

Interaction-component analysis of urea effect on amino-acid analogs

Yasuhito Karino^{ab} and Nobuyuki Matubayasi^{*ac}

Received Xth XXXXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXXXX 200X

DOI: 10.1039/b000000x

Transfer energetics from pure water to urea-water mixture is examined for a set of amino-acid analog solutes by using molecular dynamics simulation and free-energy calculation. The free energy of transfer from pure-water solvent to 8 M urea-water mixed solvent is calculated for each solute, and the urea-water mixture is shown to be a more favorable solvent than pure water. The correlation of the transfer free energy is then examined against the corresponding changes upon transfer of the average sum of solute-solvent interaction energy and its electrostatic and van der Waals components. A strong correlation is observed against the change in solute-solvent energy, with dominant contribution from the van der Waals component for neutral solutes. The electrostatic component exhibits a weak correlation due to the compensation between the urea and water contributions. The transfer free energy is further decomposed into the contributions from urea and water within an approximate framework of the energy-representation theory of solvation. It is found that urea makes a favorable contribution to the transfer free energy. The water contribution depends on the hydrophobicity/hydrophilicity of solute. Urea and water act cooperatively for hydrophobic solute, and are competitive against each other for hydrophilic solute. The effect of excluded volume is also addressed, and is seen to be minor in the transfer energetics due to the compensation between the urea and water contributions.

1 Introduction

Urea is a widely used denaturant for protein. The mechanism of urea denaturation is determined by the interplay of protein, urea, and water. When urea is added to aqueous solvent, the intermolecular interaction of protein can be operative against both urea and water, leading to two views on urea-induced denaturation.^{1–30} One view emphasizes the direct interaction between protein and urea, and is called “direct” mechanism. The other focuses on the effect of water modified by the presence of urea, and is called “indirect” mechanism.

In the “direct” mechanism, urea denatures a protein since the urea-protein interaction favors the denatured state of the protein. Among the interaction components, the electrostatic contribution operative with charged and polar parts of protein was first suggested as the primary cause of denaturation.² This line of thoughts has been further refined, with emphasis on the hydrogen-bonding interactions of urea with polar residues or peptide backbone.^{6,8,13,17–19,24} The role of apolar or van der Waals interaction is stressed more recently from structural and energetic analyses performed mainly by computer simulation.^{9,10,15,16,20–22,25,26,28–30} Compared to water, urea is found

to solvate amino-acid residues more preferably with possible exceptions of a few charged residues and to have stronger van der Waals interactions with protein and peptide.

The “indirect” mechanism is based on a notion that urea disrupts the water structure.^{31,32} Urea will then mitigate the hydrophobic effect and facilitate the exposure of hydrophobic-core residues to water. Actually, urea was demonstrated to substitute water without disrupting the hydrogen bond network^{11,20,30,33–35}, and the direct mechanism receives more support recently. The solvent-excluded volume was highlighted furthermore,^{12,14} and its change with addition of urea was shown to exert an appreciable effect on protein stability.

In the present work, we analyze the urea-induced effect from the energetic viewpoint. We focus on amino-acid analog solutes through all-atom molecular dynamics simulation and examine their free energies of transfer from pure-water solvent to urea-water mixed solvent. The transfer free energy is the difference between the solvation free energies in pure water and in urea-water mixed solvent. It is then correlated to the changes upon transfer of the electrostatic and van der Waals components in the solute-solvent interaction energy, and the decisive component is identified.

Recent simulation studies showed that the urea effects computed with currently used sets of force field are rather strong compared to experimentally observed effects.^{15,30} In this study, we focus on the correlations between the transfer free energies and the corresponding interaction components

^a Institute for Chemical Research and Elements Strategy Initiative for Catalysts and Batteries, Kyoto University, Uji, Kyoto 611-0011, Japan. Fax: +81-774-38-3074; Tel: +81-774-38-3071; E-mail: nobuyuki@scl.kyoto-u.ac.jp

^b RIKEN Quantitative Biology Center, Kobe, Hyogo 650-0047, Japan.

^c Japan Science and Technology Agency (JST), CREST, Kawaguchi, Saitama 332-0012, Japan.

obtained through numerically exact method for commonly used sets of potential functions. The correlation study is done since the solute-solvent interaction energy is more straightforward for interpretation and prediction than the free energy. It is a rule of thumb, for example, that the electrostatic interaction with solvent is more favorable for a more polar solute molecule and that the van der Waals interaction is stronger with a larger solute. The transfer free energy reflects the effects of all the interaction components in nonlinear manner, in contrast. The correlation is thus to be established for more intuitive treatment of transfer energetics.

When two species forms a mixed solvent, it is natural and common to separate the effects of the two solvent species. The notion of “direct” and “indirect” mechanisms can also belong to this kind of separation. The separated contributions of urea and water are not observable in general, however, and there needs a model of solvation to conduct the separation. To do so, we resort to the method of energy representation (ER),^{36–38} a density-functional method for evaluating the solvation free energy. Among a variety of approximate free energy methods,^{39–47} the ER method is unique in compromising the accuracy, the efficiency, and the range of applicability.^{48–53} In a previous paper, it is seen for amino-acid side-chain analog solutes in pure water that the solvation free energy calculated by the ER method deviates from the experimental value and the numerically exact value by less than 1 kcal/mol.⁵⁴ All-atom, free-energy analysis was further shown to be feasible for protein molecule with a few hundred residues.^{51–53} The free-energy functional used in the ER method is formally expressed as a sum of contributions from urea and water, and allows the decomposition into the effects of urea and water. We discuss the cooperation and competition of the urea and water effects within the framework of the ER method.

When a solute is introduced into solvent, some solvent molecules need to be removed from the region into which the solute is to be inserted. The excluded-volume effect denotes the free-energy penalty corresponding to this removal. It is part of the free-energy functional, and can be treated as another, repulsive interaction component, in addition to the electrostatic and van der Waals components. We also examine the correlation of the transfer free energy with the change upon transfer of the excluded-volume effect.

2 Methods

Pure water and 8 M ($M = \text{mol/dm}^3$) mixture of urea with water were adopted as solvent systems. The peptide-backbone analog and all the amino-acid side-chain analogs except for Gly and Pro were examined as solutes. Cyclic glycylglycine peptide (diketopiperazine) was adopted as a model compound for the backbone,^{17,18} and the halves of the calculated results were taken as the values for the peptide-backbone analog. The

chemical species identified as the side-chain analogs are listed in Table 1.^{54–57} The Arg, Lys, Asp, and Glu analogs have net charges, and the His analog is treated as a neutral species with a proton attached to the δ position.^{54,56,57} For each system, the total number of solvent molecules in the unit cell of molecular dynamics (MD) simulation was set to 1000. In the 8 M mixture of urea, 184 urea molecules and 816 water molecules were located in the unit cell, respectively.⁵⁸ No counterions were placed in the non-neutral systems, and the self-energy term of Ewald method was added to the average sum of solute-solvent electrostatic interaction energy and the solvation free energy for correcting the effect of system non-neutrality.^{59–62}

All the MD simulations were carried out with GROMACS version 4.5.4.⁶³ The Amber99sb force field⁶⁴ was adopted for the solute molecules. As described in Ref.⁵⁶, the force-field parameters were developed not for side-chain analogs but for amino-acid residues, and the force field for the amino-acid side-chain analogs was prepared in the following scheme. The side-chain analog was built by replacing the α -carbon of the corresponding amino-acid residue by a hydrogen atom. The parameters related to the added hydrogen atom were taken to be the same as those for the hydrogen which is initially connected to the β -carbon. The partial charge of β -carbon was modified to make the overall molecule (ion) neutral for the neutral analogs and ionized to the charge of +1 or –1 for the ionic analogs. The original TIP3P model was used for water.^{65,66} The Amber99sb force field⁶⁷ and the Kirkwood-Buff force field (KBFF)⁶⁸ were employed for urea. The partial charges of Amber urea were calculated with Gaussian 09 at the MP2 level and the aug-cc-pVDZ basis set,⁶⁹ and were 0.9518 (C), –0.6446 (O), –0.9430 (N), and 0.3947 (H) in this study. The KBFF was build on the basis of GROMOS96 force field.^{68,70} Although a mixed use of force fields are typically not recommended, the KBFF urea was examined since its mixing properties with water correspond well to experimental. The electrostatic interaction was handled by the particle-mesh Ewald (PME) method⁷¹ with a real-space cutoff of 13.5 Å, a spline order of 4, inverse decay length of 0.256 \AA^{-1} , and a reciprocal-space mesh size of 32 for each of the x , y , and z directions. The Lennard-Jones (LJ) interaction was truncated by applying the switching function with the switching range of 10–12 Å.⁷² The truncation was done on atom-atom basis both for the real-space part of PME interaction and the LJ interaction. The long-range correction of LJ interaction was not included during the MD run. It was applied to the solvation free energy and the average sum of solute-solvent interaction only after the MD simulations were finished. The correction was done with the standard scheme^{56,57,73} by supposing that the radial distribution functions are unity beyond the switching region.⁷⁴ The equation of motion was integrated with the leap-frog algorithm at a time step of 2 fs. For the amino-acid analogs and urea, all the bonds involving the hy-

Table 1 Correspondence between the amino-acid side chains and their analogs.^a

amino acid	analog solute	amino acid	analog solute	amino acid	analog solute
Ala	methane	Cys	methanethiol	Lys	<i>n</i> -butylamine
Val	propane	Ser	methanol	Arg	<i>n</i> -propylguanidine
Leu	<i>iso</i> -butane	Thr	ethanol	Asp	acetic acid
Ile	<i>n</i> -butane	Asn	acetamide	Glu	propionic acid
Met	methyl ethyl sulfide	Gln	propionamide		
Phe	toluene	His	4-methylimidazole		
		Tyr	<i>p</i> -cresol		
		Trp	3-methylindole		

^a Within the table, the column of Ala-Phe corresponds to nonpolar and neutral analogs, and the column of Cys-Trp is for polar and neutral analogs. The column of Lys-Glu refers to ionic analogs; Lys and Arg are cations and Asp and Glu are anions.

drogen atom were fixed with LINCS, and the other degrees of freedom such as bonding between heavy atoms, bending, torsional, and improper were flexible.^{75,76} The water molecules were kept rigid with SETTLE.⁷⁷ The simulations were carried out in the canonical ensemble with a cubic unit cell. The periodic boundary condition was employed with the minimum image convention. The temperature was set at 300 K using a stochastic dynamics integrator with a coupling constant of 1 ps.⁷⁸ The box size was set at 31.1 Å when the solvent is pure water (corresponding to the density of 1.00 g/cm³) and at 33.7 Å when the solvent is the 8 M mixture of urea with water (corresponding to the density of 1.12 g/cm³).⁷⁹

The solvation free energies in pure water and in urea-water mixture were computed by the Bennett acceptance ratio method.⁸⁰ To conduct this computation, a set of intermediate states connecting the initial state without the solute and the final state with the solute at full coupling need to be introduced. The initial, intermediate, and final states are identified by the coupling parameters λ_{elec} and λ_{LJ} ($0 \leq \lambda_{\text{elec}} \leq 1$, $0 \leq \lambda_{\text{LJ}} \leq 1$) to control the electrostatic and LJ terms of solute-solvent interaction, respectively; the solute is the amino-acid analog of interest and the solvent is water and also urea in the case of 8 M mixture. The soft-core potential $V(r)$ was employed for the site-site interaction potential as a function of λ_{elec} and λ_{LJ} in the form of^{57,81}

$$V(r) = \lambda_{\text{elec}} \frac{q_u q_v}{r} + 4\epsilon \lambda_{\text{LJ}} \left(\frac{\sigma^{12}}{[\alpha \sigma^6 (1 - \lambda_{\text{LJ}}) + r^6]^2} - \frac{\sigma^6}{\alpha \sigma^6 (1 - \lambda_{\text{LJ}}) + r^6} \right), \quad (1)$$

where r is the distance between a pair of interactions sites of solute and solvent, ϵ and σ are the LJ energy and length parameters with the Lorentz-Berthelot combination rule, respectively,⁸² and q_u and q_v are the charges on the solute and solvent sites, respectively. Eqn (1) involves the electrostatic

term with linear dependence on λ_{elec} and the LJ term with nonlinear dependence on λ_{LJ} . The parameter α defines the “softness” of the λ_{LJ} -dependent potential and was set to 0.5. The fully decoupled state ($\lambda_{\text{elec}} = 0$, $\lambda_{\text{LJ}} = 0$) corresponds to the reference solvent system without the analog solute, and the fully coupled state ($\lambda_{\text{elec}} = 1$, $\lambda_{\text{LJ}} = 1$) is the solution system of interest. In the coupling process, λ_{LJ} changes from 0 to 1 at $\lambda_{\text{elec}} = 0$ and then λ_{elec} changes with $\lambda_{\text{LJ}} = 1$. The decoupling process corresponds to the reverse variation of λ_{elec} and λ_{LJ} . The simulation was conducted at $\lambda_{\text{elec}} = i/5$ ($i = 0, 1, \dots, 5$) and $\lambda_{\text{LJ}} = i/10$ ($i = 0, 1, \dots, 10$). The number of initial, intermediate, and final states prepared was thus 16. At each λ_{elec} and λ_{LJ} , the equilibration and production runs were done for 100 ps when the solvent is pure water, and they were conducted over 0.5 and 1.0 ns, respectively, when the solvent is the urea-water mixture. Both of the coupling and decoupling calculations were performed three times, and the six values obtained were averaged to determine the solvation free energy for each pair of solute and solvent.

The solvation free energies were also calculated approximately by using the method of energy representation. In this method, a set of energy distribution functions are obtained from simulation and are substituted into a functional for the solvation free energy. For each solute, the simulation was conducted for two condensed-phase systems. One is called “solution” system, and is the state of $\lambda_{\text{elec}} = \lambda_{\text{LJ}} = 1$ in the coupling-parameter method of free energy introduced with eqn (1). It contains the solute at full coupling of solute-solvent interaction, and the distribution function of solute-solvent pair energy is to be obtained. The other is called “reference solvent” system, and corresponds to the state of $\lambda_{\text{elec}} = \lambda_{\text{LJ}} = 0$. Its simulation is done to determine the density of states of solute-solvent pair interaction and the solvent-solvent pair correlation by inserting the solute as a test particle. To prepare a set of intramolecular configurations of the solute for the test-particle insertion, the single-molecule simulation for isolated

solute was also carried out in vacuum over 200 ns and the configuration was saved every 100 fs. See previous papers for methodological details and the explicit form of the functional for the solvation free energy.^{37,50} When the solvent is pure water, the simulation length was 2 ns for the solution systems and 20 ps for the reference solvent systems. For the urea mixture, the simulation length was extended to 20 ns for the solution and to 200 ps for the reference solvent. The instantaneous configuration (snapshot) was sampled every 10 fs for the solution. 2×10^5 and 2×10^6 snapshots in total were then used for the analyses of the solution systems without and with urea, respectively. The snapshot was sampled every 100 fs for the reference solvent system. The test-particle insertion of the solute was carried out 1000 times per reference-solvent configuration sampled, leading to the generation of 2×10^5 and 2×10^6 solute-solvent configurations in total for pure water and urea-water mixture, respectively.

The MD lengths of the solution systems described above are rather long. In the energy-representation method, it is typically fine to perform the solution MD for ~ 100 ps and ~ 1 ns, respectively, when the solvent is pure water and urea-water mixture.⁵⁴ The MD was prolonged in order to obtain the average sum of solute-solvent interaction energy and its electrostatic and van der Waals components to good precision. In Section 3, they will be discussed in detail and the presented values are the ones calculated from the solution MD over 2 ns in the pure-water solvent and over 20 ns in the urea-water mixed solvent. When the method of energy representation is employed, in fact, the solvation free energy converges faster than the average solute-solvent energy and its components due to the variational principle. The latter are also a focus of the present work, though, and a long MD was conducted for their convergence.^{50,54}

3 Results and Discussion

The focus of the present work is the energetics of transfer from pure-water solvent to urea-water mixed solvent. In Subsection 3.1, we examine the correlations of the transfer free energy to the corresponding changes in the solute-solvent interaction energy and its electrostatic and van der Waals components. Numerically exact results are presented, and the strong correlation between the free energy and the van der Waals component is pointed out for neutral species. In Subsection 3.2, we investigate the “direct” and “indirect” effects of urea by decomposing the transfer free energy into the urea and water contributions. In fact, the distinct contributions from urea and water are not necessarily observable, and a model of solvation is required for the decomposition. We employ the approximate functional in the method of energy representation as the basis of the decomposition, and show that the urea contribution is favorable for all the species examined. We also analyze the

effect of excluded volume. This effect may be significantly affected by the transfer from pure water to urea-water mixture, given that urea is a few larger in size than water. The excluded-volume effect is a useful but non-observable part of the solvation free energy. We discuss the correlation between the transfer free energy and the corresponding change in the excluded-volume effect within the framework of the energy-representation method.

3.1 Correlations of the transfer free energy to the solute-solvent interaction energy and its electrostatic and van der Waals components

We first demonstrate the performance of the force field employed in the present work. Fig. 1 shows the solvation free energies $\Delta\mu^0$ of the neutral side-chain analogs in pure water, for which experimental data are available.⁵⁵ To assess the accuracy of the computational values from the Bennett acceptance ratio (BAR) method, we introduce the mean absolute deviation (MAD) by

$$\text{MAD} = \frac{1}{n} \sum_i |\Delta\mu_{i,1} - \Delta\mu_{i,2}|, \quad (2)$$

where $\Delta\mu_{i,1}$ and $\Delta\mu_{i,2}$ are the solvation free energies of the i -th solute in the first and second sets of the data, respectively, and n is the number of solutes examined. The MAD between the BAR and experimental values is 1.3 kcal/mol. It agrees well with the literature value of Shirts et al. (1.2 kcal/mol).⁵⁶ The MAD is 0.5 kcal/mol between the BAR results by us and Shirts et al. This deviation is mainly caused by the difference in the (average) density. The density of solvent was set at 1.0 g/cm³ in this study, and its average was 0.986 g/cm³ in Ref.⁵⁶ It should be noted that the solvation free energy changes by ~ 0.5 kcal/mol even with density variation of 0.01 g/cm³ for the case of the Ile analog.⁸³

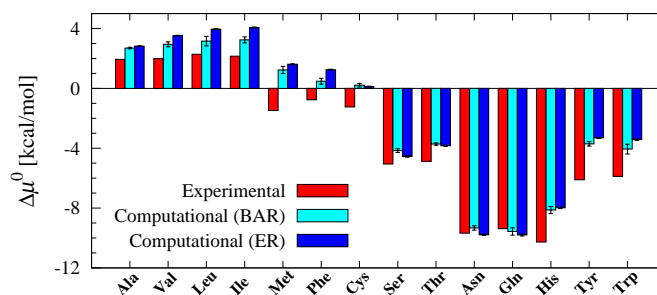


Fig. 1 Solvation free energy $\Delta\mu^0$ of amino-acid analog. BAR refers to the numerically exact value computed by the Bennett acceptance ratio method, and ER is for the approximate value by the energy-representation method. The error bar is expressed at 95% confidence level (twice the standard error).

We then examine the transfer free energy from pure water to urea-water mixture. The solvation free energy in 8 M ($M = \text{mol}/\text{dm}^3$) urea mixture is also calculated by the BAR method, and its difference from the value in pure-water solvent is the transfer free energy. Fig. 2 shows the transfer free energy ($\Delta\mu^u - \Delta\mu^0$), where the superscript u denotes the 8 M urea-water mixed solvent and the superscript 0 denotes the pure-water solvent. The transfer free energies are negative except for the Asp and Glu analogs with KBFF urea. According to Fig. 2, ($\Delta\mu^u - \Delta\mu^0$) is characterized neither by whether the solute is hydrophilic or hydrophobic nor by whether the solute is neutral or ionic. It correlates roughly with the solute size. This tendency agrees qualitatively with structural insights with contact coefficient.²⁰ In particular, ($\Delta\mu^u - \Delta\mu^0$) and the contact coefficient are both enhanced in magnitude for large analogs such as Phe and Trp. The correlation with the solute size was also observed in a previous computational study on polypeptides using the KBFF model of urea.³⁰ The free energy of transfer per amino-acid residue was obtained in Ref³⁰, and is in agreement with the corresponding value estimated from Fig. 2 as the sum of ($\Delta\mu^u - \Delta\mu^0$) of the backbone analog and the side-chain analog. The comparison to experimental value is presented in Appendix A. The computational values of transfer free energy are well correlated to the experimental. As pointed out in Refs.¹⁵ and³⁰, though, Fig. 2 and Appendix A indicate the force-field dependence and the overestimation by computation. In the following, we conduct a correlation study of transfer energetics to extract solvation features which are rather insensitive to the fine parameterization of force field.

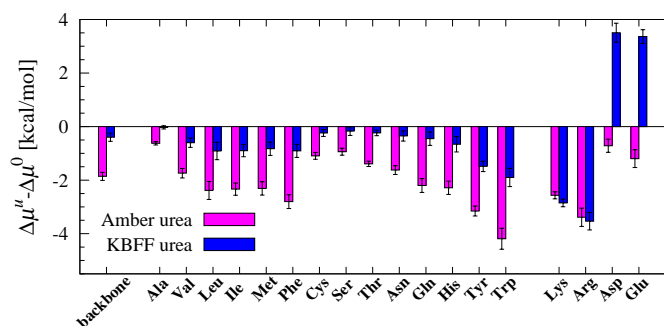


Fig. 2 Transfer free energy from pure water to 8 M urea-water mixture. The error bar is expressed at 95% confidence level (twice the standard error).

The intermolecular interaction potential v between solute and solvent is typically expressed as a sum of electrostatic and van der Waals components, as adopted in the present work. Correspondingly, the ensemble average $\langle v \rangle$ of the sum of solute-solvent interaction energy can be decomposed linearly into the electrostatic and van der Waals components.

The situation is more complicated for the solvation free energy $\Delta\mu$. It reflects the effects of all the interaction components in nonlinear manner, and the decomposition into the electrostatic and van der Waals components is possible by adopting an approximate framework or by introducing an intermediate state of solute insertion.^{39–42,45,50,84} Actually, although $\langle v \rangle$ and its components are not observable, they are more straightforward for interpretation and prediction than $\Delta\mu$.^{85–89} It is thus useful to investigate the correlation between the observable $\Delta\mu$ and the non-observable $\langle v \rangle$ and to develop physical arguments in terms of $\langle v \rangle$ and its components.

The focus of the present work is the change in thermodynamic quantity upon transfer from pure-water solvent to 8 M urea-water mixed solvent. The transfer free energy is the difference in the solvation free energy and is denoted as ($\Delta\mu^u - \Delta\mu^0$), where the superscripts u and 0 mean the ensemble averages in urea-water mixed solvent and pure-water solvent, respectively. Similarly, the transfer quantity can be defined for the average sum of solute-solvent interaction energy as ($\langle v \rangle^u - \langle v \rangle^0$), and the transfer values are introduced in parallel for the electrostatic and van der Waals components. Fig. 3 is the correlation plot of ($\Delta\mu^u - \Delta\mu^0$) with ($\langle v \rangle^u - \langle v \rangle^0$) and its electrostatic and van der Waals components. It is observed for both the urea models that ($\Delta\mu^u - \Delta\mu^0$) correlates well with ($\langle v \rangle^u - \langle v \rangle^0$). For Amber urea, the variation of ($\langle v \rangle^u - \langle v \rangle^0$) is governed by its van der Waals component. The electrostatic component varies within 2 kcal/mol, and in particular, within 1 kcal/mol for the neutral analogs. The correlation of ($\Delta\mu^u - \Delta\mu^0$) is absent against the electrostatic component of ($\langle v \rangle^u - \langle v \rangle^0$), and is operative only against the van der Waals component. For KBFF urea, the correlations are observed similarly. A difference is there for the ionic analogs. Although the van der Waals component of ($\langle v \rangle^u - \langle v \rangle^0$) is negative for all the ionic species, the electrostatic component in the case of KBFF urea has the same sign as ($\Delta\mu^u - \Delta\mu^0$) and is negative for the cations and positive for the anions. It should be noted that Fig. 3 does not represent the causality. Both of ($\Delta\mu^u - \Delta\mu^0$) and ($\langle v \rangle^u - \langle v \rangle^0$) change due to the variation of the amino-acid analog solute. The correlation plot of Fig. 3 still shows that the free energy of transfer (of neutral species) is expected (not causally deduced) to be more favorable when the solute-solvent interaction is changed to a one with stronger van der Waals interaction. It is observed in Fig. 3, furthermore, that ($\langle v \rangle^u - \langle v \rangle^0$) is larger in magnitude than ($\Delta\mu^u - \Delta\mu^0$). This means, in agreement with previous results for hydrophobic solutes,^{15,16} that the transfer value of solute-solvent entropy in the notation of Refs¹⁵ and¹⁶ is negative and that the free-energy penalty due to solvent reorganization is larger in urea-water mixture than in pure water.

Each of the electrostatic and van der Waals components of ($\langle v \rangle^u - \langle v \rangle^0$) can be further separated into the urea and water contributions. Fig. 4 shows the separated contributions

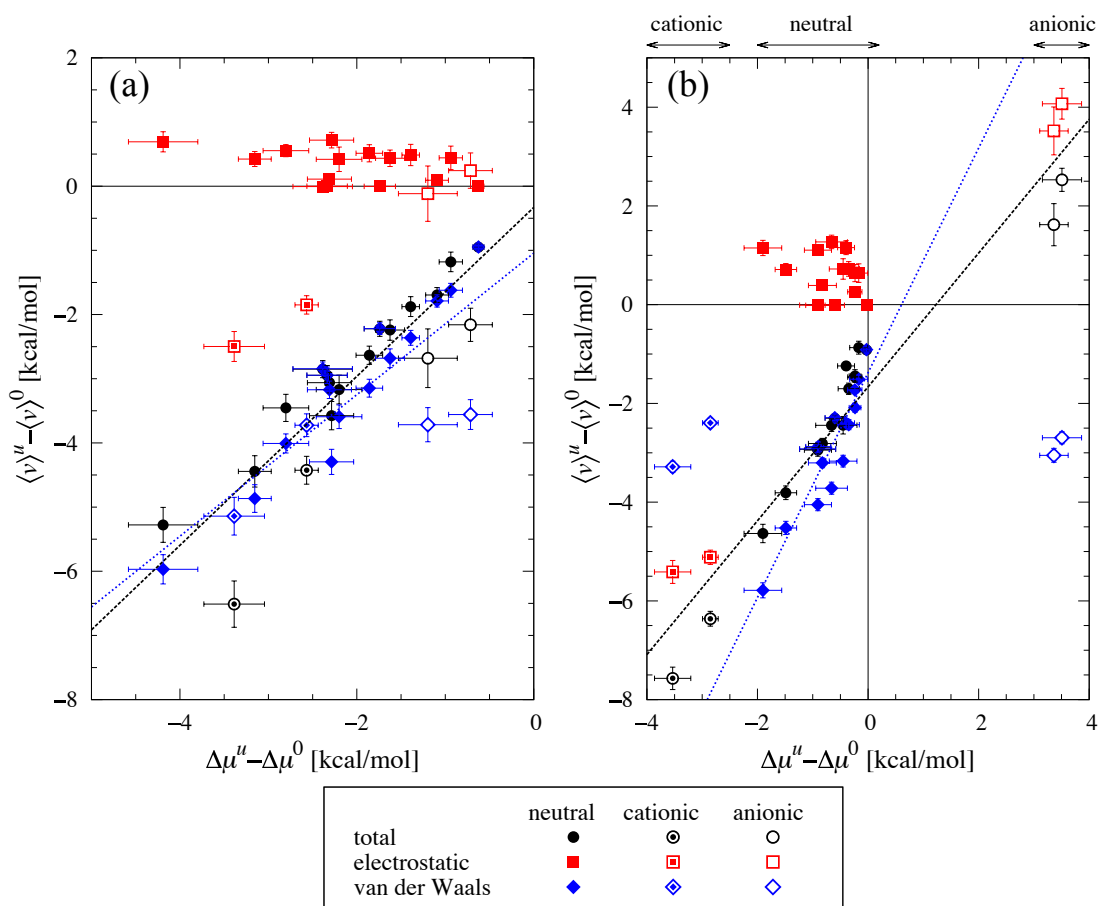


Fig. 3 Correlation plots of transfer quantities for (a) Amber urea and (b) KBFF urea. The computational values are numerically exact ones; the free energy is obtained with the Bennett acceptance ratio method. The error bar is expressed at 95% confidence level (twice the standard error). In (a), the linear regression was conducted with the least square fit for the total $\langle v \rangle^u - \langle v \rangle^0$ and its van der Waals component against $(\Delta\mu^u - \Delta\mu^0)$ over all the solute species. The regression for the total $\langle v \rangle^u - \langle v \rangle^0$ has a slope of 1.3 with the correlation coefficient of 0.90, and the regression for the van der Waals component has a slope of 1.1 with the correlation coefficient of 0.83. When the regression for the van der Waals component is restricted to the neutral species, the correlation coefficient improves to 0.95 with a slope of 1.4. In (b), the linear regression was carried out for the total $\langle v \rangle^u - \langle v \rangle^0$ over all the solute species and for the van der Waals component only over the neutral species. The slope is 1.4 for the total $\langle v \rangle^u - \langle v \rangle^0$ with the correlation coefficient of 0.96, and the slope is 2.3 for the van der Waals component with the correlation coefficient of 0.93.

from each solvent species. The compensation is evidently seen in Fig. 4a and 4c for the electrostatic component between the separated contributions from urea and water. The electrostatic interaction between solute and water is well replaced by that between solute and urea. The exception is the ionic analogs with KBFF urea. The compensation is only partial, and the electrostatic component of $\langle v \rangle^u - \langle v \rangle^0$ changes by ~ 5 kcal/mol with addition of 8 M urea. The energetic behavior in Fig. 4a and 4c is consistent with a structural insight that the total amount of hydrogen bond between peptide and solvent molecules stays almost constant upon mixing of urea.^{11,20,30,33–35} For the van der Waals component,

the general tendency of the separated contributions from urea and water is similar to the case for the electrostatic component. The difference is observed for the balance between the urea and water contributions. Although the van der Waals interaction between solute and water is less favorable in the urea-water mixture than in pure water, that between solute and urea overturns the stability loss. Fig. 3 and 4 thus support the idea that the amino-acid analog is stabilized with addition of urea through direct, van der Waals interaction with urea.^{9,10,15,16,20–22,25,26,28–30} It is to be further noted, in agreement with previous computational results,^{9,10,15,16,20,30} that the van der Waals component $(\langle v \rangle_{\text{vdW}}^u - \langle v \rangle_{\text{vdW}}^0)$ correlates

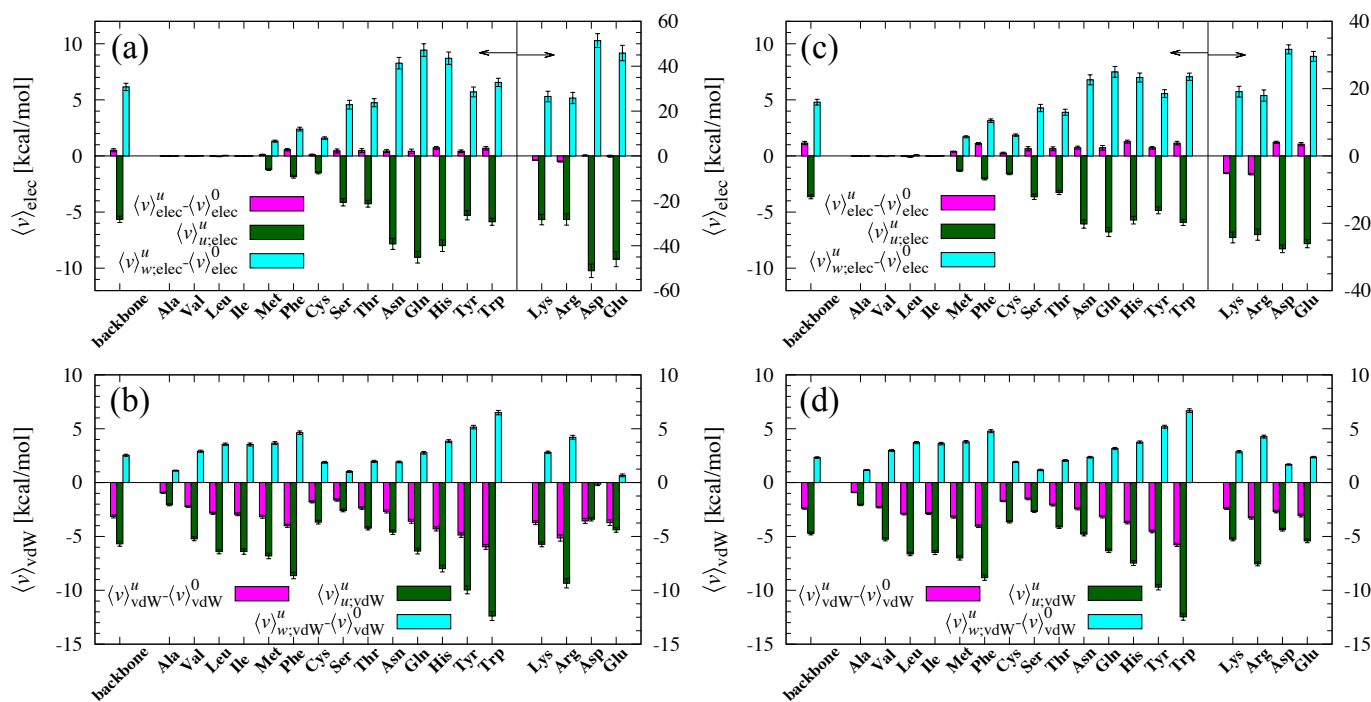


Fig. 4 Urea and water contributions to the electrostatic and van der Waals components of ($\langle v \rangle^u - \langle v \rangle^0$): (a) electrostatic component with Amber urea, (b) van der Waals component with Amber urea, (c) electrostatic component with KBFF urea, and (d) van der Waals component with KBFF urea. The error bar is expressed at 95% confidence level (twice the standard error). The superscripts u and 0 denote the urea-water mixed solvent and the pure-water solvent, respectively, and the subscripts elec and vdW mean the electrostatic and van der Waals components, respectively. The subscripts u and w refer to the urea and water contributions in the mixed solvent, respectively. For example, $\langle v \rangle_{\text{elec}}^u$ stands for the electrostatic component of solute-solvent interaction in the urea-water mixed solvent, and $\langle v \rangle_{u;\text{elec}}^u$ is the urea contribution to it. The $\langle v \rangle_{\text{elec}}^u$ values for Ala-Ile in (a) and (c) are too small in magnitude to be seen within the resolutions of the figures.

with the solute size.

3.2 Urea and water effects in approximate model of solvation

In this subsection, we analyze the separated contributions from urea and water to the transfer free energy. The decomposition into the contributions from respective solvent species is often useful for understanding and predicting a mixed-solvent effect. As pointed out at the beginning of Section 3, though, the distinct contribution from each solvent species is not necessarily observable. A model of solvation needs to be introduced for the decomposition. In the present work, the decomposition is conducted on the basis of the approximate functional for the solvation free energy in the method of energy representation.^{37,49,50} In the latter part of this subsection, furthermore, we discuss the effect of excluded volume. It is a major part of repulsive interaction with the solvent and can be a driving force for structure formation of large molecule such as protein.^{90,91} The excluded-volume effect is a non-observable part of the solvation free energy, either, and its treatment re-

quires a model of solvation. We examine the excluded-volume effect within the framework of the approximate functional in the energy-representation method.

We first demonstrate the performance of the approximate functional for the solvation free energy in the energy-representation (ER) method. The solvation free energy $\Delta\mu^0$ in pure water is compared in Fig. 1 among the approximate value from ER, the numerically exact value from the Bennett acceptance ratio (BAR) method, and the experimental value. At the beginning of Subsection 3.1, the performance of force field was examined in terms of the mean absolute deviation (MAD) introduced by eqn (2). MAD is 0.4 kcal/mol between ER and BAR and is 1.6 kcal/mol between ER and experimental. The MAD value between ER and BAR is a measure of accuracy of the approximate functional in ER. The force field is AMBER in the present study, and the MAD is comparable to that for OPLS-AA examined in a previous work.⁵⁴ When estimated against the experimental, the MAD of ER is comparable to that of BAR shown in Subsection 3.1. The ER method can thus be as useful as the exact method for practical purpose.

Fig. 5 shows the correlation of the transfer free energies calculated by the ER and BAR methods. The linearity is evident. In Appendix B, the correlation analysis is conducted for the transfer free energy approximately obtained from the ER method, and is seen to provide the same insights as those obtained in Subsection 3.1 from the BAR method. When the solute is a protein, the free-energy computation is severely demanding in the exact method, while it is readily feasible with ER.^{51–53} In subsequent paper, we carry out the correlation analysis for protein by employing the ER method.

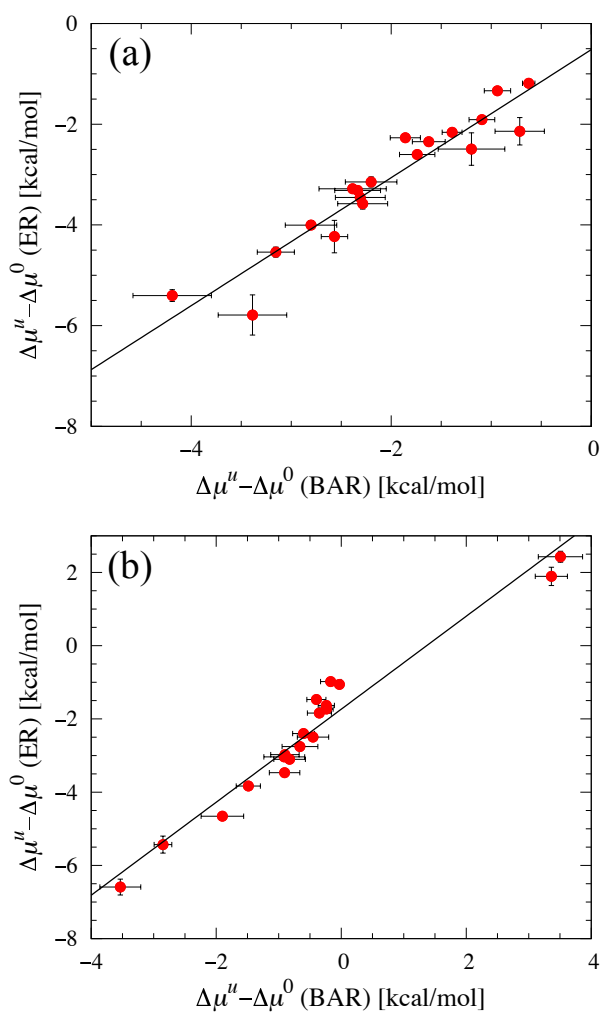


Fig. 5 Transfer free energy ($\Delta\mu^u - \Delta\mu^0$) calculated by the ER method plotted against that from the BAR method for (a) Amber urea and (b) KBFF urea. The error bar is expressed at 95% confidence level (twice the standard error), and is smaller than the size of data symbol when it is not shown. The slope and correlation coefficient are 1.3 and 0.95, respectively, for Amber urea, and they are 1.3 and 0.98, respectively, for KBFF urea.

We now turn to the decomposition of the transfer free en-

ergy into the urea and water contributions. When the mixed solvent consists of urea and water, the solvation free energy $\Delta\mu^u$ in it is written with the ER functional as

$$\Delta\mu^u = \Delta\mu_u^u + \Delta\mu_w^u, \quad (3)$$

where $\Delta\mu_u^u$ and $\Delta\mu_w^u$ are the urea and water contributions, respectively.^{61,74} The first and second terms of eqn (3) may be called “direct” and “indirect” effects of urea. It should be noted that $\Delta\mu_u^u$ and $\Delta\mu_w^u$ are expressed in terms of the correlation functions between the solute and urea and between the solute and water, respectively. Since the correlation functions themselves are affected both by the solute-urea and solute-water interactions, in turn, the decomposition in the form of eqn (3) is formally possible only within the approximate expression of free energy. In this sense, the “direct” and “indirect” effects introduced with respect to eqn (3) are specific to the ER model of solvation. A different set of definitions is possible for the “direct” and “indirect” effects, indeed.³⁰ Actually, the decomposition is done for the purpose of interpretation, and we present consistent and physically reasonable insights in the following.

Fig. 6 shows the transfer free energy ($\Delta\mu^u - \Delta\mu^0$) and the urea and water contributions evaluated with the approximate functional in the ER method.^{61,74} The overall ($\Delta\mu^u - \Delta\mu^0$) is negative (with exceptions of the anions with KBFF urea); with addition of urea, the stability in solution increases for amino-acid analog solute. When the transfer free energy is decomposed into the urea and water contributions, it is seen that the urea contribution $\Delta\mu_u^u$ is favorable (negative in the free-energy value). The “direct” effect introduced with respect to eqn (3) acts to stabilize the solute. The water contribution ($\Delta\mu_w^u - \Delta\mu^0$) depends upon the type of solute-water interaction, on the other hand. It is favorable for the hydrophobic analogs, and is unfavorable for the hydrophilic. Thus, the “direct” and “indirect” effects are cooperative with each other when the solute is hydrophobic, and they are competitive when the solute is hydrophilic. When the solvent is changed from pure water to urea-water mixture, the hydrophobicity diminishes with the water-density reduction and the hydrogen bonding of a solute with water is (partially) replaced by that with urea. These correspond to the solute-type dependence of the sign of ($\Delta\mu_w^u - \Delta\mu^0$).

Finally, we discuss the effect of excluded volume. The excluded volume is the domain of solute-solvent configuration in which the solute and solvent molecules overlap with each other and their interaction energy is prohibitively large. In the ER method, the contribution $\Delta\mu_i$ (i denotes either urea or water) from each solvent species is written in the integral form

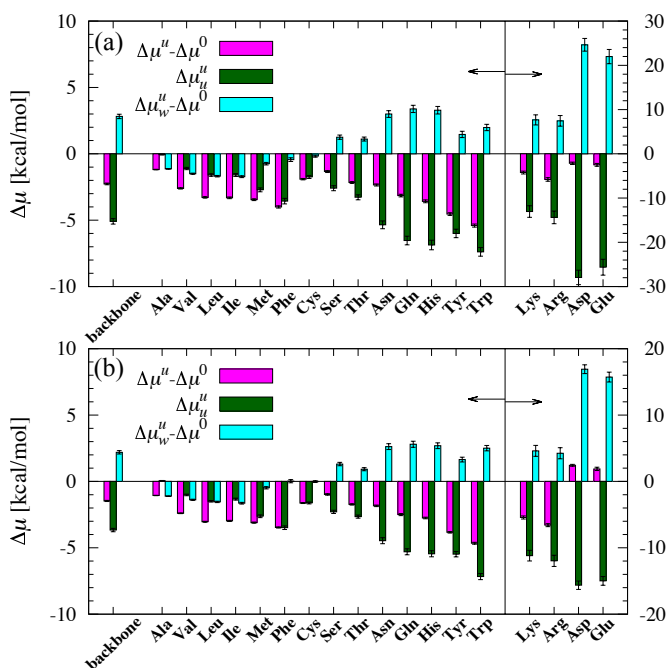


Fig. 6 Transfer free energy from pure water to 8 M urea-water mixture and the contributions from urea and water for (a) Amber urea and (b) KBFF urea.^{61,74} All the free-energy values are obtained with the energy-representation method, and the error bar is expressed at 95% confidence level (twice the standard error). The superscripts u and 0 denote the urea-water mixed solvent and the pure-water solvent, respectively, and the subscripts u and w refer to the urea and water contributions in the mixed solvent, respectively. The error of the overall $(\Delta\mu^u - \Delta\mu^0)$ is observed to be smaller than the errors of the component values $\Delta\mu_u^u$ and $(\Delta\mu_w^u - \Delta\mu^0)$. This is a reflection of the variational principle in the ER method described in Appendix C of Ref.⁵⁰.

as

$$\begin{aligned} \Delta\mu_i &= \langle v \rangle_i - \int d\varepsilon_i f(\varepsilon_i) \\ &= \int d\varepsilon_i \varepsilon_i \rho_i(\varepsilon_i) - \int d\varepsilon_i f(\varepsilon_i), \end{aligned} \quad (4)$$

where ε_i is the pair interaction energy between the solute and the i -th solvent species, $\langle v \rangle_i$ is the average sum of the interaction energy between the solute and the i -th solvent species in the solution system of interest, $\rho_i(\varepsilon_i)$ is the average distribution (histogram) of the pair energy ε_i in the solution system, and $f(\varepsilon_i)$ takes into account the effect of solvent reorganization including the excluded-volume effect. The excluded-volume component in $\Delta\mu_i$ can be introduced on the basis of eqn (4) by restricting the integral over ε_i to high-energy domain ranging from ε_i^c to infinity. ε_i^c is a threshold value for defining the excluded-volume domain. Its value is taken rather

arbitrarily within a requirement that the domain of $\varepsilon_i > \varepsilon_i^c$ corresponds to the solute-solvent overlap and is essentially inaccessible in the solution system of interest. In addition to the arbitrariness of ε_i^c , the integration over the high-energy domain is not independent of the integration over the other domain. The excluded-volume effect can thus be discussed within a specific model of solvation; this is a point similar to the case of decomposition into the “direct” and “indirect” effects described above. In the present work, we set ε_i^c so that $\rho_i(\varepsilon_i)$ vanishes numerically in $\varepsilon_i > \varepsilon_i^c$. The ε_i^c value then depends on the solute and on the solvent species (i is either urea or water). It is also possible to employ a fixed value, for example, of $\varepsilon_i^c = 20$ kcal/mol,⁹² with which the following discussion is unchanged.

Fig. 7 shows the correlation plots of $(\Delta\mu^u - \Delta\mu^0)$ obtained by the ER method with the excluded-volume component $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ as well as with the change $(\langle v \rangle^u - \langle v \rangle^0)$ of the average sum of solute-solvent interaction energy upon transfer. The correlation of the transfer free energy is weaker against the excluded-volume component than against $(\langle v \rangle^u - \langle v \rangle^0)$. The range of variation and the slope of correlation are several times smaller for $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ than for $(\langle v \rangle^u - \langle v \rangle^0)$. In particular, $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ varies over the amino-acid analogs only by ~ 1 kcal/mol, which is similar in magnitude for the variation range of the electrostatic component of $(\langle v \rangle^u - \langle v \rangle^0)$ shown in Fig. 3. According to Fig. 7 and eqn (4), therefore, the excluded-volume component plays a minor role in the transfer energetics.

To see the origin of the small change in the excluded-volume component upon introduction of urea, we decompose $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ into the urea and water contributions on the basis of eqn (3) and (4) and show them in Fig. 8a and 8c. The compensation is evident between the two contributions. When urea is added into the aqueous system, the density of water reduces. The excluded-volume effect then enhances against urea and diminishes against water, and the gain and loss balance with each other.^{93,94}

The excluded-volume component in the solvation free energy can be analyzed in structural viewpoint by employing a limiting expression. In the limiting condition that the solvent-solvent correlation is negligible, the excluded-volume component $(\Delta\mu_i)_{\text{excl}}$ from the i -th solvent species is given by

$$(\Delta\mu_i)_{\text{excl}} = k_B T N_i, \quad (5)$$

$$N_i = d_i \int_{\varepsilon_i^c}^{\infty} d\varepsilon_i \frac{\int d\psi d\mathbf{x} \delta(v_i(\psi, \mathbf{x}) - \varepsilon_i)}{\int d\psi}, \quad (6)$$

where k_B is the Boltzmann constant, T is the temperature, d_i is the number density of the i -th solvent species in the system without the solute, ψ and \mathbf{x} denote collectively the configurations of the solute and (single) solvent molecules, respectively, and v_i is the pair interaction energy between the solute and the

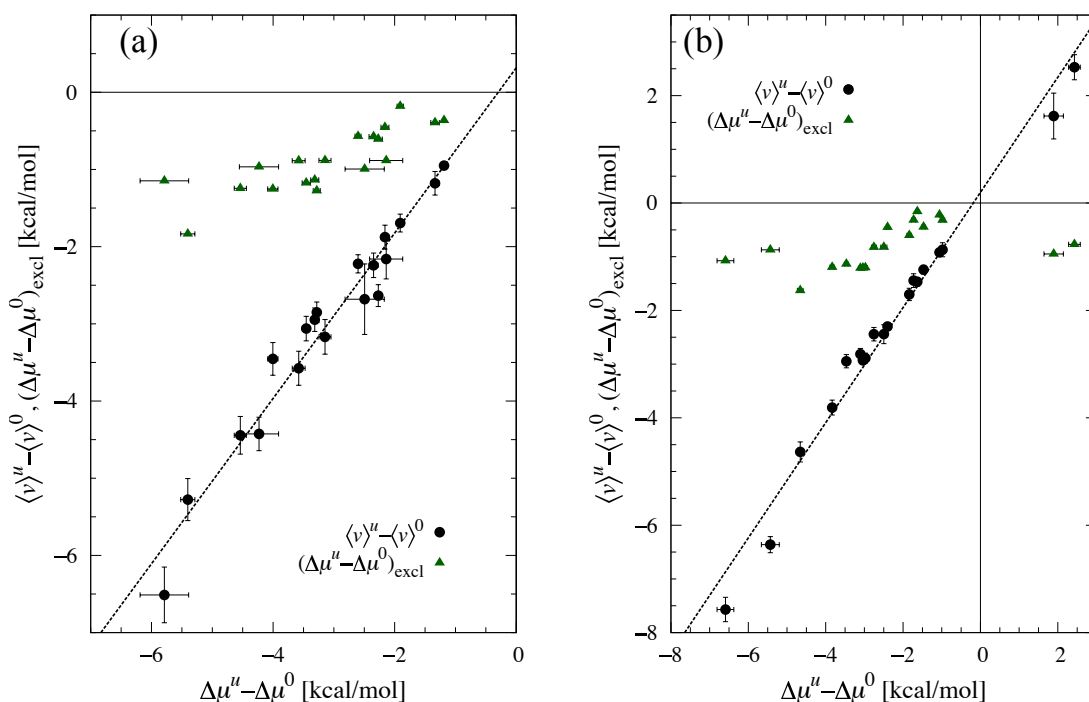


Fig. 7 Correlation plots of transfer quantities for (a) Amber urea and (b) KBFF urea. $(\Delta\mu^u - \Delta\mu^0)$ is the transfer free energy computed approximately by the energy-representation method, and $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ is the excluded-volume component. $(\langle v \rangle^u - \langle v \rangle^0)$ is the change in the sum of solute-solvent interaction energy upon transfer, and is the same as that shown in Figure 3. The error bar is expressed at 95% confidence level (twice the standard error), and is smaller than the size of data symbol when it is not shown. The linear regression against $(\Delta\mu^u - \Delta\mu^0)$ was conducted with the least square fit for $(\langle v \rangle^u - \langle v \rangle^0)$ and $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ over all the solute species. In (a), the regression for $(\langle v \rangle^u - \langle v \rangle^0)$ has a slope of 1.1 with the correlation coefficient of 0.98, and the regression for $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ has a slope of 0.27 with the correlation coefficient of 0.83. In (b), the slope is 1.1 for $(\langle v \rangle^u - \langle v \rangle^0)$ with the correlation coefficient of 0.99, and the slope is 0.09 for $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ with the correlation coefficient of 0.45.

i -th solvent species.

$$\frac{\int d\psi d\mathbf{x} \delta(v_i(\psi, \mathbf{x}) - \varepsilon_i)}{\int d\psi}$$

within eqn (6) is the density of states for the interaction potential v_i at the energy ε_i , and its integral over $\varepsilon_i > \varepsilon_i^c$ is the volume of the space enclosed by $\varepsilon_i > \varepsilon_i^c$. This space is called excluded-volume domain. N_i of eqn (6) is then the average number of the i -th solvent species found in the excluded-volume domain in the system with no solute; in other words, N_i is equal to the average number of solvent molecules removed from the excluded-volume domain upon insertion of the solute. Note that the density of states is independent of whether the solvent is pure water or urea-water mixture. Only d_i depends on the solvent system, and the change in N_i upon transfer reflects that in d_i .⁹⁵ In the following, the average number in the pure-water solvent is denoted as N^0 and is written as N_u^0 and N_w^0 for urea and water in the mixed solvent, respectively, with the sum expressed as $N^u = N_u^u + N_w^u$. In Appendix C, a strong correlation is shown between $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ and

$(N^u - N^0)$. The compensation behavior between the urea and water contributions to $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ can thus be rephrased in terms of a more structural index $(N^u - N^0)$.

In Fig. 8b and 8d, we show $(N^u - N^0)$ and the urea and water contributions. Just as is so for $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$, a compensation is evident between the urea and water contributions to $(N^u - N^0)$. Since urea is larger in size than water, the excluded-volume domain introduced as $\varepsilon_i > \varepsilon_i^c$ occupies a wider space for urea. The variation of the density d_i in eqn (6) is smaller for urea,⁹⁶ however, and the balance is realized between N_u^u and $(N_w^u - N^0)$. This balance cannot be expected in general, and may be specific to urea cosolvent.

4 Conclusions

The transfer energetics from pure water to urea-water mixture is analyzed through all-atom molecular dynamics simulation. The solvation free energy is calculated for a set of amino-acid analogs in the pure-water solvent and in the 8 M

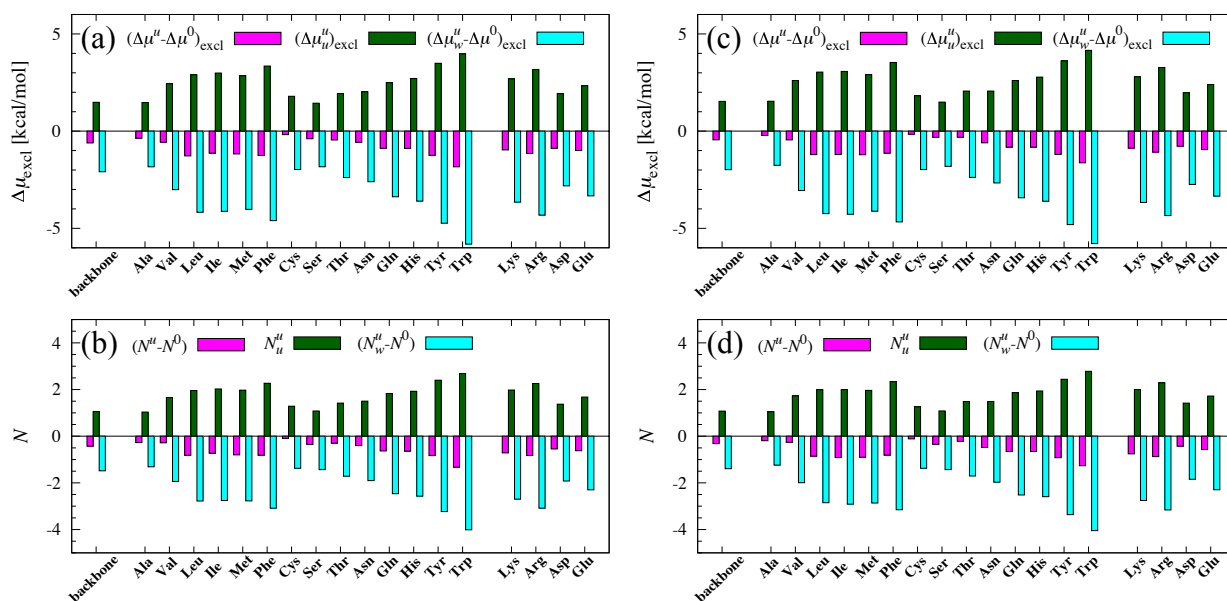


Fig. 8 (a, c) Excluded-volume component $\Delta\mu_{\text{excl}}$ and the contributions from urea and water for (a) Amber urea and (c) KBFF urea. (b, d) Average number N of solvent molecules excluded upon insertion of the solute molecule for (b) Amber urea and (d) KBFF urea. The error at 95% confidence level is too small to see in this figure. The superscripts u and 0 denote the urea-water mixed solvent and the pure-water solvent, respectively, and the subscripts u and w refer to the urea and water contributions in the mixed solvent, respectively.

urea-water mixed solvent, and the transfer free energy is obtained as the difference between the solvation free energies in the two solvent systems. According to the transfer value of free energy, the analog solute favors the mixed solvent over pure water. The transfer free energy is not determined by the strength of hydrophobicity or hydrophilicity of the solute, though, and correlates roughly with the solute size. The decomposition of the transfer free energy into the contributions from urea and water shows that urea makes a favorable contribution. The water contribution is governed by the hydrophobicity/hydrophilicity of the solute, on the other hand. It acts cooperatively with the urea contribution for hydrophobic solute, and is competitive for hydrophilic solute.

The correlation of the transfer free energy is examined against the change in the average sum of solute-solvent interaction energy upon transfer and against the corresponding changes in the interaction components such as electrostatic, van der Waals, and excluded volume. A strong correlation is observed with the total change in average solute-solvent energy, and the van der Waals component makes a dominant contribution for the neutral solute species. The contributions from the electrostatic and excluded-volume components are weak, on the other hand. When the solvent is changed from pure water to urea-water mixture, the reduced effect of water is compensated by the emerging effect of urea for the electrostatic and excluded-volume components. The van der Waals compo-

nent experiences only partial compensation between the urea and water effects, and is favorable upon transfer for all the cases examined.

The free energy examined in the present work is an observable, while the interaction components are not. Still, the electrostatic, van der Waals, and excluded-volume components are more straightforward for interpretation than the free energy, and the correlation analysis can provide a route to designing the cosolvent effect on solute stability in solution. All-atom computation is now feasible for the free energy of protein solvation with our approximate method,^{51–53} and we will conduct a correlation analysis of protein in subsequent work toward possible tuning of cosolvent effect.

Appendix A. Correlation between the computational and experimental values of transfer free energy

Fig. 9 is the correlation plot between the transfer free energy calculated by the BAR method and the corresponding experimental value.¹⁸ The computational and experimental values of transfer free energy are well correlated to each other, though the force-field dependence and the overestimation by computation are evident.^{15,30} In the present work, the main results rely upon the correlation and are rather insensitive to quantitative aspects beyond correlation.

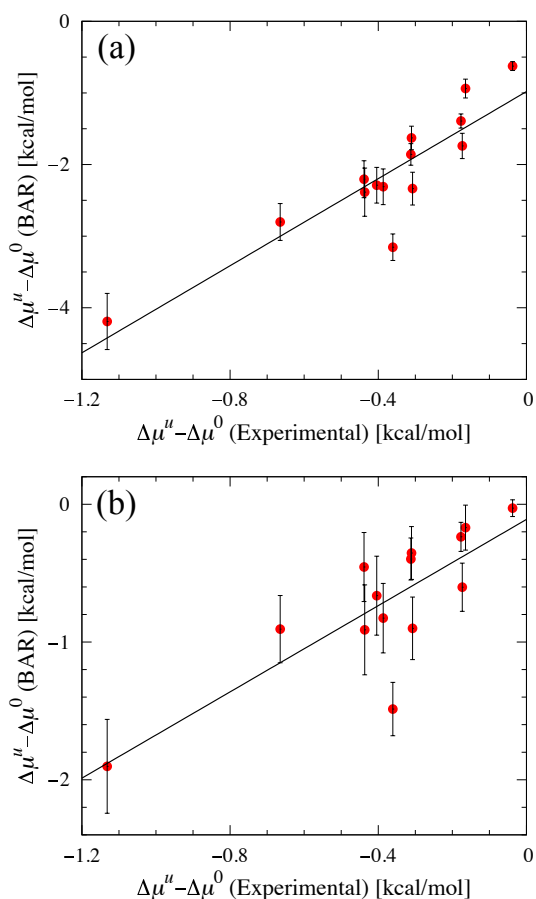


Fig. 9 Transfer free energy ($\Delta\mu^u - \Delta\mu^0$) calculated by the BAR method for (a) Amber urea and (b) KBFF urea plotted against the experimental value. The experimental values for the backbone and side-chains were determined with the group transfer model^{1,17,18}, and were taken from Ref.¹⁸. The ionic and Cys analogs in Table 1 are not shown due to the availability of experimental data. The error bar for computational value is expressed at 95% confidence level (twice the standard error). The slope and correlation coefficient are 3.0 and 0.89, respectively, for Amber urea, and they are 1.6 and 0.81, respectively, for KBFF urea.

For the transfer free energies shown in Fig. 9, the mean absolute deviation (MAD) of eqn (2) between the computational and experimental values is 1.8 kcal/mol for Amber urea and is 0.3 kcal/mol for KBFF urea. The KBFF is better in terms of accuracy and of thermodynamic properties of mixing with water.⁶⁸ The Amber force field is more convenient at present for exploring a variety of cosolvents, on the other hand, since it allows systematic construction of interaction parameters.

Appendix B. Correlation of the transfer free energy calculated approximately by the energy-representation method

Fig. 10 shows the correlation plots of $(\Delta\mu^u - \Delta\mu^0)$ obtained approximately from the energy-representation method against $(\langle v \rangle^u - \langle v \rangle^0)$ and its electrostatic and van der Waals components. As expected from Fig. 5, the correlations appear similarly between Fig. 3 and 10. Fig. 6 and 10 show (for the neutral species) that the direct effect via the van der Waals interaction between solute and urea makes a major contribution in the urea effect to stabilize the analog solutes. This observation regarding the transfer energetics is in agreement with previous, structural and dynamic insights.^{9,10,15,16,20–22,25,26,28–30}

Appendix C. Correlation between the excluded-volume component in the solvation free energy and the average number of solvent molecules excluded upon introduction of the solute

In this Appendix, we show the correlation of the excluded-volume component $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ in the transfer free energy with the corresponding change $(N^u - N^0)$ in the average number of solvent molecules excluded upon introduction of the solute. Fig. 11 is the correlation plot. The change in the average number is expressed in the form of $k_B T(N^u - N^0)$; the two quantities plotted are coincident to each other through eqn (5) when the solvent-solvent correlation is absent. A linear correlation is evident in Fig. 11, and assures the validity of the structure-based discussion at the end of Subsection 3.2. The energy value is a few times larger for $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ than for $k_B T(N^u - N^0)$, though. $k_B T(N^u - N^0)$ does not provide a good, energetic estimate of the excluded-volume effect.^{93,94}

Acknowledgements

This work is supported by the Grants-in-Aid for Scientific Research (Nos. 21300111 and 23651202) from the Japan Society for the Promotion of Science, by the Grant-in-Aid for Scientific Research on Innovative Areas (No. 20118002) and the Elements Strategy Initiative for Catalysts & Batteries from the Ministry of Education, Culture, Sports, Science, and Technology, and by the Nanoscience Program, the Computational Materials Science Initiative, and the Strategic Programs for Innovative Research of the Next-Generation Supercomputing Project. We are also grateful to Professor Yoshihiro Sambongi of Hiroshima University for valuable discussions.

Notes and references

- 1 Y. Nozaki and C. Tanford, *J. Biol. Chem.*, 1963, **238**, 4074.

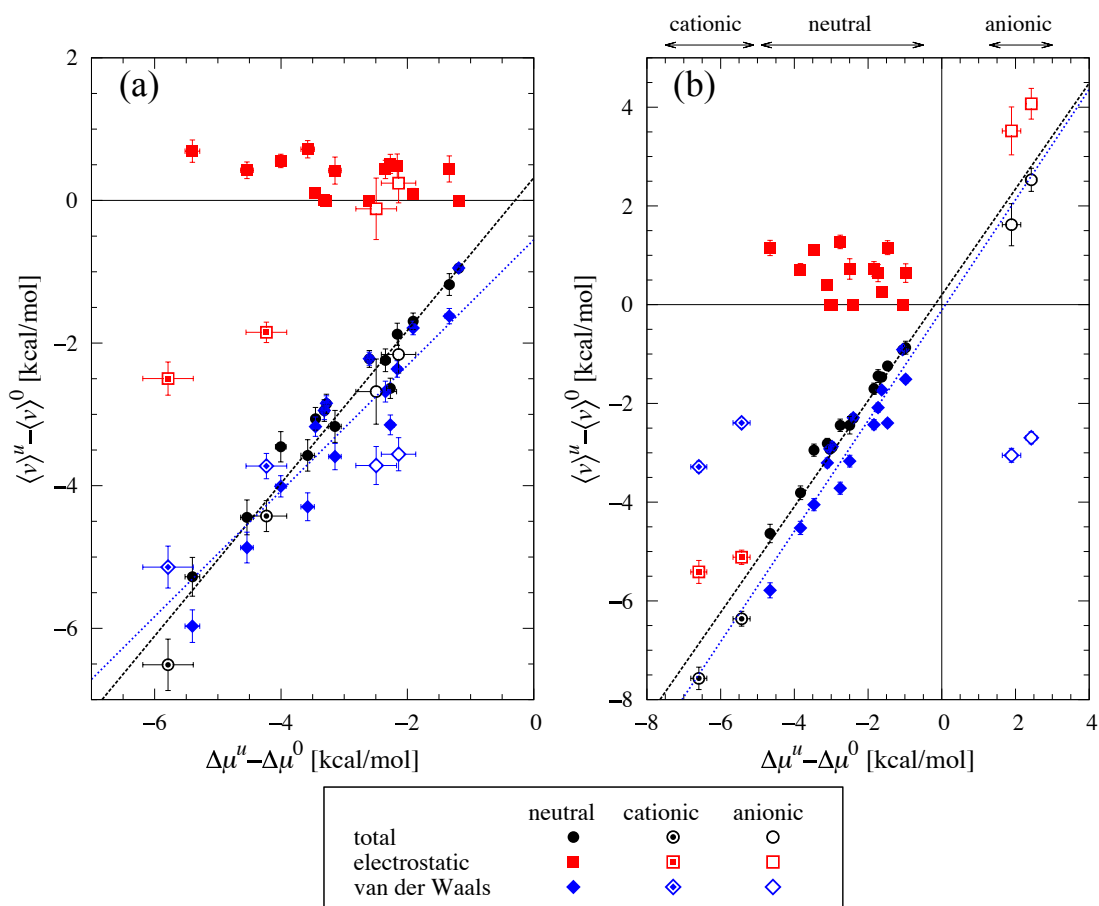


Fig. 10 Correlation plots of transfer quantities for (a) Amber urea and (b) KBFF urea. The free energies $\Delta\mu^u$ and $\Delta\mu^0$ are computed approximately by the energy-representation method. The solute-solvent interaction energies $\langle v \rangle^u$ and $\langle v \rangle^0$ and their electrostatic and van der Waals components are numerically exact and are identical to those in Fig. 3. The error bar is expressed at 95% confidence level (twice the standard error), and is smaller than the size of data symbol when it is not shown. In (a), the linear regression was conducted with the least square fit for the total ($\langle v \rangle^u - \langle v \rangle^0$) and its van der Waals component against ($\Delta\mu^u - \Delta\mu^0$) over all the solute species. The regression for the total ($\langle v \rangle^u - \langle v \rangle^0$) has a slope of 1.1 with the correlation coefficient of 0.98, and the regression for the van der Waals component has a slope of 0.88 with the correlation coefficient of 0.89. In (b), the linear regression was carried out for the total ($\langle v \rangle^u - \langle v \rangle^0$) over all the solute species and for the van der Waals component only over the neutral species. The slope is 1.1 for the total ($\langle v \rangle^u - \langle v \rangle^0$) with the correlation coefficient of 0.99, and the slope is 1.1 for the van der Waals component with the correlation coefficient of 0.94.

- 2 D. R. Robinson and W. P. Jencks, *J. Am. Chem. Soc.*, 1965, **87**, 2462.
- 3 P. K. Nandi and D. R. Robinson, *Biochemistry*, 1984, **23**, 6661.
- 4 E. Liepinsh and G. Otting, *J. Am. Chem. Soc.*, 1994, **116**, 9670.
- 5 A. Wang and D. W. Bolen, *Biochemistry*, 1997, **36**, 9101.
- 6 J. Tirado-Rives, M. Orozco and W. L. Jorgensen, *Biochemistry*, 1997, **36**, 7313.
- 7 A. Wallqvist, D. G. Covell and D. Thirumalai, *J. Am. Chem. Soc.*, 1998, **120**, 427.
- 8 A. Caflich and M. Karplus, *Structure*, 1999, **7**, 477.
- 9 M. Ikeguchi, S. Nakamura and K. Shimizu, *J. Am. Chem. Soc.*, 2001, **123**, 677.
- 10 G. Graziano, *J. Phys. Chem. B*, 2001, **105**, 2632.
- 11 A. K. Soper, E. W. Castner and A. Luzar, *Biophys. Chem.*, 2003, **105**, 649.
- 12 J. A. Schellman, *Biophys. J.*, 2003, **85**, 108.
- 13 D. Tobi, R. Elber and D. Thirumalai, *Biopolymers*, 2003, **68**, 359.
- 14 R. D. Mountain and D. Thirumalai, *J. Phys. Chem. B*, 2004, **108**, 6826.
- 15 D. Trzesniak, N. F. A. van der Vegt and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2004, **6**, 697.
- 16 N. F. A. van der Vegt, D. Trzesniak, B. Kasumaj and W. F. van Gunsteren, *ChemPhysChem*, 2004, **5**, 144.
- 17 M. Auton and D. W. Bolen, *Biochemistry*, 2004, **43**, 1329.
- 18 M. Auton and D. W. Bolen, *Proc. Natl. Acad. Sci. USA*, 2005, **42**, 15065.
- 19 E. P. O'Brien, R. I. Dima, B. Brooks and D. Thirumalai, *J. Am. Chem. Soc.*, 2007, **129**, 7346.
- 20 M. C. Stumpe and H. Grubmuller, *J. Am. Chem. Soc.*, 2007, **129**, 16126.
- 21 M. C. Stumpe and H. Grubmuller, *PLoS Comp. Biol.*, 2008, **4**, e1000221.
- 22 L. Hua, R. Zhou, D. Thirumalai and B. J. Berne, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 16928.

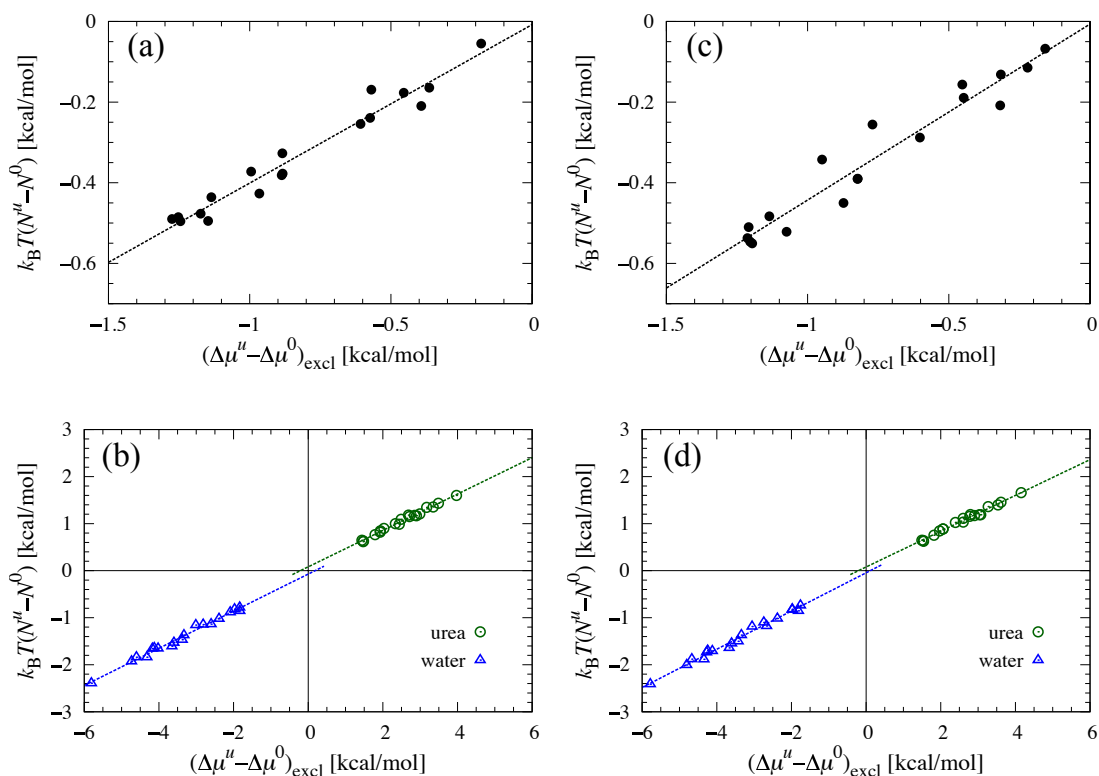


Fig. 11 Correlation plot of the excluded-volume component $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ in the transfer free energy against the change $(N^u - N^0)$ in the number of solvent molecules excluded upon insertion of the solute for (a, b) Amber urea and (c, d) KBFF urea. The ordinate is expressed in the form of $k_B T(N^u - N^0)$, and the error bar at 95% confidence level (twice the standard error) is smaller than the size of data symbol. The total $(\Delta\mu^u - \Delta\mu^0)_{\text{excl}}$ and $k_B T(N^u - N^0)$ are shown in (a) and (c), and the urea and water contributions are plotted in (b) and (d). In (a), the slope is 0.41 with the correlation coefficient of 0.98. In (b), the slope and the correlation coefficient are 0.39 and 0.99, respectively, for urea, and they are 0.39 and 0.99, respectively, for water. In (c), the slope is 0.45 with the correlation coefficient of 0.98. In (d), the slope and the correlation coefficient are 0.38 and 0.99, respectively, for urea, and they are 0.41 and 0.99, respectively, for water.

- 23 H. T. Tran, A. Mao and R. V. Pappu, *J. Am. Chem. Soc.*, 2008, **130**, 7380.
- 24 A. Das and C. Mukhopadhyay, *J. Phys. Chem. B*, 2009, **113**, 12816.
- 25 D. R. Canchi, D. Paschek and A. E. Garcia, *J. Am. Chem. Soc.*, 2010, **132**, 2338.
- 26 R. Zhou, J. Li, L. Hua, Z. Yang and B. J. Berne, *J. Phys. Chem. B*, 2011, **115**, 1323.
- 27 A. Das and C. Mukhopadhyay, *J. Phys. Chem. B*, 2011, **115**, 1327.
- 28 D. R. Canchi and A. E. Garcia, *Biophys. J.*, 2011, **100**, 1526.
- 29 S. Shimizu, *Chem. Phys. Lett.*, 2011, **517**, 76.
- 30 D. Horinek and R. R. Netz, *J. Phys. Chem. A*, 2011, **115**, 6125.
- 31 B. J. Bennion and V. Daggett, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 5142.
- 32 A. Caballero-Herrera, K. Nordstrand, K. D. Berndt and L. Nilsson, *Bio-phys. J.*, 2005, **89**, 842.
- 33 J. L. Finney, A. K. Soper and J. Turner, *Physica B*, 1989, **156-157**, 151.
- 34 R. A. Kuharski and P. J. Rossky, *J. Am. Chem. Soc.*, 1984, **106**, 5786.
- 35 H. Kokubo and B. M. Pettitt, *J. Phys. Chem. B*, 2007, **111**, 5233.
- 36 N. Matubayasi and M. Nakahara, *J. Chem. Phys.*, 2000, **113**, 6070.
- 37 N. Matubayasi and M. Nakahara, *J. Chem. Phys.*, 2002, **117**, 3605; 2003, **118**, 2446 (erratum).
- 38 N. Matubayasi and M. Nakahara, *J. Chem. Phys.*, 2003, **119**, 9686.
- 39 R. M. Levy, M. Belhadj and D. B. Kitchen, *J. Chem. Phys.*, 1991, **95**, 3627.
- 40 V. Luzhkov and A. Warshel, *J. Comput. Chem.*, 1992, **13**, 199.
- 41 J. Åqvist, C. Medina and J.-E. Samuelsson, *Protein Eng.*, 1994, **7**, 385.
- 42 H. A. Carlson and W. L. Jorgensen, *J. Phys. Chem.*, 1995, **99**, 10667.
- 43 G. N. Chuev, M. V. Fedorov and J. Crain, *Chem. Phys. Lett.*, 2007, **448**, 198.
- 44 T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko and F. Hirata, *J. Chem. Phys.*, 2007, **126**, 225102.
- 45 M. Higashi, S. Hayashi and S. Kato, *J. Chem. Phys.*, 2007, **126**, 144503.
- 46 E. L. Ratkova, G. N. Chuev, V. P. Sergiievskiy and M. V. Fedorov, *J. Phys. Chem. B*, 2010, **114**, 12068.
- 47 B. Lin and B. M. Pettitt, *J. Comput. Chem.*, 2011, **32**, 878.
- 48 H. Takahashi, N. Matubayasi, M. Nakahara and T. Nitta, *J. Chem. Phys.*, 2004, **121**, 3989.
- 49 N. Matubayasi, K. K. Liang and M. Nakahara, *J. Chem. Phys.*, 2006, **124**, 154908.
- 50 N. Matubayasi, W. Shinoda and M. Nakahara, *J. Chem. Phys.*, 2008, **128**, 195107.
- 51 H. Saito, N. Matubayasi, K. Nishikawa and H. Nagao, *Chem. Phys. Lett.*, 2010, **497**, 218.
- 52 Y. Karino and N. Matubayasi, *J. Chem. Phys.*, 2011, **134**, 041105.

- 53 K. Takemura, H. Guo, S. Sakuraba, N. Matubayasi and A. Kitao, *J. Chem. Phys.*, 2012, **137**, 215105.
- 54 Y. Karino, M. V. Fedorov and N. Matubayasi, *Chem. Phys. Lett.*, 2010, **496**, 351.
- 55 R. Wolfenden, L. Andersson, P. M. Cullis and C. C. B. Southgate, *Biochemistry*, 1981, **20**, 849.
- 56 M. R. Shirts, J. W. Pitera, W. C. Swope and V. S. Pande, *J. Chem. Phys.*, 2003, **119**, 5740.
- 57 M. R. Shirts and V. S. Pande, *J. Chem. Phys.*, 2005, **122**, 134508.
- 58 The MD box size is so large that the approach to unity at large distance is safely seen for the radial distribution functions of the solute with urea and water.
- 59 F. Figueirido, G. S. Del Buono and R. M. Levy, *J. Chem. Phys.*, 1995, **103**, 6133.
- 60 G. Hummer, L. R. Pratt and A. E. García, *J. Phys. Chem.*, 1996, **100**, 1206.
- 61 The self energy of Ewald (PME) method is the interaction energy of a charge with its own periodic images and the neutralizing background. It does not come from the solute-solvent interaction. When the self energy is written as $\Delta\mu_s^u$ in the urea-water mixed solvent, eqn (3) is actually not true and needs to be expressed as $\Delta\mu^u = \Delta\mu_u^u + \Delta\mu_w^u + \Delta\mu_s^u$. When the self energy is $\Delta\mu_s^0$ in pure water, $\Delta\mu^0$ is similarly given by $\Delta\mu^0 = \Delta\mu_w^0 + \Delta\mu_s^0$, where $\Delta\mu_w^0$ is the contribution from the solute-water intermolecular interaction. The difference ($\Delta\mu_u^u - \Delta\mu_s^0$) is negligible (less than 0.01 kcal/mol in magnitude) for the neutral analogs and is 1.1-1.2 kcal/mol for the ionic analogs. It should be noted in Fig. 6, due to the presence of self-energy term, that ($\Delta\mu^u - \Delta\mu^0$) is not rigorously equal to the sum of $\Delta\mu_u^u$ and ($\Delta\mu_w^u - \Delta\mu^0$). To be exact, furthermore, $\Delta\mu^0$ within ($\Delta\mu^u - \Delta\mu^0$) is the "total" free energy ($\Delta\mu_w^0 + \Delta\mu_s^0$) including the self-energy term, while $\Delta\mu^0$ within ($\Delta\mu_w^u - \Delta\mu^0$) refers only to the intermolecular term $\Delta\mu_w^0$. The points concerning the self-energy term apply in parallel to $\langle v \rangle^u$, $\langle v \rangle^0$, and their urea and water contributions. The discussions for the transfer energetics in the text are unchanged even when the self-energy correction in Ewald method is not taken into explicit account.
- 62 The interaction of an ion with its neutralizing background mimics the effect of counterion.
- 63 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput*, 2008, **4**, 435.
- 64 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct. Funct. Bioinf.*, 2006, **65**, 712.
- 65 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
- 66 Water is treated as rigid and the hydrogen atom does not have the Lennard-Jones term.
- 67 G. A. Özpinar, W. Peukert and T. Clark, *J. Mol. Model.*, 2010, **16**, 1427.
- 68 S. Weerasinghe and P. E. Smith, *J. Phys. Chem. B*, 2003, **107**, 3891.
- 69 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.02*, Gaussian Inc. Wallingford CT 2009.
- 70 W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guid*, Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.
- 71 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577.
- 72 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.
- 73 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, 1987.
- 74 The long-range correction of the LJ interaction is part of solute-solvent interaction. The values of $\Delta\mu_u^u$, $\Delta\mu_w^u$, $\Delta\mu^0$, $\langle v \rangle_u^u$, $\langle v \rangle_w^u$, $\langle v \rangle^0$, and their van der Waals components are the ones corrected for the respective long-range interactions with urea or water.
- 75 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463.
- 76 J. M. Hermida-Ramón, A. Öhrn and G. Karlström, *J. Phys. Chem. B*, 2007, **111**, 11511.
- 77 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952.
- 78 W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Simul.*, 1988, **1**, 173.
- 79 K. Kawahara and C. Tanford, *J. Biol. Chem.*, 1966, **241**, 3228.
- 80 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245.
- 81 T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber and W. F. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 529.
- 82 J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic Press, London, 1986.
- 83 The solvation free energies of the Ile analog at solvent densities of 0.98, 1.03, and 1.07 g/cm³ are 2.4, 4.6, and 6.6 kcal/mol, respectively, when calculated with the Bennett acceptance ratio method. The solvation free energy varies by ~0.5 kcal/mol with the density change of 0.01 g/cm³.
- 84 H. Kokubo, C. Y. Hu and B. M. Pettitt, *J. Am. Chem. Soc.*, 2011, **133**, 1849.
- 85 For example, the internal energy or enthalpy changes with addition of solute, and $\langle v \rangle$ can be a major part of the change. It should be noted, however, that the solvent reorganization term needs to be taken into account to obtain the thermodynamically observable change in energy or enthalpy (Refs.⁸⁶⁻⁸⁹).
- 86 N. Matubayasi, L. H. Reed and R. M. Levy, *J. Phys. Chem.*, 1994, **98**, 10640.
- 87 N. Matubayasi, E. Gallicchio and R. M. Levy, *J. Chem. Phys.*, 1998, **109**, 4864.
- 88 D. Ben-Amotz and B. Widom, *J. Phys. Chem. B*, 2006, **110**, 19839.
- 89 K. Koga, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19749.
- 90 M. Kinoshita, *Front. Biosci.*, 2009, **14**, 3419.
- 91 R. Akiyama, Y. Karino, H. Obama and A. Yoshifuku, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3096.
- 92 In the present simulations, the maximum pair energy with numerically non-zero $\rho_i(\epsilon_i)$ between the analog solute and urea or water in the solution system was 19 kcal/mol.
- 93 In Fig. 7, 8, and 11, the excluded-volume component ($\Delta\mu^u - \Delta\mu^0$)_{excl} in the transfer free energy is seen to be negative and the free-energy penalty of cavity formation is less positive in the urea-water mixed solvent than in the pure-water solvent. This observation is in contrast to Graziano's analysis on the basis of scaled-particle theory⁹⁴ that the free-energy penalty is more positive in the urea-water mixture.¹⁰ ($\Delta\mu^u - \Delta\mu^0$)_{excl} becomes positive, in fact, when the threshold energy ϵ_i^c of eqn (6) is modified from the one in the text and is taken so that $N^u > N^0$ is satisfied through eqn (6).
- 94 R. A. Pierotti, *J. Phys. Chem.*, 1965, **69**, 281.
- 95 To be exact, the threshold energy ϵ_w^c for water is different between pure water and urea-water mixed solvent when it is introduced by $\rho_i(\epsilon_i) = 0$ in $\epsilon_i > \epsilon_i^c$. It is possible to set the same ϵ_w^c values in the pure-water solvent and the mixed solvent, for example, by applying the ϵ_w^c value in either of the two solvents to both solvent systems. The effect of the ϵ_w^c difference is then found to be negligible in our discussion.
- 96 The density of water is 35.6 M in its 8 M mixture with urea. From pure

water to the 8 M urea-water mixture, the density d_i varies by 8 M for urea and by -20 M for water.