Solvation Free Energies of Amino Acids Calculated by Molecular Dynamics/Free Energy Perturbation Method

Noriyuki YAMAOTSU, Ikuo MORIGUCHI, and Shuichi HIRONO*

School of Pharmaceutical Sciences, Kitasato University, 5–9–1, Shirokane, Minato-ku, Tokyo 108, Japan.
Received August 3, 1994; accepted January 24, 1995

It has been accepted that free energy differences between amino acids calculated using computer simulations are in good agreement with the corresponding experimental values. In recent years, however, Sharp *et al.* [*Biochemistry*, 30, 9686 (1991)] pointed out that the experimental solvation free energies had been underestimated and suggested reevaluation of the extent of agreement between the experiment and computer simulations. We calculated the free energy differences of transfer from vapor to water between neutral amino acids using molecular dynamics/free energy perturbation method and compared the calculated values ($\Delta\Delta G_{calc}$) with both the uncorrected values ($\Delta\Delta G_{uncorr}$) obtained by Wolfenden *et al.* [*Science*, 206, 575 (1979); *Biochemistry*, 20, 849 (1981)] and the corrected values ($\Delta\Delta G_{corr}$) by Sharp *et al.* considering the effect of solute-solvent size differences. For the uncorrected values, the correlation coefficient was $r=0.938$ and the simple regression equation was $\Delta\Delta G_{uncorr}=0.823\Delta\Delta G_{calc}-0.235$. For the corrected values, the correlation coefficient and the simple regression equation were $r=0.987$ and $\Delta\Delta G_{corr}=1.042\Delta\Delta G_{calc}-0.172$, respectively. It was shown that the calculated values are in better agreement with the corrected values than with the uncorrected values.

Key words hydrophobicity; solvation; free energy; α-amino acid; molecular dynamics

The side chain hydrophobicity of an amino acid residue has provided a major unifying concept in understanding the conformation,[1] the stability of proteins[2] and the affinity of a substrate to enzymes.[3] A quantitative measure of the hydrophobicity is usually obtained from the solvation free energy differences between amino acids. These free energies were first measured by Nozaki and Tanford and have been obtained in many other studies.[4] In recent years, Sharp *et al.* pointed out that the experimental solvation free energies had been underestimated.[5] For transfer from vapor to solution, the standard form of the expression used to relate solvation free energy to observed concentrations is $\Delta G=-RT\ln K$, where K is a partition coefficient, R is the molar gas constant, and T is the absolute temperature. Experimental solvation free energies are usually based on this form. When the solute and solvent molecules are different in size, the effect of solute-solvent size differences enters into the expression, $\Delta G=-RT\ln K+RT(1-V_s/V_v)$, where V_s and V_v are the molar volumes for the solute and solvent, respectively. This leads to much larger solvation free energies and, as a result, to a different hydrophobicity.

The solvation free energy of a given compound cannot always be determined by experiments. Therefore, a method to determine the free energy difference without experiments is indeed important. The molecular dynamics (MD)/free energy perturbation (FEP) method may be a very useful tool for this determination. Free energy differences obtained by computer simulations were compared with the values based on the standard form,[6] but Sharp *et al.* suggested the need to reevaluate the extent of agreement between the experiment and computer simulations.[5]

In order to do the reevaluation, we calculated the free energy differences of transfer from vapor to water between neutral, mainly aliphatic, amino acids using the MD/FEP method. These neutral amino acids were used to clarify the contributions of the effect of solute-solvent size differences to the free energy differences. The calculated values were then compared with the uncorrected values[4b] and the corrected values.[5]

Materials and Methods

Theoretical Background

$$\Delta G=-RT\ln\langle\exp(-\Delta H/RT)\rangle_A \quad (1)$$

The relative free energy is described as Eq. 1, where ΔG ($=G_B-G_A$) is the free energy difference between two states A and B, ΔH ($=H_B-H_A$) is the difference between the Hamiltonians of states A and B estimated for the same set of coordinates and momenta, and $\langle\ \rangle_A$ indicates that an ensemble average is to be taken over the reference state A. In practice, the momentum contribution to the free energy change is assumed to be negligible, and the Hamiltonian is replaced by an empirical potential energy function, U, as in Eq. 2.

$$\Delta G=-RT\ln\langle\exp(-\Delta U/RT)\rangle_A \quad (2)$$

To obtain an accurate value of the ensemble average in Eq. 2, it is inappropriate to have a large difference between states A and B. To calculate the large free energy difference between the two A and B physical states, the change from A to B is divided into a number of windows that are close enough to allow for the use of Eq. 2 within a practical amount of sampling. Since free energy is a function of a state, the required total free energy, ΔG, is calculated by Eq. 3, where N is the total number of windows:

$$\Delta G=G_B-G_A=\sum_{i=1}^{N-1}\Delta G_i \quad (3)$$

We introduced a new variable, λ, to the analytic potential function for the windows such that $U(r,\lambda)$ changes monotonously from U_A to U_B for $0\leqq\lambda\leqq1$, $U(r,\lambda=0)=U_A(r)$ and $U(r,\lambda=1)=U_B(r)$, then ΔG_i is obtained from Eq. 4, where ΔU_i is given as Eq. 5 when $i=0$ and $i=N$, $\lambda_0=0$ and $\lambda_N=1$.

$$\Delta G_i=-RT\ln\langle\exp(-\Delta U_i/RT)\rangle_\lambda \quad (4)$$

$$\Delta U_i=U(r,\lambda_{i+1})-U(r,\lambda_i) \quad (5)$$

λ is slowly changed enough that it can be assumed that the system remains in equilibrium at every step. Thus, the ensemble average in Eq. 4 is replaced by its instantaneous value as Eq. 6, and then Eq. 3 becomes Eq. 7.

* To whom correspondence should be addressed.

$$\Delta G_i = -RT\ln\{\exp(-\Delta U_i/RT)\} = \Delta U_i \quad (6)$$

$$\Delta G = G_B - G_A = \sum_{i=1}^{N-1} \Delta U_i \quad (7)$$

Computational Procedure The differences between solvation free energies of amino acids were determined using the thermodynamic cycle (Chart 1).[6b] In the thermodynamic cycle where the amino acids A and B are either in the gas phase [$A_{(g)}$ and $B_{(g)}$] or in aqueous solution [$A_{(aq)}$ and $B_{(aq)}$], the free energy of solvation of amino acids A and B [ΔG_A and ΔG_B], can be measured. Although the free energy of solvation cannot be directly determined by free energy perturbation (FEP) calculations, the free energy of mutation of amino acid A into B in the gas phase or solution [$\Delta G_{AB(g)}$ or $\Delta G_{AB(aq)}$] can be easily calculated. Because free energy is a function of a state, the sum of the individual ΔG's of a cycle is zero. For the cycle given in Chart 1, Eq. 8 shows the free energy relationship. Rearrangement of Eq. 8 to solve the relative free energy difference of solvation results in Eq. 9, so that the results from computer calculations can be related to experimental values.

$$\Delta G_{AB(g)} + \Delta G_B - \Delta G_A - \Delta G_{AB(aq)} = 0 \quad (8)$$

$$\Delta\Delta G = \Delta G_B - \Delta G_A = \Delta G_{AB(aq)} - \Delta G_{AB(g)} \quad (9)$$

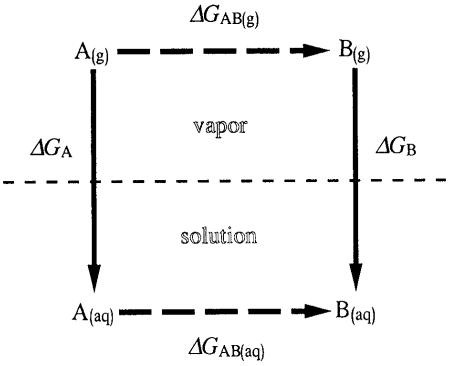

Chart 1

The MD/FEP calculations were performed with AMBER 4.0[7b] on Toshiba SPARC LT AS1000/L10 workstations. For the calculations in the gas phase, *N*-acetyl-*N'*-methylamides of the amino acids in an extended conformation were used as the starting structure. For the calculations in aqueous solution, the molecule was solvated by about seven hundred TIP3P water molecules[8] in a box and the calculations were performed under a periodic boundary condition. The systems were minimized and were then equilibrated for 10 ps using the MD calculation. The MD/FEP calculations were done at constant temperature (298 K) in both states and also at constant pressure (1 atm) in the solution.[9] SHAKE was employed to restrain the bond lengths in order to remove the high frequency motions.[10] This allowed us to use a time step of 1 fs. The non-bonded cutoff for interactions was 9 Å. The non-bonded pair list was updated every 20 time-steps during the calculations. The FEP calculations were performed using the slow-growth procedure in the AMBER program and the free energies were averaged from the forward ($\lambda = 0 \rightarrow 1$) and backward ($\lambda = 1 \rightarrow 0$) directions. During the FEP process, atoms of amino acid A were gradually replaced by atoms of amino acid B. As shown in Chart 2, disappearing atoms were substituted by dummy atoms which do not have non-bonded interactions and appearing atoms were changed from dummy to real ones. The total perturbation time was 80 ps in both directions. The perturbed group was taken to be the whole mutation residue. In AMBER 3.0 used by Kollman and his group in 1987,[6] interactions between an amino acid and water could be included in the calculated free energies but the intra-perturbed group energies of an amino acid could not be calculated.[7a] Thus, only $\Delta G_{AB(aq)}$ was

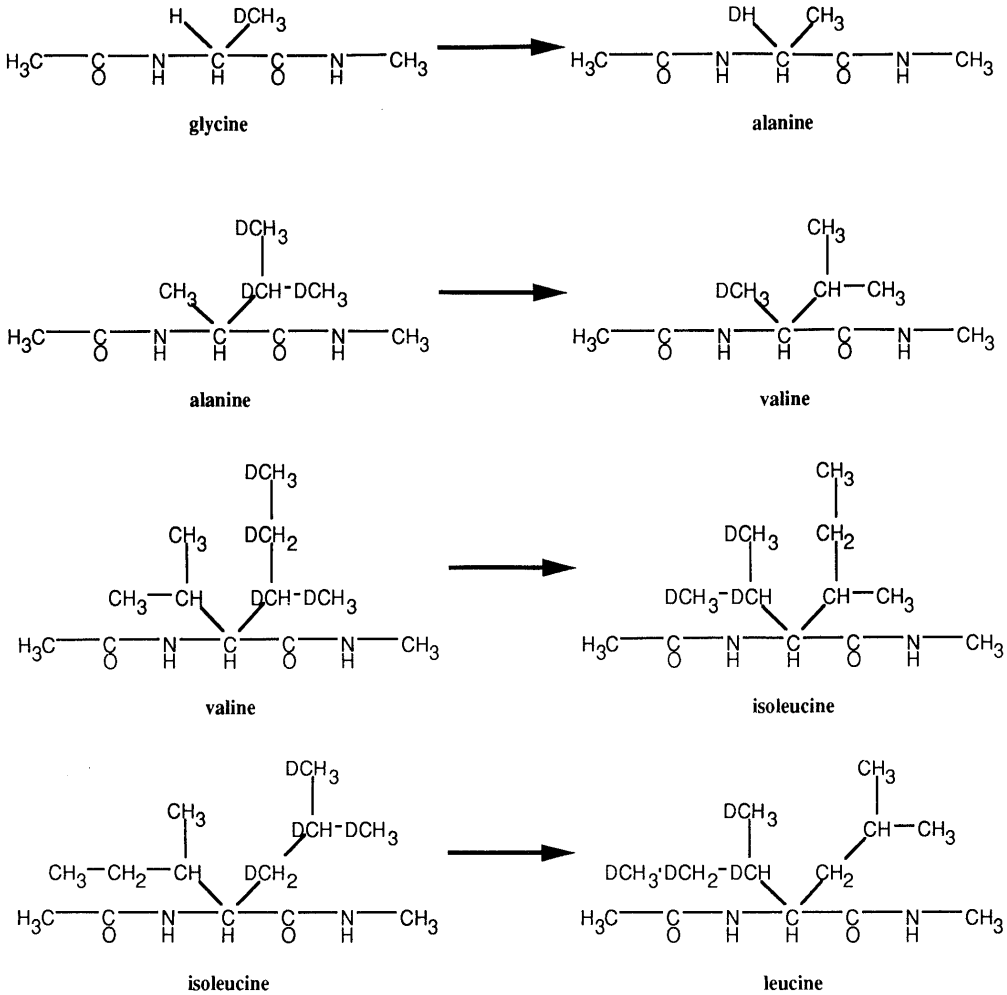

Chart 2

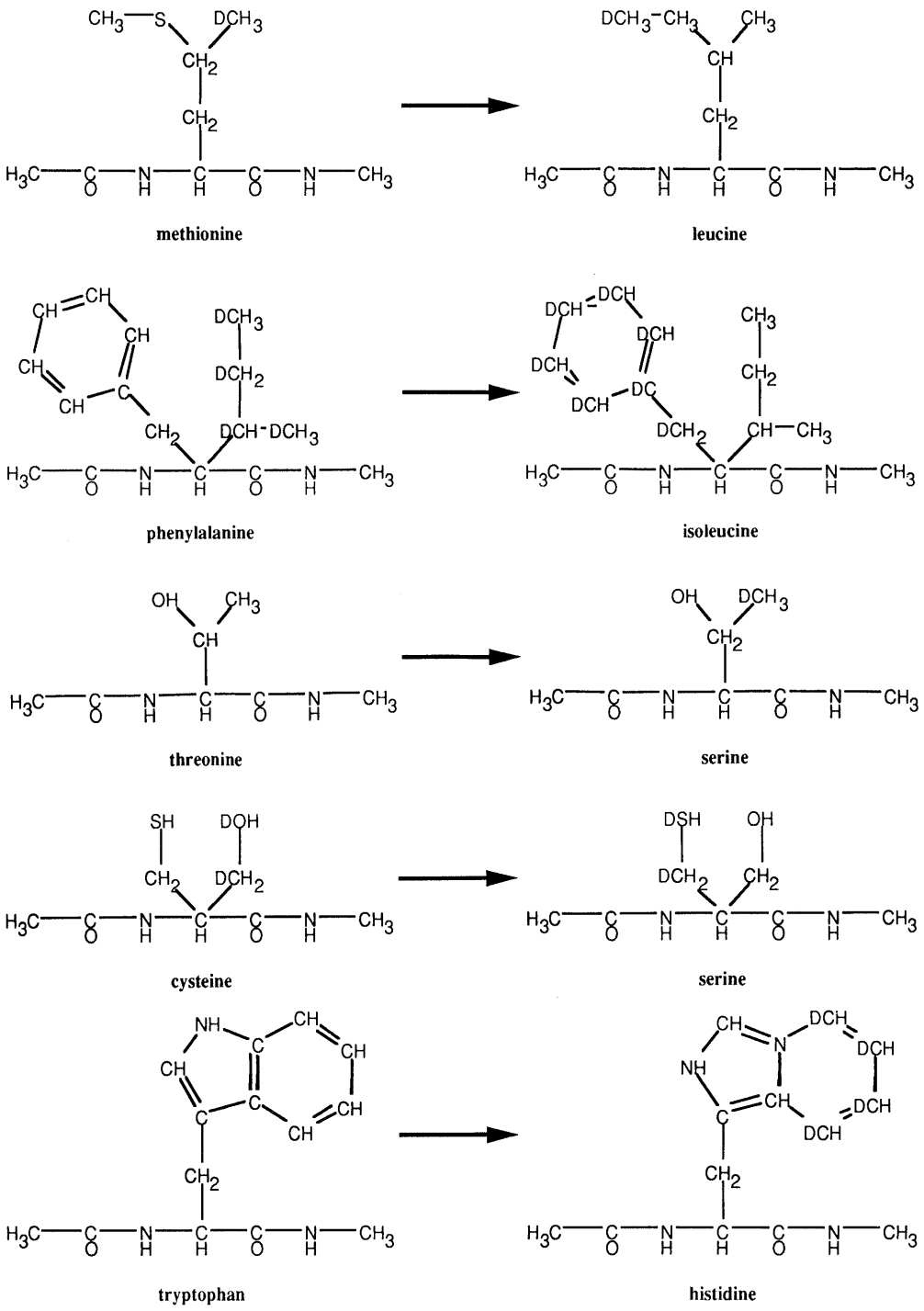

Chart 2.
DCH_n, DH, and DSH: atom group whose name begins with D consists of dummy atoms.

calculated because $\Delta G_{AB(g)}$ was considered zero. In AMBER 4.0, the intra-perturbed group energies of an amino acid can be calculated.[7b] We included the intra-perturbed group non-bond energies in the calculated free energies, and both $\Delta G_{AB(aq)}$ and $\Delta G_{AB(g)}$ were calculated.

Results and Discussion

In Table 1, the results of simulations are compared with the uncorrected[4b] and corrected values.[5] The relations of the calculated values with the corresponding uncorrected and corrected values are shown in Fig. 1.

For the uncorrected values (see Fig. 1a), the correlation coefficient was $r=0.938$ and the simple regression equation (Eq. 10) was developed, where $\Delta\Delta G_{uncorr}$ is the uncorrected value and $\Delta\Delta G_{calc}$ is the calculated value.

$$\Delta\Delta G_{uncorr}=0.823\Delta\Delta G_{calc}-0.235(\pm 0.115) \quad (10)$$

For the corrected values (see Fig. 1b), the correlation coefficient was $r=0.987$ and the simple regression equation (Eq. 11) was developed, where $\Delta\Delta G_{corr}$ is the corrected value.

$$\Delta\Delta G_{corr}=1.042\Delta\Delta G_{calc}-0.172(\pm 0.065) \quad (11)$$

The calculated values were in better agreement with the corrected values than with the uncorrected values. The

slope of Eq. 11 was closer to one than that of Eq. 10, and the intercept of Eq. 11 was closer to zero than that of Eq. 10. This means that the calculated value can be absolutely compared with the corrected value, *i.e.* $\Delta\Delta G_{corr} \fallingdotseq \Delta\Delta G_{calc}$. Our result differs from that obtained by Kollman and his group,[6] because the intra-perturbed group energies of an amino acid are computed in our calculation.

Table 1. Comparison between Calculated and Experimental Free Energy Differences

Process	Free energy differences[a]		
A → B	$\Delta\Delta G_{calc}$[b]	$\Delta\Delta G_{uncorr}$[c]	$\Delta\Delta G_{corr}$[d]
Gly → Ala	1.14	−0.45	0.21
Ala → Val	1.57	0.05	1.44
Val → Ile	1.43	0.16	0.82
Ile → Leu	0.71	0.13	0.31
Met → Leu	3.43	3.76	4.27
Phe → Ile	2.65	2.91	2.71
Thr → Ser	−1.08	−0.18	−0.77
Cys → Ser	−3.76	−3.82	−4.32
Trp → His	−5.74	−4.39	−5.85

a) Free energy differences (in kcal/mol at 298 K) are given by Eq. 9. *b*) $\Delta\Delta G_{calc}$ are calculated free energy differences. *c*) $\Delta\Delta G_{uncorr}$ are experimental free energy differences obtained from Ref. 4*b*. *d*) $\Delta\Delta G_{corr}$ are corrected experimental free energy differences obtained from Ref. 5.

When the free energy difference between amino acids A and B is represented by the expression considering the effect of solute-solvent size differences, $\Delta G = -\boldsymbol{R}T\ln K + \boldsymbol{R}T(1 - V_s/V_v)$, the relation of the corrected value with the uncorrected value is shown as Eq. 12, where V_A, V_B and V_{WAT} are molar volumes for amino acids A, B and water, respectively. The actual free energy difference ($\Delta\Delta G_{corr}$) is larger by $\boldsymbol{R}T(V_A - V_B)/V_{WAT}$ than the value ($\Delta\Delta G_{uncorr}$) obtained using the standard expression, $\Delta G = -\boldsymbol{R}T\ln K$.

$$\begin{aligned}\Delta\Delta G_{corr} &= \Delta G_B - \Delta G_A \\ &= \{-\boldsymbol{R}T\ln K_B + \boldsymbol{R}T(1 - V_B/V_{WAT})\} \\ &\quad -\{\boldsymbol{R}T\ln K_A + \boldsymbol{R}T(1 - V_A/V_{WAT})\} \\ &= \{(-\boldsymbol{R}T\ln K_B) - (-\boldsymbol{R}T\ln K_A)\} + \boldsymbol{R}T(V_A - V_B)/V_{WAT} \\ &= \Delta\Delta G_{uncorr} + \boldsymbol{R}T(V_A - V_B)/V_{WAT} \end{aligned} \quad (12)$$

The volume for the solute molecule changed in our FEP calculations because the atoms of the side chain appeared and disappeared (Chart 2). Therefore, the effect of solute-solvent size differences, $\boldsymbol{R}T(V_A - V_B)/V_{WAT}$, may be exactly estimated.

Contributions to the free energy differences are shown in Table 2. Because of the complexity of the system, the free energy changes in solution were generally less accurate

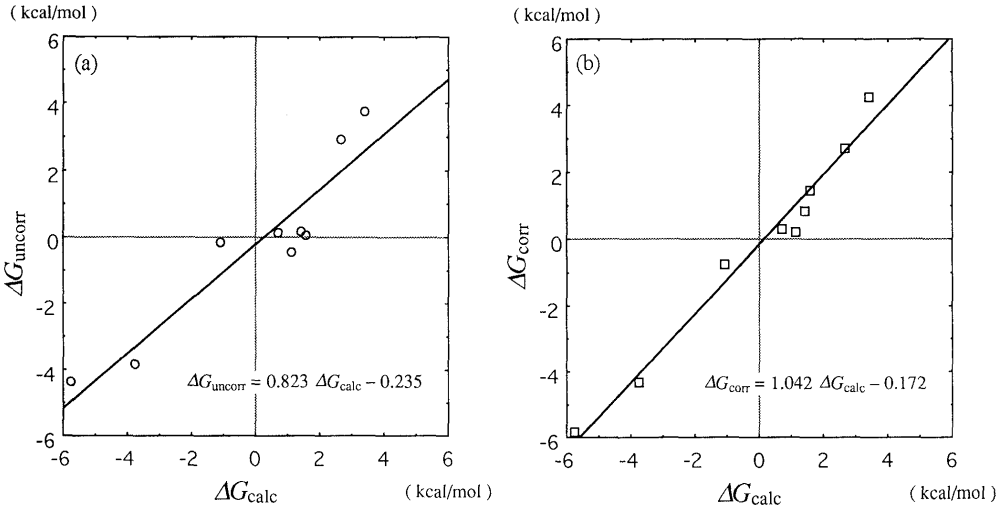

Fig. 1. Correlation between Calculated and Experimental Free Energy Difference between Amino Acids

(a), correlation between calculated values and uncorrected values; (b), correlation between calculated values and corrected values.

Table 2. Free Energy Differences between Amino Acids Obtained by Calculation

Process	$\Delta G_{AB(aq)}$[a]			$\Delta G_{AB(g)}$[a]			$\Delta\Delta G$[a,b]		
A → B	Elec	VdW	Total	Elec	VdW	Total	Elec	VdW	Total
Gly → Ala	1.19±0.00	1.50±0.11	2.69±0.12	1.17±0.00	0.37±0.05	1.54±0.05	0.02±0.00	1.13±0.12	1.14±0.13
Ala → Val	1.10±0.00	5.03±0.47	6.13±0.47	1.35±0.32	3.22±0.08	4.57±0.24	−0.25±0.32	1.82±0.48	1.57±0.53
Val → Ile	−0.69±0.08	2.32±0.53	1.63±0.61	−0.78±0.00	0.98±0.14	0.20±0.14	0.10±0.08	1.34±0.55	1.43±0.63
Ile → Leu	−0.95±0.07	−1.69±0.56	−2.64±0.49	−1.20±0.16	−2.15±0.09	−3.35±0.24	0.25±0.17	0.46±0.57	0.71±0.54
Met → Leu	12.94±0.02	3.92±0.20	16.86±0.18	11.54±0.35	1.90±0.31	13.43±0.05	1.41±0.36	2.03±0.37	3.43±0.19
Phe → Ile	14.15±1.30	−7.46±2.22	6.69±0.92	1.94±0.09	2.10±0.85	4.04±0.94	12.22±1.30	−9.56±2.37	2.65±1.32
Thr → Ser	3.41±0.74	−2.29±0.16	1.12±0.90	3.70±0.47	−1.49±0.02	2.20±0.49	−0.29±0.88	−0.79±0.16	−1.08±1.02
Cys → Ser	29.97±0.24	−3.83±0.26	26.14±0.50	31.86±0.02	−1.96±0.02	29.90±0.04	−1.89±0.24	−1.87±0.26	−3.76±0.50
Trp → His	−3.61±0.38	−5.43±0.34	−9.04±0.04	−1.68±0.02	−1.62±0.08	−3.30±0.10	−1.93±0.38	−3.81±0.35	−5.74±0.11

a) Free energies and free energy differences (in kcal/mol at 298 K) given are the average values and the standard deviation for forward and backward simulations. Elec are electrostatic contributions. VdW are van der Waals contributions. *b*) Free energy differences are given by Eq. 9.

than the free energy changes in the gas phase. Accordingly, a longer perturbation time for the solution is necessary in order to obtain a more accurate free energy difference. Van der Waals contributions to the free energy differences tended to be less accurate than electrostatic contributions. This might be related to the functional form that represents the van der Waals interactions and the amount of sampling performed.[6b] The van der Waals contributions to the free energy differences were roughly correlated with the volume change in the side chain. Free energy differences between aliphatic amino acids came mainly from the van der Waals contributions, not the electrostatic contributions. This is related to a small amount of polarization of the side chain of the aliphatic amino acids.

The free energy perturbation method can reproduce solvation free energies of natural amino acids considering the effect of solute-solvent size differences defined by Sharp *et al.*[5] Furthermore, this suggests that the free energy perturbation method can predict an accurate solvation free energy change when a natural amino acid residue of a bio-active peptide is substituted by an unnatural amino acid residue. Consequently, the free energy perturbation method is a useful tool for the rational design of proteins and bio-active peptides.

References and Notes

1) Kauzmann W., *Adv. Protein Chem.*, **14**, 1 (1959).
2) Yutani K., Ogasahara K., Tsujita T., Sugino Y., *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4441 (1987); Matsumura M., Becktel W. J., Matthews B. W., *Nature* (London), **334**, 406 (1988).
3) Okajima T., Tanizawa K., Yoneya T., Fukui T., *J. Biol. Chem.*, **266**, 11442 (1991).
4) *a*) Tanford C., *J. Am. Chem. Soc.*, **84**, 4240 (1962); Nozaki Y., Tanford C., *J. Biol. Chem.*, **246**, 2211 (1971); *b*) Wolfenden R. V., Cullis P. M., Southgate C. C. F., *Science*, **206**, 575 (1979); Wolfenden R., Andersson L., Cullis P. M., Southgate C. C. B., *Biochemistry*, **20**, 849 (1981); *c*) Fauchere J.-L., Pliska V., *Eur. J. Med. Chem.*, **18**, 369 (1983).
5) Sharp K. A., Nicholls A., Friedman R., Honig B., *Biochemistry*, **30**, 9686 (1991).
6) *a*) Singh U. C., Brown F. K., Bash P. A., Kollman P. A., *J. Am. Chem. Soc.*, **109**, 1607 (1987); *b*) Bash P. A., Singh U. C., Langridge R., Kollman P. A., *Science*, **236**, 564 (1987).
7) *a*) Singh U. C., Weiner P. K., Caldwell J. W., Kollman P. A., "AMBER 3.0," University of California, San Francisco, 1986; *b*) Pearlman D. A., Case D. A., Caldwell J. C., Seibel G. L., Singh U. C., Weiner P., Kollman P. A., "AMBER 4.0," University of California, San Francisco, 1991.
8) Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L., *J. Chem. Phys.*, **79**, 926 (1983).
9) Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R., *J. Chem. Phys.*, **81**, 3684 (1984).
10) Ryckaert J.-P., Ciccotti G., Berendsen H. J. C., *J. Comp. Phys.*, **23**, 327 (1977).