoms and with solvent molecules in this case. This allows the simulation in the reference state to sample the conformational space important to the differently perturbed peptides with CH_3 and H substitutions, which is a prerequisite for obtaining accurate free enthalpies through one-step perturbation. Two values of the soft-core parameter α_{LJ} in the Lennard-Jones interaction (see Supporting Information) of the reference state were used: 1.0 and 1.51, referred to as $R_{\alpha 1.0}$ and $R_{\alpha 1.51}$ below. Use of the larger value of $\alpha_{\mathrm{LJ}} = 1.51$, i.e., softer X,Y atoms, leads to increased sampling of unfolded, i.e., non-3_{14}-helical conformations (see Figure 2). Simulations of these two unphysical reference states at $T = 310$ K and 1 atm in explicit methanol solvent were run for 200 ns (the simulation details

[†] Swiss Federal Institute of Technology.
[‡] University of Science and Technology of China.

 10.1021/ja100879k

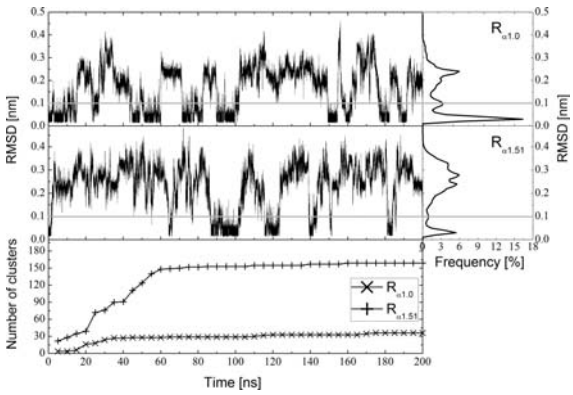


Figure 2. Time evolution of the backbone atom-positional rmsd of residues 2−6 of the hepta-β-peptide with respect to the 3_{14}-helix and the corresponding distribution (upper panels). Number of conformational clusters (lower panel). The gray lines represent the criterion used to distinguish folded and unfolded conformations.

Table 1. Folding Free Enthalpies and Perturbation Free Enthalpies (kJ mol^{-1}) Associated with Perturbations from the Two Reference States to the Four Peptides with Different Side-Chain Substitutions[a]

	(X,Y)			
	(S,S) (CH_3,CH_3)	(S,S) (CH_3,H)	(S,S) (H,CH_3)	(S,S) (H,H)
$\Delta G^{sim}_{F,U}$[b]	-2.1 ± 0.2	2.9 ± 0.4	5.4 ± 0.4	8.3 ± 0.6
		$R_{\alpha 1.0}$		
$\Delta G^{ref}_{F,U}$		1.8 ± 0.4		
$\Delta G^{F}_{pert,ref}$	14.4 ± 0.1	12.2 ± 0.6	12.1 ± 0.4	19.9 ± 1.3
$\Delta G^{U}_{pert,ref}$	16.0 ± 0.2	8.9 ± 1.2	11.1 ± 0.5	19.0 ± 0.3
$\Delta G^{pert}_{F,U}$	0.2 ± 0.4	5.1 ± 1.4	2.8 ± 0.8	2.7 ± 1.4
		$R_{\alpha 1.51}$		
$\Delta G^{ref}_{F,U}$		4.3 ± 0.7		
$\Delta G^{F}_{pert,ref}$	20.3 ± 0.1	17.6 ± 0.1	17.9 ± 0.1	26.0 ± 0.2
$\Delta G^{U}_{pert,ref}$	24.0 ± 0.3	17.2 ± 0.2	17.8 ± 0.3	20.9 ± 0.4
$\Delta G^{pert}_{F,U}$	0.6 ± 0.8	4.7 ± 0.7	4.4 ± 0.8	9.4 ± 0.8

[a] The statistical uncertainties were estimated using block averaging.[15] [b] $\Delta G^{sim}_{F,U}$ calculated through eq 1 from 500 ns simulations.[5]

are described in the Supporting Information). The backbone atom-positional root-mean-square deviations of the peptide in the two reference states with respect to an ideal 3_{14}-helix are shown in Figure 2 as function of time. We chose an rmsd of 0.1 nm (gray line) as our criterion to separate conformations into C_F and C_U. Use of other properties, such as number of intrasolute hydrogen bonds, solute radius of gyration, or solute dipole moment, to separate C_F from C_U does not significantly influence the results[14] obtained for this peptide when varying force-field parameters in a procedure similar to that used here for side-chain substitutions. The value of α_{LJ}, however, has a large influence on the folding equilibrium in the reference state. The free enthalpy of folding $\Delta G^{ref}_{F,U}$ obtained through eq 1 is 1.8 kJ mol^{-1} and 4.3 kJ mol^{-1} for simulations $R_{\alpha 1.0}$ and $R_{\alpha 1.51}$, respectively. The number of conformational clusters, defined using a 0.1 nm rmsd criterion, which make up 95% of the trajectories are also shown in Figure 2 as function of time. Few new clusters appear after 100 ns of simulation, indicating that the sampling has converged. Due to a larger fraction of unfolded conformations, simulation $R_{\alpha 1.51}$ sampled a larger conformational space than $R_{\alpha 1.0}$. Free enthalpy differences of perturbing from the two reference states to the four peptides with different side-chain substitutions (eq 3) are listed in Table 1, together with the predicted folding free enthalpies (eq 4). The statistical uncertainties were estimated using block averaging.[15] Reference simulation $R_{\alpha 1.0}$ predicts $\Delta G_{F,U}$ of the peptides with (S,S) (CH_3,CH_3), (CH_3,H), and (H,CH_3) within k_BT (2.6 kJ mol^{-1}) but yields a deviation of -5.6 kJ mol^{-1} for peptide (H,H). Reference simulation $R_{\alpha 1.51}$ predicts slightly worse for peptide (S,S) (CH_3,CH_3), slightly better for peptide (CH_3,H), and much better for the other two peptides. Especially for peptide (S,S) (H,H), for which simulation $R_{\alpha 1.0}$ completely fails to correctly predict $\Delta G_{F,U}$, simulation $R_{\alpha 1.51}$ yields (eq 4) an accuracy of 1.1 kJ mol^{-1} compared with the simulation result (eq 1). Using $R_{\alpha 1.51}$, the average absolute deviation of predicted folding free enthalpies compared with the ones obtained from long-time MD simulations is 1.6 kJ mol^{-1}, which is half the value of 3.2 kJ mol^{-1} obtained using $R_{\alpha 1.0}$. The former value is of the same order of magnitude as the estimated error of the folding free enthalpy obtained from long-time MD simulations.[16]

Experimental data[11] from CD and NMR measurements on three of the four peptides investigated here confirm the decreasing degree of structural stability of the 3_{14}-helical fold for peptides (S,S) (CH_3,CH_3), (CH_3,H), and (H,H) predicted using the one-step perturbation methodology.

These results show that creation or deletion of one or two atoms is not a too big perturbation when calculating a free enthalpy of folding. One-step perturbation could correctly predict the folding free enthalpy of the peptides with such differences. However, the choice of reference state is very important, as is illustrated by the failure of $R_{\alpha 1.0}$. First, the number of clusters being sampled using Hamiltonian $R_{\alpha 1.0}$ (Figure 2) is obviously insufficient to sample the conformational spaces important to peptides (H,CH_3) and (H,H), which comprise mostly unfolded structures (Table 1). Second, the soft-core parameter value $\alpha_{LJ} = 1.0$ induces high energy when solute atoms or solvent molecules overlap with soft-core solute atoms.[13] When solute atoms or solvent molecules cannot overlap with a soft-core solute atom, perturbing from a soft-core atom to a H atom will result in a cavity, which is an unfavorable configuration. Too limited sampling in the reference simulation causes inaccuracy of the one-step perturbation result. The slightly worse prediction of peptide with (CH_3,CH_3) by $R_{\alpha 1.51}$ compared to $R_{\alpha 1.0}$ has the same cause. Using $R_{\alpha 1.51}$ solute atoms and solvent molecules can overlap with soft-core solute atoms, and not many folded configurations are sampled, which results in a slight overestimate of the free enthalpy of the folded conformation when perturbing from a soft-core atom to CH_3. Taken together, this means that both folded and unfolded conformations must be sufficiently sampled, and this requires a proper reference state Hamiltonian and a temperature for which there are many more unfolded than folded configurations. What constitutes a sufficient fraction of unfolded conformations and a sufficient length of the reference MD simulation is dictated by the type of molecule, i.e., the size of its unfolded conformational space.

It has been shown[1] that the number of different conformers making up the denatured (unfolded) state is very, very low. The denatured state of peptides is orders of magnitude smaller than expected on the basis of considerations given by Levinthal,[17] only a tiny fraction of the possible conformations is populated in the temperature range between 300 and 360 K (see Figure 2). This implies that it is possible to sample a sufficient fraction of unfolded conformations, given a force field that accurately describes the characteristics of this limited denatured state.[1]

In summary, the evaluated one-step perturbation methodology constitutes an efficient technique to predict folding equilibria of peptides with an accuracy of about k_BT. In addition, it could be used to predict other observables or quantities than $\Delta G_{F,U}$ for a particular system, which makes it a powerful molecular simulation methodology that reduces the number of required separate simulations by an order of magnitude.

Acknowledgment. This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-121913 of the Swiss National Science Foundation, by grant number 228076 of the European Research Council (ERC), and by grants IZLCZ2-123884 (Swiss) and GJHZ0906 (Chinese) of the Sino-Swiss Science and Technology Cooperation Program, which are gratefully acknowledged.

Supporting Information Available: Description of the computational methods used in the present work and studies of phase space overlap between reference states and real peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) van Gunsteren, W. F.; Bürgi, R.; Peter, C.; Daura, X. *Angew. Chem., Int. Ed.* **2001**, *40*, 351–355.
(2) Brooks, C. L. *Acc. Chem. Res.* **2002**, *35*, 447–454.
(3) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.
(4) van Gunsteren, W. F.; Gattin, Z. In *Simulation of Folding Equilibria in Foldamers: Structure, Properties, and Applications*; Hecht, S., Huc, I., Eds.; Wiley: Weinheim, Germany, 2007; pp 173−192.
(5) Keller, B.; Gattin, Z.; van Gunsteren, W. F. *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 1677−1690.
(6) Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1996**, *100*, 9485–9494.
(7) Hünenberger, P. H.; Granwehr, J. K.; Aebischer, J. N.; Ghoneim, N.; Haselbach, E.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 7533–7544.
(8) Oostenbrink, C.; van Gunsteren, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6750–6754.
(9) Cao, Z. X.; Liu, H. Y. *J. Chem. Phys.* **2008**, *129*, 15101–15109.
(10) Hritz, J.; Oostenbrink, C. *J. Phys. Chem. B* **2009**, *113*, 12711–12720.
(11) Seebach, D.; Ciceri, P. E.; Overhand, M.; Jaun, B.; Rigo, D.; Oberer, L.; Hommel, U.; Amstutz, R.; Widmer, H. *Helv. Chim.* **1996**, *79*, 2043–2066.
(12) Daura, X.; van Gunsteren, W. F.; Rigo, D.; Jaun, B.; Seebach, D. *Chem.—Eur. J.* **1997**, *3*, 1410–1417.
(13) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
(14) Lin, Z. X.; Liu, H. Y.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, in press.
(15) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(16) Boned, R.; van Gunsteren, W. E.; Daura, X. *Chem.—Eur. J.* **2008**, *14*, 5039–5046.
(17) Levinthal, C. *J. Chim. Phys.* **1968**, *65*, 44–45.

JA100879K