HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2020 May 14.

Published in final edited form as:

J Chem Theory Comput. 2019 May 14; 15(5): 2896–2912. doi:10.1021/acs.jctc.8b01309.

Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method

Yoshiki Ishii[†], **Naoki Yamamoto**[†], **Nobuyuki Matubayasi**[*,†,‡], **Bin W. Zhang**[¶], **Di Cui**[¶], and **Ronald M. Levy**[¶]

[†]Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

[‡]Elements Strategy Initiative for Catalysts and Batteries, Kyoto University, Katsura, Kyoto 615-8520, Japan

[¶]Center for Biophysics and Computational Biology, Department of Chemistry, and Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122, United States

Abstract

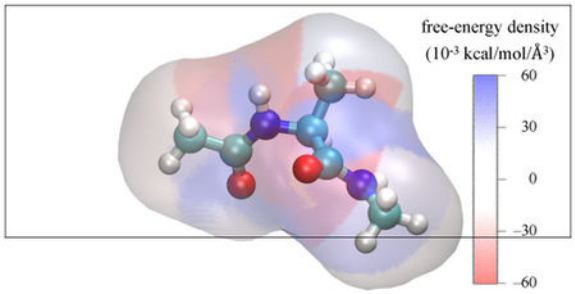
A spatially resolved version of the density-functional method for solvation thermodynamics is presented by extending the free-energy functional previously established in the 1-dimensional, energy representation and formulating a new expression in a mixed 4-dimensional representation (3 dimension for position and 1 dimension for energy). The space was further divided into a set of discrete regions with respect to the relative position of a solvent molecule from the solute, and the spatially-decomposed energetics of solvation was analyzed for small molecules with a methyl, amine, or hydroxyl group and alanine dipeptide in solvent water. It was observed that the density of the solvation free energy is weakly dependent on the solute site in the excluded-volume region and is distinctively favorable in the first shells of the solute atoms that can readily form hydrogen bonds with water. The solvent-reorganization term reduces faster with the separation from the solute than the direct interaction between the solute and solvent, and the latter governs the energetics in the second shell and outer regions. The sum of the contributions to the free energy from the excluded volume and first shell was found to be deviate significantly from the total sum over all the regions, implying that the solvation free energy is not spatially localized near the solute in a quantitative sense. Still, a local description was shown to be valid as confirmed by the correlation of the total value of free energy with the corresponding value obtained by integrating the free-energy density to the second shell. The theoretical framework developed in the present work to spatially decompose the solvation free energy can thus be useful to identify stabilizing or destabilizing regions of solvent proximate to a solute and to analyze the role that the displacement of interfacial water plays in the thermodynamics of molecular association.

Graphical Abstract

[*] nobuyuki@cheng.es.osaka-u.ac.jp.

Supporting Information Available

Numerical values of $\Delta\mu(i)/v(i)$, $u(i)/v(i)$, $\Delta\mu(i)$, and $u(i)$; volume of each region $v(i)$; correlation plots between $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ in the first and second shells.



Introduction

Solvation free energy quantifies the influence of solvent on the stability of a solute. It is a thermodynamic variable, and describes the effect of solute-solvent interaction as a whole. A molecular picture is established, on the other hand, by identifying chemically important groups of atoms and pursuing local structure and energetics around them. The focus is then the solvent molecules located proximately to the solute, and a theoretical framework is required to bridge the spatially resolved information to a spatially averaged observable such as free energy. The scheme of spatial decomposition meets this requirement. It expresses a thermodynamic, transport, or spectroscopic quantity as a spatial integral of relevant, molecular correlation functions and sets a basis for specifying the important regions that determine the observable. The spatial decomposition was conducted to analyze the energetics and volumetrics of hydration[1-4] and was adopted to investigate the ion-pair contribution to the electrical conductivity and the solvation-shell contributions to dynamic and spectroscopic observables.[5-16]

The spatially resolved energetics receives much attention recently in connection to ligand design for proteins.[17-28] The inhomogeneous solvation theory sets a basis for investigating the energy and entropy of water molecules at the biomolecular surface, and a practical strategy has been proposed to identify the locations of water that are to be displaced by ligands possibly with improved affinities. The empirical scheme based on the accessible surface area and the method of 3D-RISM (three-dimensional reference interaction site model) can also be viewed as ones with spatial (regional) decomposition.[29-32] A set of thermodynamic parameters are assigned to or determined for each group of atoms (functional group), and in this procedure, the space around the solute is effectively divided by specifying the contributions from the atom groups. The density functional theory (DFT) is an alternative scheme for expressing the solvation free energy as an integral over the space.[33-43] The local signature for assessing the solvent's effect on the solute stability is then the indirect (solvent-mediated) part of the solute-solvent potential of mean force,[35,42,44] which describes the solvent-solvent correlation modified by the presence of the solute.

The present work performs the spatial decomposition of the solvation free energy on the basis of the endpoint density-functional formula. In the DFT formalism, the solvation free energy is expressed as a functional of the solvent density around the solute,[33,34] and with introduction of approximations, the free energy can be evaluated from distribution functions in pure solvent and in the solution system of interest; note that these two are the initial and

final (endpoint) states of solute insertion. When the solute is placed at the origin with fixed orientation, the solvent distribution around it can be represented as a function of the position and orientation of the solvent molecule. Such a six-dimensional representation is not numerically advantageous, however, and a scheme of dimensionality reduction is required for implementing a density-functional method in practice. In previous works,[35,41,44-47] an endpoint DFT was formulated by adopting the pair-interaction energy between the solute and solvent as the one-dimensional coordinate for the distribution functions. This is called energy-representation method, and was applied to a variety of system including protein and lipid membrane.[41,48-53] The solute-solvent configuration is then projected onto the energy coordinate, and thus the spatial information of the solvent around the solute is lost. To restore a spatially decomposed view of the solvation free energy within the DFT framework, therefore, it is necessary to formulate a representation that retains both the spatial information and the numerical feasibility.

In the present work, we formulate a functional for the solvation free energy by adopting a mixed representation with the position and energy. The energy refers to the pair energy between the solute and solvent and acts as a proxy for the solvent orientation relative to the solute. The coordinate for the distribution functions is then 4-dimensional with further reduction possible by the division of the space into a set of discrete regions, and a DFT expression can be developed in parallel to the cases of the six-dimensional (full position and orientation) and one-dimensional (energy) representations.[42] We provide a formalism based on the Kirkwood charging formula and conduct the spatial decomposition for the free energies of solvation in water of small molecules with a methyl, amine, or hydroxyl group and of alanine dipeptide at representative conformations.

When the solvation free energy is spatially decomposed, the contribution from a region and/or an atomic group can be identified, for example, as hydrophobic or hydrophilic in solvent water in terms of its sign. This is a useful feature for quantitatively assessing the role of each contribution in the solvation effect, whereas the decomposed values are not observable in general. In the present work, we treat the solute-solvent interaction in each region examined and discuss the hydrophobicity or hydrophilicity of the methyl, amine, hydroxyl, and phenyl groups in small molecules and of the atomic sites in alanine dipeptide. Another feature of the spatial-decomposition analysis is that the extent of spatial localization can be addressed for the free energy of solvation. The total value of the free energy is then compared to the partial contribution from the excluded-volume, first-shell, and second-shell regions of the solute, and it is seen that the solvation free energy is not localized in the sense that the partial contribution deviates significantly from the total. Still, a good correlation is observed between the partial and total values, showing that a local view is adequate for describing the free-energetics of solvation.

Theory

In this section, we describe the DFT formalism for the solvation free energy $\Delta\mu$. We briefly summarize the developments in the six-dimensional (full position and orientation) and one-dimensional (energy) representations and provide a spatially-decomposed expression for $\Delta\mu$. We suppose that the solute and solvent are both rigid and that the solution is at infinite

dilution with a single-component solvent. Actually, we will note at the end of the section that the supposition of rigidity is not necessary. The following developments can also be extended straightforwardly when the solute is at finite concentration and/or the solvent is a mixed one with more than a single species.

The solvation free energy $\Delta\mu$ is the free-energy change for turning on the solute-solvent interaction. When the solute is located at the origin with fixed orientation, $\Delta\mu$ is expressed as

$$\exp(-\beta\Delta\mu) = \frac{\int d\mathbf{X} \exp\left(-\beta\left\{\sum_l v(\mathbf{x}_l) + U(\mathbf{X})\right\}\right)}{\int d\mathbf{X} \exp\left(-\beta U(\mathbf{X})\right)}, \quad (1)$$

where $\mathbf{x}_l$ is the configuration of the lth solvent molecule, $\mathbf{X}$ denotes the solvent configuration collectively, $U(\mathbf{X})$ is the total energy among the solvent, $v(\mathbf{x})$ is the pair-interaction potential between the solute and solvent, and β is the inverse of k_BT with the Boltzmann constant k_B and the temperature T. Note that $\mathbf{x}$ is a six-dimensional coordinate representing the position and orientation of the solvent molecule relative to the solute. To formulate an endpoint DFT expression for $\Delta\mu$, we introduce a set of intermediate states connecting the initial and final states of the insertion process of the solute, where the initial and final states are the pure-solvent system without solute and the solution system of interest, respectively. Let λ be the coupling parameter identifying the state and $u_\lambda(\mathbf{x})$ be the solute-solvent interaction potential at the coupling parameter λ. When $\lambda = 0$, the system is the pure solvent and $u_0 = 0$ (no solute-solvent interaction). The solute molecule at $\lambda = 0$ is a virtual particle, and is located in the system as a test particle. When $\lambda = 1$, the solute interacts with the solvent at full coupling and $u_1 = v$, where v is the solute-solvent interaction of interest. $\Delta\mu$ is then given by the Kirkwood charging formula of

$$\Delta\mu = \int_0^1 d\lambda \int d\mathbf{x} \frac{\partial u_\lambda(\mathbf{x})}{\partial\lambda} \rho_\lambda^f(\mathbf{x}) \quad (2)$$

where ρ_λ^f is the one-body distribution of the solvent around the solute at the coupling parameter λ. It is the ensemble average of the instantaneous distribution of $\mathbf{x}$ defined as

$$\hat{\rho}^f(\mathbf{x}) = \sum_l \delta(\mathbf{x} - \mathbf{x}_l), \quad (3)$$

with the superscript f meaning that the function is represented over the full coordinate of position and orientation, and an ensemble average at λ is written as

$$\langle\cdots\rangle_\lambda = \frac{\int d\mathbf{X}(\cdots) \exp\left(-\beta\left\{\sum_l u_\lambda(\mathbf{x}_l) + U(\mathbf{X})\right\}\right)}{\int d\mathbf{X} \exp\left(-\beta\left\{\sum_l u_\lambda(\mathbf{x}_l) + U(\mathbf{X})\right\}\right)}. \quad (4)$$

The solvation free energy $\Delta\mu$ consists of the solute-solvent interaction energy and the solvent-reorganization term. The latter is the free-energy penalty corresponding to the change in the solvent distribution upon insertion of the solute, and is introduced from the Kirkwood charging formula Eq. 2 through partial integration as

$$\Delta\mu = \int d\mathbf{x} v(\mathbf{x})\rho^f(\mathbf{x}) - \int_0^1 d\lambda \int d\mathbf{x} u_\lambda(\mathbf{x}) \frac{\partial \rho_\lambda^f(\mathbf{x})}{\partial \lambda} \quad (5)$$

where $\rho^f(\mathbf{x})$ denotes $\rho_\lambda^f(\mathbf{x})$ at $\lambda = 1$ and is the density of solvent in the solution system of interest. The second term of Eq. 5 is the density-functional and is the free energy of solvent reorganization. To separate it further to the solute-solvent pair contribution and the term incorporating the solvent-solvent correlations, we define the indirect (solvent-mediated) part $\omega_\lambda^f(\mathbf{x})$ of the potential of mean force as

$$\rho_\lambda^f(\mathbf{x}) = \rho_0^f(\mathbf{x}) \exp\left(-\beta\left\{u_\lambda(\mathbf{x}) + \omega_\lambda^f(\mathbf{x})\right\}\right) \quad (6)$$

where ρ_0^f stands for ρ_λ^f at $\lambda = 0$. Equation 5 can then be modified into[42,44,47]

$$\Delta\mu = \int d\mathbf{x} v(\mathbf{x})\rho^f(\mathbf{x}) - k_B T \int d\mathbf{x}\left[\rho^f(\mathbf{x}) - \rho_0^f(\mathbf{x}) - \rho^f(\mathbf{x}) \log\left(\frac{\rho^f(\mathbf{x})}{\rho_0^f(\mathbf{x})}\right)\right] + \int_0^1 d\lambda \int d\mathbf{x} \omega_\lambda^f(\mathbf{x}) \frac{\partial \rho_\lambda^f(\mathbf{x})}{\partial \lambda}$$

(7)

This is an exact expression. Its first term is the average sum of the solute-solvent interaction energy at $\lambda = 1$, and the second term is the pair entropy that quantifies the deviation of the solvent distribution in the solution system ($\lambda = 1$) from that in the pure solvent ($\lambda = 0$). The third term corresponds to the change in the solvent-solvent correlations upon gradual insertion of the solute with the coupling parameter λ. As noted in Refs. 42 and 43, the integrand of Eq. 7 is of $o(1/V)$ at large distances from the solute, where V refers to the system size and $o(1/V)$ denotes a quantity that vanishes faster than $1/V$ in the thermodynamic limit ($V \to \infty$). There is then no contribution to the integral from the bulk far from the solute, and the integral value of Eq. 7 is assured to be independent of the choice of ensemble (for example, canonical vs. isothermal-isobaric). This is because the ensemble dependence arises from the presence of an $O(1/V)$ correlation in the bulk region. [3]

The difficulty in using Eq. 7 at the computation of $\Delta\mu$ is the high dimensionality of the full coordinate (position and orientation) $\mathbf{x}$. To make feasible the DFT route to $\Delta\mu$, a formulation

is required which expresses $\Delta\mu$ only in terms of distribution functions over a coordinate of reduced dimension. The method of energy representation was formulated as a DFT scheme over a one-dimensional coordinate.[35,44] In this method, the solute-solvent pair interaction $v(\mathbf{x})$ in the solution system of interest plays the key role and the instantaneous distribution of its value ϵ is introduced as

$$\hat{\rho}^e(\epsilon) = \sum_l \delta(\epsilon - v(\mathbf{x}_l)), \quad (8)$$

where the sum is taken over the solvent molecules and the superscript e is attached to emphasize that a function is represented over the energy coordinate. A density-functional treatment can then be implemented by restricting the set of solute-solvent interaction potentials $u_\lambda(\mathbf{x})$ to those which are constant over equi-energy surfaces of $v(\mathbf{x})$. With this setup, u_λ is a composite function with the form of $u_\lambda(v(\mathbf{x}))$ and can be written as $u_\lambda(\epsilon)$ when ϵ denotes the value of $v(\mathbf{x})$. At the endpoints ($\lambda = 0$ and 1), $u_0(\epsilon) = 0$ and $u_1(\epsilon) = \epsilon$ since $v(\mathbf{x})$ is the potential function in the solution system of interest.

The Kirkwood charging formula in the energy representation is then given by

$$\Delta\mu = \int_0^1 d\lambda \int d\epsilon \frac{\partial u_\lambda(\epsilon)}{\partial\lambda} \rho_\lambda^e(\epsilon) \quad (9)$$

where ρ_λ^e is the distribution function in the presence of u_λ and is the ensemble average of Eq. 8 through Eq. 4. The indirect part $\omega_\lambda^e(\epsilon)$ of the potential of mean force can also be defined in parallel to Eq. 6 as [35,42,44,47]

$$\rho_\lambda^e(\epsilon) = \rho_0^e(\epsilon) \exp(-\beta\{u_\lambda(\epsilon) + \omega_\lambda^e(\epsilon)\}) \quad (10)$$

and Eq. 9 reduces to

$$\Delta\mu = \int d\epsilon \epsilon \rho^e(\epsilon) - k_B T \int d\epsilon \left[\rho^e(\epsilon) - \rho_0^e(\epsilon) - \rho^e(\epsilon) \log\left(\frac{\rho^e(\epsilon)}{\rho_0^e(\epsilon)}\right)\right] + \int_0^1 d\lambda \int d\epsilon \omega_\lambda^e(\epsilon) \frac{\partial \rho_\lambda^e(\epsilon)}{\partial\lambda}$$

(11)

where $\rho_0^e(\epsilon)$ and $\rho^e(\epsilon)$ denote $\rho_\lambda^e(\epsilon)$ at $\lambda = 0$ and 1, respectively. Equation 11 is exact and it has a similar structure to Eq. 7. The first term is equal to the average sum of the solute-solvent interaction energy in the solution ($\lambda = 1$), and the second term expresses the pair entropy in the energy representation. The third term takes into account the effects of solvent-

solvent correlations, and was approximated by a combined Percus-Yevick (PY)-type and hypernetted-chain (HNC)-type relationship in previous works.[44,47,49] As in Eq. 7, the solvent-reorganization term in Eq. 11 is the sum of the second and third terms. It should be noted that although Eq. 7 is not easy to implement due to the high dimensionality of the full coordinate **x**, Eq. 11 is straightforward to handle since the energy coordinate ϵ is one-dimensional.

A notable feature of Eq. 8 is that the spatial information of the solvent around the solute is projected out. To restore the spatial resolution at the lowest possible dimensionality, a mixed space-energy representation can be proposed by using the solute-solvent pair energy ϵ as a proxy for the solvent orientation relative to the solute and defining the instantaneous distribution as

$$\hat{\rho}^m(\mathbf{r},\epsilon) = \sum_l \delta(\mathbf{r}-\mathbf{r}_l)\delta(\epsilon - v(\mathbf{x}_l)), \quad (12)$$

where **r** is the position of a solvent molecule and the superscript m means the mixed representation. Equation 12 is a marginal distribution of Eq. 3 obtained by projecting the full coordinate **x** to (**r**, ϵ). A typical choice of **r** can correspond to the center of mass of the solvent molecule or the oxygen site in the case of water. A DFT expression for $\Delta\mu$ can be formulated in a similar manner as above by considering the set of solute-solvent interactions with the form of $u_\lambda(\mathbf{r},\epsilon)$; $u_0(\mathbf{r},\epsilon) = 0$ at $\lambda = 0$ and $u_1(\mathbf{r},\epsilon) = \epsilon$ holds at $\lambda = 1$. u_λ can depend explicitly on the position **r** only at the intermediate states ($0 < \lambda < 1$). When the ensemble average of Eq. 12 at the coupling parameter λ is written as $\rho_\lambda^m(\mathbf{r},\epsilon)$ and $\rho_0^m(\mathbf{r},\epsilon)$ denotes $\rho_\lambda^m(\mathbf{r},\epsilon)$ at $\lambda = 0$, the indirect part $\omega_\lambda^m(\mathbf{r},\epsilon)$ of the potential of mean force can be introduced as

$$\rho_\lambda^m(\mathbf{r},\epsilon) = \rho_0^m(\mathbf{r},\epsilon)\exp\left(-\beta\left\{u_\lambda(\mathbf{r},\epsilon) + \omega_\lambda^m(\mathbf{r},\epsilon)\right\}\right) \quad (13)$$

in the mixed representation. The Kirkwood charging formula then reads as

$$\Delta\mu = \int_0^1 d\lambda \int d\mathbf{r}d\epsilon \frac{\partial u_\lambda(\mathbf{r},\epsilon)}{\partial\lambda}\rho_\lambda^m(\mathbf{r},\epsilon) \quad (14)$$

and modifies further into[42]

$$\Delta\mu = \int d\mathbf{r}d\epsilon\epsilon\rho^m(\mathbf{r},\epsilon) - k_BT\int d\mathbf{r}d\epsilon\left[\rho^m(\mathbf{r},\epsilon) - \rho_0^m(\mathbf{r},\epsilon) - \rho^m(\mathbf{r},\epsilon)\log\left(\frac{\rho^m(\mathbf{r},\epsilon)}{\rho_0^m(\mathbf{r},\epsilon)}\right)\right] \quad (15)$$
$$+ \int_0^1 d\lambda\int d\mathbf{r}d\epsilon\omega_\lambda^m(\mathbf{r},\epsilon)\frac{\partial\rho_\lambda^m(\mathbf{r},\epsilon)}{\partial\lambda}$$

where $\rho^m(\mathbf{r}, \epsilon)$ means $\rho_\lambda^m(\mathbf{r}, \epsilon)$ at $\lambda = 1$ for notational brevity. It should also be noted that the distribution function $\rho_\lambda^{\mathrm{pos}}(\mathbf{r})$ of the solvent position $\mathbf{r}$ satisfies

$$\rho_\lambda^{\mathrm{pos}}(\mathbf{r}) = \int d\epsilon \rho_\lambda^m(\mathbf{r}, \epsilon), \quad (16)$$

and $\rho_\lambda^e(\epsilon)$ in the 1-dimensional, energy representation is related to $\rho_\lambda^m(\mathbf{r}, \epsilon)$ through

$$\rho_\lambda^e(\epsilon) = \int d\mathbf{r} \rho_\lambda^m(\mathbf{r}, \epsilon) \quad (17)$$

by virtue of Eqs. 8 and 12. Equations 16 and 17 show that $\rho_\lambda^{\mathrm{pos}}(\mathbf{r})$ and $\rho_\lambda^e(\epsilon)$ are marginal distributions of $\rho_\lambda^m(\mathbf{r}, \epsilon)$, which is the solvent density around the solute in a mixed representation of position and energy. In the following, $\rho_\lambda^{\mathrm{pos}}(\mathbf{r})$ at $\lambda = 1$ is denoted as $\rho^{\mathrm{pos}}(\mathbf{r})$.

As is so for Eqs. 7 and 11, Eq. 15 is an exact expression. The spatial decomposition of $\Delta\mu$ can then be conducted by rewriting Eq. 15 as

$$\Delta\mu = \int d\mathbf{r} \Delta\mu(\mathbf{r}) \quad (18)$$

$$\Delta\mu(\mathbf{r}) = \int d\epsilon \epsilon \rho^m(\mathbf{r}, \epsilon) - k_B T \int d\epsilon \left[\rho^m(\mathbf{r}, \epsilon) - \rho_0^m(\mathbf{r}, \epsilon) - \rho^m(\mathbf{r}, \epsilon) \log\left(\frac{\rho^m(\mathbf{r}, \epsilon)}{\rho_0^m(\mathbf{r}, \epsilon)} \right) \right] \quad (19)$$
$$+ \int_0^1 d\lambda \int d\epsilon \omega_\lambda^m(\mathbf{r}, \epsilon) \frac{\partial \rho_\lambda^m(\mathbf{r}, \epsilon)}{\partial \lambda}.$$

$\Delta\mu(\mathbf{r})$ is a free-energy density at $\mathbf{r}$. When the first term of Eq. 19 is divided by the spatial density $\rho^{\mathrm{pos}}(\mathbf{r})$, the quotient is the average sum at $\lambda = 1$ of the interaction energy with the solute of the solvent molecules located at the position $\mathbf{r}$. Actually, $\Delta\mu(\mathbf{r})$ is not zero even when $\rho^{\mathrm{pos}}(\mathbf{r})$ is vanishingly small. $\rho^{\mathrm{pos}}(\mathbf{r})$ is exactly zero when the solute-solvent interaction energy is infinite at $\mathbf{r}$, and is numerically zero in a simulation of finite length when $\mathbf{r}$ is contained in the excluded-volume domain. In either case, $\Delta\mu(\mathbf{r})$ does not vanish and will be unfavorable (positive) to reflect the excluded-volume effect.

In numerical implementations, the spatial position is expressed with finite bins. A molecular picture may also be obtained by focusing on the solvation shells around chemically important groups of atoms. When the space is divided into a set of discrete regions, the contribution from each region to $\Delta\mu$ can be formulated in the DFT framework by introducing

$$\hat{\rho}^d(\epsilon; i) = \sum_l \Theta(\mathbf{r}_l; i)\delta(\epsilon - v(\mathbf{x}_l)), \quad (20)$$

where $\Theta(\mathbf{r}; i)$ is the characteristic function for the ith region in the space and is taken to satisfy

$$\sum_i \Theta(\mathbf{r}; i) = 1 \quad (21)$$
$$\Theta(\mathbf{r}; i)\Theta(\mathbf{r}; j) = 0 \qquad \text{for } i \neq j$$

for any position $\mathbf{r}$ in space. With Eq. 21, $\Theta(\mathbf{r}; i)$ at a given $\mathbf{r}$ is unity for a single i and vanishes for the other i's. The division of the space is then unique. A position in space is always contained in one of the regions and can never belong to multiple regions. Equation 20 is actually a marginal distribution of Eq. 3 since it is obtained through the projection from the full coordinate $\mathbf{x}$ in Eq. 3 to the continuous position and energy $(\mathbf{r}, \epsilon)$ in Eq. 12 and a further marginalization expressed as

$$\hat{\rho}^d(\epsilon; i) = \int d\mathbf{r}\Theta(\mathbf{r}; i)\hat{\rho}^m(\mathbf{r}, \epsilon) = \int_{\text{region } i} d\mathbf{r}\hat{\rho}^m(\mathbf{r}, \epsilon)\,. \quad (22)$$

An example of division is shown in Fig. 1. Three centers A, B, and C are specified, and the space is first divided into three regions in terms of the distances from A, B, and C with the boundaries given by the dashed lines. Each of the three regions is then divided with respect to the distance from the center, and the example of Fig. 1 consists of 12 regions represented by distinct colors. The regions in Fig. 1 satisfy Eq. 21 by construction, and may not be spherical in spite of the distance criteria used for the division. In RESULTS AND DISCUSSION section, we will divide the space as above. The centers A, B, and C in Fig. 1 do not have to refer to three atomic positions. They may only be three sites arbitrarily chosen within a solute molecule, that is not necessary to be a triatomic one. When A, B, and C are positions of atoms, the shells around them can be used to define the regions. Each region may be a non-spherical part of a (spherical) shell, though, and when the molecule is not triatomic, the shell structures around the atoms other than A, B, and C are not reflected to introduce the regions in Fig. 1. A position in space is specified only with the distances from A, B, and C, and its distances with the other atoms are disregarded. It should be remarked again that the regions can be introduced arbitrarily as far as Eq. 21 is assured. The radial distances from a set of specific atoms are only a convenient option for division with Eq. 21. When the solute is a protein, for example, its cavity or active site may be adopted as a region. A grid can be employed, too, since it satisfies Eq. 21.

The DFT expression for $\Delta\mu$ in the disretized space-energy representation is formulated in parallel to that in the continuous representation based on Eq. 12. The indirect part $\omega_\lambda^d(\epsilon; i)$ of the potential of mean force is defined as

$$\rho_\lambda^d(\epsilon;i) = \rho_0^d(\epsilon;i)\exp\left(-\beta\left\{u_\lambda(\epsilon;i) + \omega_\lambda^d(\epsilon;i)\right\}\right) \quad (23)$$

and the Kirkwood charging formula is

$$\Delta\mu = \sum_i \int_0^1 d\lambda \int d\epsilon \frac{\partial u_\lambda(\epsilon;i)}{\partial\lambda}\rho_\lambda^d(\epsilon;i)\,. \quad (24)$$

In these equations, $\rho_\lambda^d(\epsilon;i)$ is the ensemble average of Eq. 20 in the presence of $u_\lambda(\epsilon;i)$, where $u_\lambda(\epsilon;i)$ is the solute-solvent interaction at the coupling parameter λ in region i. $u_0(\epsilon;i) = 0$ and $u_1(\epsilon;i) = \epsilon$ are satisfied at the endpoints ($\lambda = 0$ and 1), and the explicit dependence of u_λ on i appears only at the intermediate states. Equation 24 is then rewritten as

$$\Delta\mu = \sum_i \Delta\mu(i) \quad (25)$$

$$\Delta\mu(i) = \int d\epsilon\epsilon\rho^d(\epsilon;i) - k_BT\int d\epsilon\left[\rho^d(\epsilon;i) - \rho_0^d(\epsilon;i) - \rho^d(\epsilon;i)\log\left(\frac{\rho^d(\epsilon;i)}{\rho_0^d(\epsilon;i)}\right)\right] + \int_0^1 d\lambda\int d\epsilon\omega_\lambda^d(\epsilon;i)\frac{\partial\rho_\lambda^d(\epsilon;i)}{\partial\lambda}, \quad (26)$$

where $\rho^d(\epsilon;i)$ denotes $\rho_\lambda^d(\epsilon;i)$ at $\lambda = 1$. $\Delta\mu(i)$ is the contribution from region i to the (total) free energy $\Delta\mu$ of solvation. The first term of Eq. 26 is the averaged interaction energy of the solvent in region i with the solute in the solution ($\lambda = 1$), and the rest is the solvent-reorganization term as in Eqs. 7, 11, and 19. Actually, the energy-representation expression for $\Delta\mu$ in a mixed solvent (solvent system involving more than a single species) without spatial decomposition[47,49] has the form of Eqs. 25 and 26. In other words, distinct regions in space can be viewed as distinct species of a mixed solvent in the energy-representation formalism.

Equations 25 and 26 are exact, and the structure of the latter is similar to those of Eqs. 7, 11, and 19. The integration over the coupling parameter λ is involved in the third term of Eq. 26, and needs to be approximated to formulate an endpoint expression. In the present work, we employ the combined PY-type and HNC-type relationship as adopted in previous works to approximate the third term of Eq. 11 in the (one-dimensional) energy-representation method.[44,47,49] The approximate expression for $\Delta\mu$ is given in Appendix A.

At the beginning of the present section, we suppose that the solute and solvent are rigid. This supposition is actually not necessary. When the solute and/or solvent is flexible, our

developments are valid when all the intramolecular degrees of freedom are incorporated into **x**. The problem is then the high dimensionality of **x**, whereas in the energy representation introduced by Eq. 8, the coordinate of the distribution functions is kept one-dimensional even with intramolecular degrees of freedom of the solute and/or solvent. This is a useful feature of the energy-representation formalism, and ϵ in the mixed space-energy representation serves as a proxy for all the variables other than the position, including the orientation and the intramolecular flexibility. The dimension is at least 2 for the orientation, and grows for the intramolecular part with the number of intramolecular degrees of freedom in the solute and/or solvent molecule. Equations 8, 12, and 20 are thus marginal distributions of Eq. 3 over coordinate sets with (largely) reduced dimension since many (usually infinite) orientational and intramolecular configurations are projected to a single value of ϵ.

Methods

The solvent in the present work was TIP3P water,[54] and the solute species were ethane, methylamine, methanol, toluene, aniline, phenol, and alanine dipeptide. The force field for the solutes was OPLS-AA/L,[55,56] and in the following, those solutes except for alanine dipeptide are called small solutes. The N- and C-termini of alanine dipeptide were capped with -CO-CH_3 and -NH-CH_3, respectively, and the peptide was treated as a neutral species. The system configurations were generated through all-atom molecular dynamics (MD) simulation using GROMACS 2016.3,[57,58] and to obtain the solvation free energy $\Delta\mu$, three kinds of MD were conducted for each solute species: the solution system of interest, pure solvent (pure water), and an isolated solute in vacuum. The force fields were the same among the three systems, and an isothermal MD of the isolated solute was performed at 300 K to prepare a set of intramolecular configurations. The solution and pure solvent were simulated in the isothermal-isobaric (*NPT*) ensemble at 300 K and 1 bar with the periodic boundary condition and minimum image convention. The unit cell of MD was then cubic, and contained 2000 water molecule with a single solute for the solution. The simulation length was 5 and 2 ns for the solution and pure solvent, respectively, with the sampling interval of 0.1 and 1 ps. See Appendix B for the detailed procedures of MD simulation.

4 conformational states were examined for alanine dipeptide by using a restraining potential, whereas the small solutes were simulated without any restraint. The structure of alanine dipeptide is shown in Fig. 2 with the dihedral angles ϕ and ψ, and the conformations are denoted as follows.[59-70]

$$\begin{aligned}
&\alpha_{\mathrm{R}}: \quad \phi = -65^{\circ}, \quad \psi = -45^{\circ} \qquad (27)\\
&\mathbf{P}_{\mathrm{II}}: \quad \phi = -65^{\circ}, \quad \psi = 145^{\circ}\\
&\alpha_{\mathrm{L}}: \quad \phi = 45^{\circ}, \quad \psi = 65^{\circ}\\
&\mathbf{C}_7^{\mathrm{ax}}: \quad \phi = 55^{\circ}, \quad \psi = -85^{\circ}
\end{aligned}$$

Alanine dipeptide was restrained by using flat-bottomed potential functions operative on ϕ and ψ; the restraint was harmonically active with a force constant of 0.1 kcal/mol/degree^2

when the dihedral angle deviates from the corresponding reference listed in Eq. 27 by more than 10°. The intramolecular contribution to the relative stability of each conformation was also determined by performing an replica-exchange MD of alanine dipeptide isolated in vacuum, as described in Appendix B.

Results and Discussion

Small Solutes

Figure 3 shows the radial distribution functions g(r) of water around the small solutes examined; the abscissa r refers to the distance of the oxygen site of a solvent water molecule from the methyl carbon (C), amine nitrogen (N), hydroxyl oxygen (O), or the center of mass of the 6 carbon atoms forming the phenyl ring (Ph). A sharper first peak appears at a shorter distance for N and O than for C and Ph. This simply reflects the hydrogen bonding with water, and the peak is stronger around O than around N. A clear second peak is present for N of methylamine and O of methanol. Aniline and phenol have a bulky phenyl group next to the amine or hydroxyl group, and the peak structure of g(r) is less distinct. The major peak for Ph is located broadly at ~5 Å, with a small shoulder observed only for toluene at ~3 Å. Ph is taken to be the center of the phenyl ring, and with this setup, the water coordination is seen mainly at farther distances than in the cases of atomic C, N, and O. Actually, g(r) around C or Ph is affected by the interaction of the neighboring site with water. The first peak for C is weaker around ethane than around methylamine and methanol, and the major peak for Ph exhibits a similar tendency.

Let R be the minimum of the distances of the C, Ph, N and O sites in the solute with the oxygen site of the solvent water molecule. On the basis of this R and Fig. 3, we divide the space into a set of regions as listed in Table 1. The solute molecule has two sites for which distances with a solvent molecule are considered, and the space is first divided into two regions each of which is closer to one of the sites. The two regions are then divided further in terms of the values of R, and the number of regions is 9 for ethane and toluene and 10 for methylamine, methanol, phenol, and aniline. The scheme of division is actually similar to that in Fig. 1, and as illustrated in Fig. 1, the regions in Table 1 satisfy Eq. 21 and are not spherical. The region with the smallest R for each site corresponds to the excluded volume. It is the spatial region in which water overlaps with a solute atom and the solute-solvent interaction energy is prohibitively large. Upon dissolution of the solute, a number of solvent molecules need to be displaced from the region to be occupied by the solute, and the free-energy penalty corresponding to this displacement is denoted as the excluded-volume effect. The second and third regions of R around each site in Table 1 refer to the first and second shells, respectively. A small shoulder is present at ~3 Å for Ph of toluene, and is treated as part of the first shell in Table 1. The far-separated region is where the solvent molecule is separated by more than 10 Å from all of the C, Ph, N, and O sites in the solute, and it is defined as a single region with $R > 10$ Å. A notable feature of our spatial-decomposition scheme is that it is not necessary to introduce a region with respect to the position of a solute atom. Ph is the center-of-mass site of 6 carbon atoms, for example, and does not correspond to the position of a specific atom. A “coarse-grained” unit such as the phenyl group may be employed for the decomposition, and the regions can be defined in accordance with the

spatial resolution of interest. We divided the space with Table 1 to highlight the roles of such atomic groups as methyl, phenyl, amine, and hydroxyl.

The contribution $\Delta\mu(i)$ from region i to the solvation free energy was evaluated from Eq. 26 with the approximate form in Appendix A. It should be noted that the volume is different among the regions, and to correct the effect of this difference, we also computed the free-energy density $\Delta\mu(i)/v(i)$. $v(i)$ refers to the volume of region i and was computed at test-particle insertion of the solute of which the intramolecular motions were sampled in vacuum (see METHODS section). Figure 4 shows $\Delta\mu(i)/v(i)$ for the small solutes examined; the numerical value of $\Delta\mu(i)/v(i)$ is listed in Supporting Information. It is evident that $\Delta\mu(i)/v(i)$ is unfavorable (positive) in the excluded-volume region for each site. The solvent molecules need to be displaced from that region to avoid the overlap with the solute, and this displacement corresponds to the unfavorable density of free energy in Fig. 4. Actually, $\Delta\mu(i)/v(i)$ in the excluded-volume region is weakly dependent on the site and the solute species. The free-energy density in the excluded volume is not affected strongly by the chemical structure of the solute, though it is larger for Ph. When the solvent water is located outside the excluded-volume region, $A\mu(i)/v(i)$ is favorable (negative) and reduces in magnitude toward outer regions except at the Ph site of phenol. The $A\mu(i)/v(i)$ value in the first shell is distinctively more favorable around the amine N and the hydroxyl O than around the methyl C and the phenyl Ph, as expected from the hydrogen bonding with water. The free-energy density is comparable in magnitude only between the excluded-volume regions and the first shells of N and O, and is smaller in the other regions.

The negative $A\mu(i)/v(i)$ signifies the attractive interaction between the solute and the solvent water, and to see the repulsive and attractive effects in more detail, we further examine the energetics on the basis of Eq. 26. Its first term, denoted as $u(i)$ hereafter, is the average sum of the interaction energy of the solute with the solvent molecules located in region i in the solution system of interest, and the rest of the right-hand side of Eq. 26 describes the free-energy change due to the solvent reorganization. Figure 4 then depicts the energy density $u(i)/v(i)$; see Supporting Information for the numerical value of $u(i)/v(i)$. The energy density vanishes in the excluded-volume region since there is no solvent in that region when the solute-solvent interaction is operative in the solution system. $u(i)/v(i)$ is favorable, on the other hand, throughout the regions outside the excluded volume. The magnitude of $u(i)/v(i)$ decreases in outer shells, reflecting simply that the intermolecular interaction is weaker at farther distances. When compared among the first- or second-shell contribution, $A\mu(i)/v(i)$ in Fig. 4 depends on the site and solute species in fair correspondence with $u(i)/v(i)$, as shown in the correlation plot in Supporting Information. The preference order is parallel between $A\mu(i)/v(i)$ and $u(i)/v(i)$ within each of the first and second shells, and the distinction between the first shell and outer regions will be seen next with regard to the solvent reorganization; $(A\mu(i) - u(i))$ is equal to the second and third terms of Eq. 26 and is called solvent-reorganization term.

Figure 4 shows $A\mu(i)/v(i) \geq u(i)/v(i)$, and accordingly, the solvent-reorganization term is unfavorable (positive). The effect of solvent reorganization is particularly strong in the region adjacent to the solute. $A\mu(i)/v(i)$ is less than half of $u(i)/v(i)$ in magnitude in the first shell of each site, and the direct interaction between the neighboring solute and solvent

molecules is negated by more than half extent due to the solvent reorganization. Actually, the solvent-reorganization term diminishes with the distance from the solute faster than $u(i)/v(i)$. Around the C and Ph sites, $\Delta\mu(i)/v(i)$ is significantly different from $u(i)/v(i)$ only in the excluded volume and first shell and $\Delta\mu(i)/v(i) \simeq u(i)/v(i)$ holds in the second-shell and outer regions. The contribution from the solvent reorganization to the free energy vanishes essentially beyond the first shell of the nonpolar site. Around the N and O sites, the difference between $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ persists to longer distances. It is appreciable in the second shell, indicating that the solvent-reorganization effect in the solvation free energy is not localized when the solute-solvent interaction is strong.[4] The decays of the direct-interaction and solvent-reorganization terms are further discussed at the end of Appendix A on the basis of the mathematical expression for the free-energy functional.

According to Fig. 4, $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ at the methyl C site are more favorable (more negative) for methylamine and methanol than for ethane. This is consistent with a structural result in Fig. 3, and indeed, the radial distribution function around C has a stronger peak structure when the polar amine N or hydroxyl O is bonded than when the nonpolar methyl group is attached. At the phenyl Ph site, on the other hand, $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ in the first shell are larger in magnitude for toluene than for aniline and phenol. It is considered that the strengthened energetics for toluene corresponds to the shoulder present at short distances. Around the N and O sites, $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ are more favorable in the first shell for methylamine and methanol than for aniline and phenol and are less in the second shell and outer regions. The favorable interaction of N or O with water is observed at farther distances near a bulky phenyl group.

$\Delta\mu(i)$ and $u(i)$ are depicted in Fig. 5. Their signs are of course coincident with those of $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$ in Fig. 4. In the excluded-volume region, the dependence of $\Delta\mu(i)$ on the site and solute species is governed by the volume of the region $v(i)$ since it was observed in Fig. 4 that $\Delta\mu(i)/v(i)$ is almost constant throughout the excluded-volume regions. Within methylamine, methanol, aniline, or phenol, $v(i)$ of the excluded volume is larger around C or Ph than around N or O as shown in Supporting Information, and the excluded-volume contribution to the solvation free energy $\Delta\mu(i)$ is correspondingly larger at the nonpolar site. Outside the excluded-volume region, $\Delta\mu(i)$ and $u(i)$ decay more slowly with the separation from the solute than the densities $\Delta\mu(i)/v(i)$ and $u(i)/v(i)$, reflecting the fact that $v(i)$ increases toward outer regions. At the C and Ph sites, the sum of $\Delta\mu(i)$ over the first shell, second shell, and outer region is overwhelmed by the excluded-volume $\Delta\mu(i)$. Although the former sum contributes favorably to the solvation free energy, it cancels the latter only partially. The excluded-volume contribution is overturned by the first-shell contribution for N of methylamine and for O of methanol and phenol, and it is negated by the sum of the first- and second-shell contributions for N of aniline. A favorable (negative) free energy of solvation is thus carried by the amine or hydroxyl group for the polar compounds examined in the present work.

According to Fig. 5, the contribution to the solvation free energy is appreciable also from the second shell. This is due to the factor of $v(i)$, and the solute-solvent energetics is not spatially localized in the first solvation shell. In Ref. 4, the excess partial molar energy was addressed for a series of ions in water and the contribution from the proximate region of the

solute was examined by taking its correlation to the total value of the excess energy. Figure 6 provides the correlation plot of the total sum of $\Delta\mu(i)$ over all the regions around each solute against the partial sums over the excluded-volume and first-shell regions and over the excluded-volume to second-shell regions. The preference order of the total sum is well reproduced by the partial sum to the second shell. The difference of the partial sum from the total is positive for all of the small solutes and amounts to 1–2 kcal/mol. This amount of attractive interaction is spread in the outer regions beyond the second shell, where the solvent-reorganization term is small as seen in Fig. 5 and the tables for the small solutes in Supporting Information. The direct interaction between the solute and solvent outside the second shell thus needs to be taken into account to quantitatively evaluate the solvation free energy. The correlation to the total sum of $\Delta\mu(i)$ is worse for the partial sum over the excluded-volume and first-shell regions. The partial sum to the first shell correlates to the total only separately over ethane, methylamine, and methanol and over toluene, aniline, and phenol. Even the sign of the total sum is often not reproduced by the partial sum, which means that the first-shell contribution is not always enough to overturn the unfavorable contribution from the excluded volume. Figure 6 thus shows that the solvation free energy is not localized in the excluded-volume to second-shell regions, while the preference order of solvation is governed by the contributions to the second shell. The free-energetics of solvation can be described by the interactions within the second nearest neighbor through correlations.

Alanine Dipeptide

We now turn to alanine dipeptide. Figure 7 depicts the radial distribution functions $g(r)$ of water around the C_1, N_1, C_2, C_3, and N_2 atoms in Fig. 2 at the 4 conformations listed in Eq. 27 and Fig. 2. The peak structure is clearer around C_1 and C_3 than around C_2. This corresponds to the amide structure in alanine dipeptide, and the N_1 and N_2 atoms may stay at ~3 Å due to the hydrogen bonding between the amine and water. It is further seen among the 4 conformations that the hydration structure is least distinct for C_7^{ax}. At the C_7^{ax} conformation, the carbonyl oxygen at C_1 forms an intramolecular hydrogen bond with the amine hydrogen at N_2 and the coordination of a water molecule is prohibited with the two amide groups. A peak is not present at ~3 Å around N_2, in particular, and the first peaks are also weak for C_1 and C_3. When α_R and α_L are compared, the latter has the first peaks of C_2 and C_3 at shorter distances. The steric hindrance is weaker for α_L, and in addition, the peak structure is strongest for P_{II} around N_2.

We then divide the space on the basis of Figs. 2 and 7. R is the minimum of the distances of the C_1, N_1, C_2, C_3, and N_2 atoms in alanine dipeptide with the oxygen site of the solvent water molecule, and the space is divided similarly to those in Fig. 1 and in Table 1 for the small solutes. 5 regions are first introduced by identifying the closest one of the 5 atoms, and as shown in Table 2, the division is done further around C_1, N_1, C_2, C_3, and N_2 in terms of R. The number of regions is then 26 for all of the 4 conformations. The excluded volume, first shell, and second shell are introduced around each of C_1, N_1, C_2, C_3, and N_2, and the solvent molecule is far-separated when $R > 10$ Å holds from all of those atoms. In the second paragraph of “Small Solutes” subsection, we noted that it is not necessary to set the spatial region with respect to the position of a solute atom. We remark further at this point

that only a partial set of solute atoms may be employed to define the regions. The main-chain atoms of C_1, N_1, C_2, C_3, and N_2 were chosen for the analyses in the present subsection since they specify the ϕ and ψ angles in Fig. 2 and we examine the dependence on the conformation identified by ϕ and ψ through Eq. 27. The space is still divided uniquely, as noted with respect to Eq. 21. Each position in space belongs to one and only one of the regions, and there is no double-counting. For example, any position within the excluded-volume and first-shell regions around the methyl groups bonded to C_1, C_2, and N_2 is contained in one of the regions in Table 2. Since these methyl groups are not employed to introduce the spatial regions in Table 2, they are not referenced when a position in space is assigned to a region in the table.

Figure 8 shows $\Delta\mu(i)/\nu(i)$ and $u(i)/\nu(i)$ for alanine dipeptide at the 4 conformations of Eq. 27; the numerical values of $\Delta\mu(i)/\nu(i)$ and $u(i)/\nu(i)$ are provided in Supporting Information. As was observed for the small solutes in Fig. 4, $\Delta\mu(i)/\nu(i)$ in the excluded-volume region is unfavorable (positive) for each of the C_1, N_1, C_2, C_3, and N_2 atoms and depends weakly on the atomic site and the conformation of alanine dipeptide. $u(i)/\nu(i)$ is favorable outside the excluded volume, on the other hand, and its magnitude decreases monotonically with the separation from the solute atom. Actually, $\Delta\mu(i)/\nu(i)$ is monotonic only in the second-shell and outer regions when seen throughout the sites and conformations. The first shells of the C_1, C_2, and N_2 atoms overlap with the excluded-volume regions of the neighboring methyl groups, and these overlaps contribute unfavorably to the first-shell $\Delta\mu(i)/\nu(i)$ for those atoms (note that any position around the methyl groups belongs uniquely to one of the regions in Table 2 and not to a distinct region other than those in the table, as mentioned in the preceding paragraph).

$\Delta\mu(i)/\nu(i)$ and $u(i)/\nu(i)$ are distinctively favorable in the first shells of the N_1 and C_3 atoms. A similar tendency was observed in the first shells of the N and O atoms in Fig. 4, and the hydrogen-bond availability with the solvent water is reflected. The region around the C_2 atom is hydrophobic in the sense that $\Delta\mu(i)/\nu(i)$ is positive in its first shell. Indeed, C_2 is the C_α atom of the alanine residue and the unfavorable interaction from the methyl sidechain is reflected in the energetics in the first shell; note that alanine is regarded as a hydrophobic amino acid. When compared over the 5 atomic sites and 4 conformations of alanine dipeptide, the preference order of $\Delta\mu(i)/\nu(i)$ corresponds fairly to that of $u(i)/u(i)$ in the first or second shell, as seen in the correlation plot in Supporting Information. This feature holds separately for the first- and second-shell contributions, and is common to the case for the small solutes described in the preceding subsection. The solvent-reorganization term behaves similarly, too. It is always unfavorable ($A\mu(i)/\nu(i) \geq u(i)/u(i)$) and is large in the first solvation shell. The decay with the separation from the solute is faster than of $u(i)/\nu(i)$, and the effect is small beyond the second shell.

Figure 9 depicts $\Delta\mu(i)$ and $u(i)$ of alanine dipeptide. In the excluded-volume region, $\Delta\mu(i)$ is larger at the terminal atoms of C_1 and N_2 than of N_1, C_2, and C_3 due to the volume size $\nu(i)$. Outside the excluded-volume region, $u(i)$ reduces monotonically in magnitude with the distance from the solute around the C_1, N_1, and C_3 sites. The monotonic dependence is observed only beyond the second shell around the C_2 and N_2 atoms, on the other hand, and the steric hindrance due to the neighboring methyl group is evidenced there as a factor to

weaken the interaction with the first-shell water. Actually, $\Delta\mu(i)$ is non-monotonic also around C_1. Although Fig. 8 shows that the magnitude of $\Delta\mu(i)/v(i)$ is larger in the first shell than in the second shell except for the C_7^{ax} conformation, the larger $v(i)$ of the second shell leads to the less favorable $\Delta\mu(i)$ in the first shell. The sum of $\Delta\mu(i)$ over the excluded-volume to outer regions is plotted for each atom in Fig. 10. At the C_2 site, the sum is unfavorable (positive) and the contribution from the excluded volume is larger in magnitude than that from the rest. The region around C_2 is thus hydrophobic as a whole, while the other atoms provide hydrophilic environments with the negative sums of $\Delta\mu(i)$ over the excluded-volume to outer regions.

To address the extent of localization of the solvation free energy, the correlation plot is provided in Fig. 11 for the $\Delta\mu(i)$ sum over all the regions around alanine dipeptide against the partial sums over the excluded-volume and first-shell regions of all the atomic sites and over the excluded-volume to second-shell regions. The partial sums correlate well with the total sum. Although they deviate significantly from the total, the conformation dependence of the total free energy of solvation is well reproduced. Actually, the sum of the excluded-volume and first-shell contributions is positive and the sum to the second shell is different from the total by several tens of %. The solvation free energy is thus not spatially localized near the alanine-dipeptide solute in view of its value, whereas a local description is valid to describe the free-energetics of conformational variation by virtue of the correlations evidenced in Fig. 11.

Figure 10 shows the sum of $\Delta\mu(i)$ over the excluded-volume to outer regions for each atom, and when compared among the 4 conformations of α_R, P_{II}, α_L, and C_7^{ax}, it is evident that the $\Delta\mu(i)$ sum depends on the conformation strongly at the C_1, C_3, and N_2 atoms. Although $\Delta\mu(i)/v(i)$ in Fig. 8 is more favorable in the first shells of N_1 at the α_R and α_L conformations than at P_{II} and C_7^{ax}, the corresponding volumes $v(i)$ are smaller at the former conformations and the contribution from N_1 is minor in determining the conformation dependence of the total free energy of solvation. The hydrophobic region around the C_2 atom also makes a minor contribution. Figure 9 shows that $\Delta\mu(i)$ around C_2 depends weakly on the conformation in each region and does not reflect the conformational difference. When the contributions from C_1, C_3, and N_2 are compared, the excluded-volume $\Delta\mu(i)$ are similar among the 4 conformations and the difference in the $\Delta\mu(i)$ sum for each of C_1, C_3, and N_2 comes from the first-shell and second-shell contributions. The total sums of $\Delta\mu(i)$ and $u(i)$ (the sums over all the regions and the atoms) are the least favorable (most positive) with C_7^{ax} among the conformations examined, and this is due to the contributions from the C_1 and N_2 atoms. Indeed, the C_7^{ax} conformation has an intramolecular hydrogen bond between the carbonyl oxygen at C_1 and the amine hydrogen at N_2 and the hydration structures around C_1 and N_2 are weakest as was seen in Fig. 7. The most favorable conformation is α_L, for which the C_1 and C_3 contributions are larger in magnitude than for the others. The carbonyl groups on C_1 and C_3 are the closest for α_L, and the interaction of a water molecule that is proximate to C_1 or C_3 is further strengthened by the interaction with C_3 or C_1, respectively; note that the interaction energy of a solvent molecule with the whole solute is counted when the

spatial region is assigned through Eq. 20 to construct the distribution function. Still, the hydration structures represented by the radial distribution functions in Fig. 7 are not distinct at the α_L conformation. The orientational features are averaged out in Fig. 7, while they are sensitively reflected in the energetics shown in Figs. 8, 9, and 10. The preference order of the α_R and P_{II} conformations is governed by the balance between the C_3 and N_2 contributions. Although the free-energy density $\Delta\mu(i)/v(i)$ and the solute-solvent energy density $u(i)/v(i)$ around the C_3 atom are comparable between the two conformations in Fig. 8, the C_3 contributions in Fig. 9 are more favorable for α_R than for P_{II} since the space assigned to C_3 is larger with the former.

According to Fig. 10, the total value of the solvation free energy (total sum of $\Delta\mu(i)$) is in the order of $\alpha_L < \alpha_R < P_{II} < C_7^{ax}$. The interaction with water stabilizes the α_L conformation most, and to determine the overall stabilities of the 4 conformations, the intramolecular contribution needs to be added. The preference order is then $P_{II} > \alpha_R(1.2) > \alpha_L(5.3) > C_7^{ax}(6.9)$ by counting both the intra- and intermolecular effects, where the value in parenthesis is the free-energy difference relative to P_{II} in kcal/mol. When the solvation free energy is computed without spatial decomposition through Eq. 11, the order is $P_{II} > \alpha_R(1.3) > \alpha_L(5.4) > C_7^{ax}(6.8)$, and when the exact calculation is done as described in Supporting Information, we have $P_{II} > \alpha_R(0.7) > \alpha_L(4.7) > C_7^{ax}(6.4)$. The order of conformational preference is thus unchanged with the spatial decomposition and the approximation of the energy-representation method.

Conclusion

A spatial-decomposition formula was developed for the solvation free energy. The density-functional method was adopted in the mixed representation of position and energy, and a free-energy functional was formulated with distribution functions of the relative position and pair interaction energy of the solute with the solvent. The free energy of solvation was then analyzed in solvent water for small molecules with methyl, amine, or hydroxyl group and for alanine dipeptide at 4 conformations. The radial distribution functions were employed to define the excluded-volume, first-shell, second-shell, and outer regions around a set of sites within the solute molecule, and it was observed that the density of the solvation free energy in the excluded-volume region depends weakly on the site and on the species or structure of the solute. In the first or second shell, the preference order is in fair correspondence between the free-energy density and the density of the direct interaction between the solute and solvent. The difference of the solvation free energy from the solute-solvent energy is the solvent-organization term, and it reduces in magnitude toward outer regions faster than the direct interaction. The extent of spatial localization of the free-energetics of solvation was examined by comparing the total free energy of solvation with the partial sum consisting of the excluded-volume, first-shell, and second-shell contributions. It was seen that the solvation free energy is not spatially localized in the sense that the partial sum deviates significantly from the total value. Still, a good correlation was found between the partial and total free energies and a local description of the solvation free energy can be justified on the basis of the correlation.

In the present paper, we formulated the spatial-decomposition analysis of the solvation free energy and provided illustrative results for small solutes and alanine dipeptide in solvent water. The formulation is based on the density-functional method over the position and energy coordinates, and is parallel to that in the one-dimensional energy-representation formalism for a mixed solvent (solvent system with more than a single species) when the space is divided into discrete regions. Actually, the division scheme is arbitrary and may be prepared according to the analysis contents sought. A set of atomic groups can be introduced, for example, to divide the space in terms of the closest distance within the set. In such a case, the solvation free energy is expressed as a sum of the contributions from the atomic groups and is in formal correspondence to the accessible-surface-area scheme.[29] As done in the analysis of alanine dipeptide, furthermore, it is possible to adopt only the atomic sites that are used to describe the conformation.

The key equation for the spatial-decomposition analysis is Eq. 26, and its first and second terms (those without the integration over the coupling parameter λ) is rewritten as

$$\int d\epsilon \epsilon \rho^d(\epsilon;i) - k_B T \int d\epsilon \left[\rho^d(\epsilon;i) - \rho_0^d(\epsilon;i) - \rho^d(\epsilon;i) \log\left(\frac{\rho^d(\epsilon;i)}{\rho_0^d(\epsilon;i)} \right) \right] \quad (28)$$

$$= -k_B T \int d\epsilon \left(\rho^d(\epsilon;i) - \rho_0^d(\epsilon;i) \right) - \int d\epsilon \rho^d(\epsilon;i) \omega^d(\epsilon;i)$$

$$= -k_B T \left(N^d(i) - N_0^d(i) \right) - N^d(i) \langle \omega^d \rangle_i$$

where $\omega^d(\epsilon; i)$ denotes the indirect part of potential of mean force in the solution system ($\lambda = 1$), $\langle \omega^d \rangle_i$ is its average in region i, and $N^d(i)$ and $N_0^d(i)$ are the average numbers of solvent molecules contained in region i for the solution and the pure solvent ($\lambda = 0$), respectively; the indirect part of potential of mean force was defined by Eq. 23 and ω_1^d is written as ω^d for notational brevity. In a previous work,[43] we discussed the role of interfacial water in protein-ligand binding and categorized the thermodynamic signature of water in terms of the local density and the indirect part of potential of mean force in the 6-dimensional representation over the position and orientation. A lower-dimensional alternative for the signature may be proposed on the basis of Eqs. 26 and 28, that are formulated by adopting the energy as a proxy for the orientation and are numerically easier to handle. As noted in INTRODUCTION, the spatially resolved energetics of water is considered to be useful for ligand design,[17-28,31,32,43] and the present work employed the DFT in the mixed representation of position and energy to elucidate which region of solvent (water) leads to stabilization or destabilization of the solute. It is thus expected that the spatial-decomposition extension of the energy-representation formalism can be a basis for addressing the role of water toward modification of ligand affinities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work is supported by the Grants-in-Aid for Scientific Research (Nos. 15K13550 and 17J01006) from the Japan Society for the Promotion of Science, by the Elements Strategy Initiative for Catalysts and Batteries and the Post-K Supercomputing Project from the Ministry of Education, Culture, Sports, Science, and Technology, by an NIH grant (GM30580), and by an NSF grant (1665032). The simulations were conducted partly using ITO at Kyushu University, OCTOPUS at Osaka University, and the HPCI systems of CX400 at Nagoya University and the K computer at RIKEN Advanced Institute for Computational Science through the HPCI System Research Project (Project IDs: hp180030 and hp180173).

Appendix A: Approximate functional for free energy

The approximate form for the solvation free energy $\Delta\mu$ is given by a set of definitions and equations listed as

$$\omega^d(\epsilon;i) = -k_BT\log\left(\frac{\rho^d(\epsilon;i)}{\rho_0^d(\epsilon;i)}\right) - \epsilon \quad (29)$$

$$\chi_0^d(\epsilon;i,\eta;j) = \langle\hat{\rho}^d(\epsilon;i)\hat{\rho}^d(\eta;j)\rangle_0 - \langle\hat{\rho}^d(\epsilon;i)\rangle_0\langle\hat{\rho}^d(\eta;j)\rangle_0 \quad (30)$$

$$\sigma_0^d(\epsilon;i) = -k_BT\frac{\rho^d(\epsilon;i)-\rho_0^d(\epsilon;i)}{\rho_0^d(\epsilon;i)} + k_BT\sum_j\int d\eta\left(\chi_0^d\right)^{-1}(\epsilon;i,\eta;j)\left(\rho^d(\eta;j)-\rho_0^d(\eta;j)\right) \quad (31)$$

$$\beta\int_0^1 d\lambda\int d\epsilon\omega_\lambda^d(\epsilon;i)\frac{\partial\rho_\lambda^d(\epsilon;i)}{\partial\lambda} = \alpha^d(\epsilon;i)F^d(\epsilon;i) + \left(1-\alpha^d(\epsilon;i)\right)F_0^d(\epsilon;i) \quad (32)$$

$$F^d(\epsilon;i) = \begin{cases} \beta\omega^d(\epsilon;i) + 1 + \dfrac{\beta\omega^d(\epsilon;i)}{\exp\left(-\beta\omega^d(\epsilon;i)\right)-1} & (\textbf{when}\ \omega^d(\epsilon;i)\le 0) \\ \dfrac{1}{2}\beta\omega^d(\epsilon;i) & (\textbf{when}\ \omega^d(\epsilon;i)\ge 0), \end{cases} \quad (33)$$

$$F_0^d(\epsilon;i) = \begin{cases} -\log\left(1-\beta\sigma_0^d(\epsilon;i)\right) + 1 + \dfrac{\log\left(1-\beta\sigma_0^d(\epsilon;i)\right)}{\left(\beta\sigma_0^d(\epsilon;i)\right.} & (\textbf{when}\ \sigma_0^d(\epsilon;i)\le 0) \\ \dfrac{1}{2}\beta\sigma_0^d(\epsilon;i) & (\textbf{when}\ \sigma_0^d(\epsilon;i)\ge 0), \end{cases} \quad (34)$$

$$\alpha^d(\epsilon;i)=\begin{cases}1 & (\textbf{when}\ \rho^d(\epsilon;i)\geq\rho_0^d(\epsilon;i))\\ 1-\left(\dfrac{\rho^d(\epsilon;i)-\rho_0^d(\epsilon\, i)}{\rho^d(\epsilon;i)+\rho_0^d(\epsilon;i)}\right)^2 & (\textbf{when}\ \rho^d(\epsilon;i)\leq\rho_0^d(\epsilon;i)),\end{cases} \quad (35)$$

where $\omega^d(\epsilon;i)=\omega_1^d(\epsilon;i)$ for the brevity of notation and $\langle\cdots\rangle_0$ means the ensemble average in the pure solvent ($\lambda=0$). $\langle\cdots\rangle_0$ is implemented by virtually placing the solute molecule as a test particle without disturbing the solvent configuration. χ_0^d describes the solvent-solvent correlation at two-body level over the coordinates introduced by the solute at $\lambda=0$, where the instantaneous distribution of Eq. 20 (or Eq. 3, 8, or 12) is constructed for the solvent molecule against the solute virtually present in the system. σ_0^d is the solvent-mediated part of the response function of the solute-solvent distribution to the solute-solvent interaction at $\lambda=0$, and the procedure for (pseudo-)inverting χ_0^d is described later with respect to Eq. 42. The first lines of Eqs. 33 and 34 are the PY-type expressions, and the second lines are HNC-type. Equations 33 and 34 are the combined PY-type and HNC-type approximations written in terms of $\omega^d(\epsilon;i)$ and $\sigma_0^d(\epsilon;i)$ respectively, and are mixed with the weighting function α^d of Eq. 35. Equations 29-31 and 33-35 define the variables in their left-hand sides, and with Eqs. 25 and 26, Eq. 32 evaluates $\Delta\mu$ approximately through a combined PY-type and HNC-type relationship.

In the currently employed version of the energy-representation method (in one dimension),[44,47] the third term of Eq. 11 is further turned into a simpler form by adopting a specific $u_\lambda(\epsilon)$ with which $\rho_\lambda^e(\epsilon)$ varies linearly with the coupling parameter λ; no approximation is involved at this point. The integration over λ is then performed analytically with PY-type and HNC-type approximations for the λ dependence of the indirect part of the potential of mean force $\omega_\lambda^e(\epsilon)$ of Eq. 10. In the PY-type approximation, $(\exp(-\beta\omega_\lambda^e(\epsilon))-1)$ is taken to be linear with respect to λ, and in the HNC-type, $\omega_\lambda^e(\epsilon)$ is linear with λ. No special treatment is necessary for the direct interaction potential. The value ϵ of the solute-solvent pair interaction energy is adopted as the abscissa for the distribution functions, and the approximations are formulated only in terms of the λ dependence of $\omega_\lambda^e(\epsilon)$. The approximation in the present work is expressed as Eqs. 32-35, and its formulation is also done with the λ dependence of $\omega_\lambda^d(\epsilon;i)$ of Eq. 23.

As in the one-dimensional case,[44,47] $u_\lambda(\epsilon;i)$ is chosen so that $\rho_\lambda^d(\epsilon;i)$ varies linearly with λ. The integral over λ in the third term of Eq. 26 then simplifies to

$$\int d\epsilon\left(\rho^d(\epsilon;i)-\rho_0^d(\epsilon;i)\right)\int_0^1 d\lambda\omega_\lambda^d(\epsilon;i) \quad (36)$$

and the derivative at $\lambda = 0$ is related to $\sigma_0^d(\epsilon; i)$ of Eq. 31 as

$$\left.\frac{\partial \omega_\lambda^d(\epsilon; i)}{\partial \lambda}\right|_{\lambda = 0} = \sigma_0^d(\epsilon; i)\,. \quad (37)$$

The PY-type and HNC-type approximations are formulated by adopting the linear dependencies on λ of $(\exp(-\beta\omega_\lambda^d(\epsilon; i)) - 1)$ and $\omega_\lambda^d(\epsilon; i)$, respectively. When the linear dependence is expressed with the derivative at $\lambda = 0$, the PY-type and HNC-type relationships are given by

$$\exp(-\beta\omega_\lambda^d(\epsilon; i)) = 1 - \beta\lambda\sigma_0^d(\epsilon; i) \quad (38)$$

and

$$\omega_\lambda^d(\epsilon; i) = \lambda\sigma_0^d(\epsilon; i) \quad (39)$$

respectively, since $\omega_\lambda^d(\epsilon; i) = 0$ at $\lambda = 0$ by definition of Eq. 23. When the linearity is written in terms of the value at $\lambda = 1$,

$$\exp(-\beta\omega_\lambda^d(\epsilon; i)) = 1 + \lambda\left(\exp(-\beta\omega^d(\epsilon; i)) - 1\right) \quad (40)$$

and

$$\omega_\lambda^d(\epsilon; i) = \lambda\omega^d(\epsilon; i) \quad (41)$$

are PY-type and HNC-type, respectively, where $\omega_\lambda^d(\epsilon; i)$ at $\lambda = 1$ is denoted as $\omega^d(\epsilon; i)$ of Eq. 29. The integration over λ in Eq. 36 can be carried out analytically with Eqs. 38-41, and the approximate form of Eqs. 32-35 is obtained by using the PY-HNC combination and the weighting function which are parallel to those in the one-dimensional formulation.[44,47]

In our simulation setup, χ_0^d has a non-degenerate, null eigenvalue. Its pseudo-inverse, expressed as $\left(\chi_0^d\right)^{-1}$ in Eq. 31, was then obtained by referring to the procedure in Appendix B of Ref. 46 and Appendix B of Ref. 47. An auxiliary variable $\bar{u}$ was determined up to an additive constant through

$$\sum_j \int d\eta \chi_0^d(\epsilon; i, \eta; j) \bar{u}(\eta; j) = -k_B T \left(\rho^d(\epsilon; i) - \rho_0^d(\epsilon; i) \right), \quad (42)$$

and the additive constant was fixed by setting $\bar{u}(\epsilon, i) = 0$ at the discretized energy coordinate corresponding to $\epsilon = 0$ for the farthest region i from the solute (the far-separated regions in Tables 1 and 2).

It was observed in RESULTS AND DISCUSSION section that the solvent-reorganization term reduces faster with the separation from the solute than the direct-interaction term between the solute and solvent. This observation can be justified on the basis of Eq. 26 by considering that the indirect part of the potential of mean force ω_λ^d decays faster than the direct interaction. ρ^d then converges to ρ_0^d at the rate corresponding to the dependence of the direct interaction on the distance from the solute, given that ρ^d is related to ρ_0^d through Eq. 23 (note that $u_1 = \epsilon$ and $\rho_1^d = \rho^d$). Accordingly, the solvent-reorganization term of Eq. 26 (second and third terms) is of higher order than the first term with respect to $(\rho^d - \rho_0^d)$ since $\left[\rho^d - \rho_0^d - \rho^d \log(\rho^d / \rho_0^d) \right]$ is of second order.

The computed value of the (total) solvation free energy $\Delta\mu$ can be different between Eqs. 11 and 25 when the combined PY-type and HNC-type approximation is introduced. The difference is thus a measure of the performance of the approximate treatment since Eqs. 11 and 25 themselves are exact. The $\Delta\mu$ value obtained through the energy-representation method without spatial decomposition is listed in Supporting Information, as well as the (numerically) exact one from the method of Bennett acceptance ratio.[71,72] When the approximate $\Delta\mu$ with and without the spatial decomposition is compared, the difference is within 0.1 kcal/mol for the small solutes and is 0.4–0.6 kcal/mol for alanine dipeptide. The relative free energies among the 4 conformations still agree within 0.2 kcal/mol. When the comparison is done between the approximate and exact $\Delta\mu$, the deviation is seen to be within 1.1 and 1.6 kcal/mol for the small molecules and for the 4 conformations of alanine dipeptide, respectively.

Appendix B: Detailed procedures of simulation

As noted in METHODS section, MD was performed for the solution system of interest, pure solvent (pure water), and an isolated solute in vacuum. The solution system was simulated in the *NPT* ensemble by locating a single solute and 2000 water molecules in the MD unit cell; the cell size was ~39 Å. The leap-frog stochastic dynamics algorithm was employed to integrate the equation of motion at a time step of 2 fs and an inverse friction constant of 2 ps, [73] and the pressure was regulated by the Parrinello-Rahman barostat at a coupling time of 2 ps and isothermal compressibility of 4.5×10^{-5} bar^{-1} with the isotropic coupling.[74] Each water molecule was kept rigid with SETTLE, and the LINCS method was used for the solute molecule to fix the lengths of all the bonds.[75,76] The electrostatic interaction was handled by the smooth particle-mesh Ewald (PME) scheme at a real-space cutoff of 12 Å, a spline order

of 4, a relative tolerance of 10^{-5} (inverse decay length of 0.26 $Å^{-1}$), and a reciprocal-space mesh size of 48 for each of the *x*, *y*, and *z* directions.[77] The Lennard-Jones (LJ) interaction was truncated with the switching function in a range of 10-12 Å.[78] The geometric mean was adopted both for the energy and length parameters to combine the LJ interaction between unlike pairs of atoms, and the truncation was done on atom-atom basis for LJ as well as for the real-space part of PME. The long-range correction for LJ was not included in MD and was incorporated into the calculations of the solvation free energy and solute-solvent interaction energy by supposing that the solvent water is of bulk distribution beyond 10 Å from any atom within the solute molecule. The pure-solvent system was only of 2000 water molecules, and the other MD setups were identical to those described above for the solution system.

To simulate the isolated solute in vacuum, a single solute molecule was subject to an isothermal MD for 200 ns at a sampling interval of 0.1 ps. In this simulation, the solute center of mass was fixed at the origin and the electrostatic potential was treated as its bare form of $1/r$ without cutoff; the other procedures were the same as those for the solution and pure solvent. The solute was inserted as a test particle into the pure-solvent system at random position and orientation, and the insertion was done without disturbing the solvent configuration after the MD of pure water had been performed independently. The number of insertions was 1000 per pure-water configuration sampled, leading to the generation of 2×10^6 solute-solvent configurations in total for the free-energy calculation. Actually, the MD of the isolated solute was carried out for 200 ns simply to prepare 2×10^6 intramolecular configurations that were sampled at an interval of 0.1 ps and were used for test-particle insertion.

For alanine dipeptide, a replica-exchange MD was also performed in vacuum to determine the intramolecular contribution to the free energy of each conformation. A single molecule was simulated at 9 temperatures of 300, 350, …, 650, and 700 K with an exchange interval of 1 ps,[79] and the simulation length was 4 μs at each temperature. No restraint was employed, and the other simulation setups were the same as those described in the preceding paragraph. The reweighting was then conducted to obtain the probability for finding each conformation of Eq. 27 with allowances of ±10°; for example, the probability of $-75° \le \phi \le -55°$ and $-55° \le \psi \le -35°$ was evaluated for α_R.

References

(1). Matubayasi N; Reed LH; Levy RM Thermodynamics of Hydration Shell. 1. Excess Energy of a Hydrophobic Solute. J. Phys. Chem. 1994, 98, 10640–10649.

(2). Matubayasi N; Levy RM Thermodynamics of Hydration Shell. 2. Excess Volume and Compressibility of a Hydrophobic Solute. J. Phys. Chem. 1996, 100, 2681–2688.

(3). Matubayasi N; Gallicchio E; Levy RM On the Local and Nonlocal Components of Solvation Thermodynamics and Their Relation to Solvation Shell Models. J. Chem. Phys. 1998, 109, 4864.

(4). Mogami G; Suzuki M; Matubayasi N Spatial-Decomposition Analysis of Energetics of Ionic Hydration. J. Phys. Chem. B 2016, 120, 1813. [PubMed: 26784859]

(5). Abseher R; Schreiber H; Steinhauser O The Influence of a Protein on Water Dynamics in Its Vicinity Investigated by Molecular Dynamics Simulation. Proteins: Structure, Function, and Genetics 1996, 25, 366–378.

(6). Bizzarri AR; Cannistraro S Molecular Dynamics of Water at the Protein-Solvent Interface. J. Phys. Chem. B 2002, 106, 6617–6633.

(7). Heyden M; Sun J; Funkner S; Mathias G; Forbert H; Havenith M; Marx D Dissecting the THz Spectrum of Liquid Water from First Principles via Correlations in Time and Space. Proc. Natl. Acad. Sci. USA 2010, 107, 12068–12073. [PubMed: 20566886]

(8). Kubota Y; Yoshimori A; Matubayasi N; Suzuki M; Akiyama R Molecular Dynamics Study of Fast Dielectric Relaxation of Water around a Molecular-Sized Ion. J. Chem. Phys. 2012, 137, 224502. [PubMed: 23249012]

(9). Śmiechowski M; Forbert H; Marx D Spatial Decomposition and Assignment of Infrared Spectra of Simple Ions in Water from Mid-Infrared to THz Frequencies: Li+(aq) and F−(aq). J. Chem. Phys. 2013, 139, 014506. [PubMed: 23822313]

(10). Tu K-M; Ishizuka R; Matubayasi N Spatial-Decomposition Analysis of Electrical Conductivity in Concentrated Electrolyte Solution. J. Chem. Phys. 2014, 141, 044126. [PubMed: 25084900]

(11). Jones RE; Ward DK; Templeton JA Spatial Resolution of the Electrical Conductance of Ionic Fluids Using a Green-Kubo Method. J. Chem. Phys. 2014, 141, 184110. [PubMed: 25399135]

(12). Tu K-M; Ishizuka R; Matubayasi N Spatial-Decomposition Analysis of Electrical Conductivity in Ionic Liquid. J. Chem. Phys. 2014, 141, 244507. [PubMed: 25554167]

(13). Śmiechowski M; Sun J; Forbert H; Marx D Solvation Shell Resolved THz Spectra of Simple Aqua Ions - Distinct Distance- and Frequency-Dependent Contributions of Solvation Shells. Phys. Chem. Chem. Phys. 2015, 8323–8329. [PubMed: 25579399]

(14). Tu K-M; Kim K; Matubayasi N Spatial-Decomposition Analysis of Viscosity with Application to Lennard-Jones Fluid. J. Chem. Phys. 2018, 148, 094501.

(15). Pluhařová E; Jungwirth P; Matubayasi N; Marsalek O Structure and Dynamics of the Hydration Shell: Spatially Decomposed Time Correlation Approach. J. Chem. Theory Comput. 2019, 15, 803–812. [PubMed: 30537825]

(16). Matubayasi N Spatial-Decomposition Analysis of Electrical Conductivity. Chem. Rec. in press, DOI: 10.1002/tcr.201800116.

(17). Lazaridis T; Paulaitis ME Entropy of Hydrophobic Hydration: A New Statistical Mechanical Formulation. J. Phys. Chem. 1992, 96, 3847–3855.

(18). Lazaridis T Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. J. Phys. Chem. B 1998, 102, 3531–3541.

(19). Li Z; Lazaridis T The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. J. Phys. Chem. B 2005, 109, 662–670. [PubMed: 16851059]

(20). Li Z; Lazaridis T Water at Biomolecular Binding Interfaces. Phys. Chem. Chem. Phys. 2007, 9, 573–581. [PubMed: 17242738]

(21). Young T; Abel R; Kim B; Berne B; Friesner R Motifs for Molecular Recognition Exploring Hydrophobic Enclosure in Protein-Ligand Binding. Proc. Natl. Acad. Sci. USA 2007, 104, 808–813. [PubMed: 17204562]

(22). Abel R; Young T; Farid R; Berne B; Friesner R Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. J. Am. Chem. Soc. 2008, 130, 2817–2831. [PubMed: 18266362]

(23). Pearlstein R; Hu Q; Zhou J; Yowe D; Levell J; Dale B; Kaushik V; Daniels D; Hanrahan S; Sherman W; Abel R New Hypotheses about the Structure-Function of Proprotein Convertase Subtilisin/Kexin Type 9: Analysis of the Epidermal Growth Factor-like Repeat a Docking Site Using WaterMap. Proteins: Struct., Funct., Genet. 2010, 78, 2571–2586. [PubMed: 20589640]

(24). Snyder P; Mecinovic J; Moustakas D; Thomas I,S; Harder M; Mack E; Lockett M; Héroux A; Sherman W; Whitesides G Proc. Natl. Acad. Sci. USA 2011, 108, 17889–17894. [PubMed: 22011572]

(25). Li Z; Lazaridis T Computing the Thermodynamic Contributions of Interfacial Water. Methods Mol. Biol. 2012, 819, 393–404. [PubMed: 22183549]

(26). Nguyen C; Cruz A; Gilson M; Kurtzman T Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. J. Chem. Theory Comput. 2014, 10, 2769–2780. [PubMed: 25018673]

(27). Huggins D Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. Biophys. J. 2015, 108, 928–936. [PubMed: 25692597]

(28). Persson F; Söderhjelm P; Halle B The Geometry of Protein Hydration. J. Chem. Phys. 2018, 148, 215101. [PubMed: 29884063]

(29). Ooi T; Oobatake M; Némethy G; Scheraga HA Accessible Surface Areas as a Measure of the Thermodynamic Parameters of Hydration of Peptides. Proc. Natl. Acad. Sci. USA 1987, 84, 3086–3090. [PubMed: 3472198]

(30). Hirata F, Ed. Molecular Theory of Solvation; Kluwer Academic Publishers: Dordrecht, Netherlands, 2003.

(31). Chong SH; Ham S Distinct Role of Hydration Water in Protein Misfolding and Aggregation Revealed by Fluctuating Thermodynamics Analysis. Acc. Chem. Res. 2015, 48, 956–965. [PubMed: 25844814]

(32). Silveira RL; Stoyanov SR; Gusarov S; Skaf MS; Kovalenko A Supramolecular Interactions in Secondary Plant Cell Walls: Effect of Lignin Chemical Composition Revealed with the Molecular Theory of Solvation. J. Phys. Chem. Lett. 2015, 6, 206–211. [PubMed: 26263115]

(33). Henderson D, Ed. Fundamentals of Inhomogeneous Fluids; Marcel Dekker: New York, 1992.

(34). Hansen J-P; McDonald IR Theory of Simple Liquids, 4th ed.; Academic Press: Oxford, 2013.

(35). Matubayasi N; Nakahara M Theory of Solutions in the Energetic Representation. I. Formulation. J. Chem. Phys. 2000, 113, 6070–6081.

(36). Gendre L; Ramirez R; Borgis D Classical Density Functional Theory of Solvation in Molecular Solvents: Angular Grid Implementation. Chem. Phys. Lett. 2009, 474, 366–370.

(37). Jeanmairet G; Levesque M; Vuilleumier R; Borgis D Molecular Density Functional Theory of Water. J. Phys. Chem. Lett. 2013, 4, 619–624. [PubMed: 26281876]

(38). Ratkova EL; Palmer DS; Fedorov MV Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. Chem. Rev. 2015, 115, 6312–6356. [PubMed: 26073187]

(39). Sergiievskyi V; Jeanmairet G; Levesque M; Borgis D Solvation Free-Energy Pressure Corrections in the Three Dimensional Reference Interaction Site Model. J. Chem. Phys. 2015, 143, 184116. [PubMed: 26567655]

(40). Remsing RC; Liu S; Weeks JD Long-Ranged Contributions to Solvation Free Energies from Theory and Short-Ranged Models. Proc. Natl. Acad. Sci. USA 2016, 113, 2819–2826. [PubMed: 26929375]

(41). Matubayasi N Free-Energy Analysis of Protein Solvation with All-Atom Molecular Dynamics Simulation Combined with a Theory of Solutions. Curr. Opn. Str. Bio. 2017, 43, 45–54.

(42). Levy RM; Cui D; Zhang BW; Matubayasi N Relationship between Solvation Thermodynamics from IST and DFT Perspectives. J. Phys. Chem. B 2017, 121, 3825–3841. [PubMed: 28186751]

(43). Cui D; Zhang BW; Matubayasi N; Levy RM The Role of Interfacial Water in Protein-Ligand Binding: Insights from the Indirect Solvent Mediated Potential of Mean Force. J. Chem. Theory Comput. 2018, 14, 512–526. [PubMed: 29262255]

(44). Matubayasi N; Nakahara M Theory of Solutions in the Energy Representation. II. Functional for the Chemical Potential. J. Chem. Phys. 2002, 117, 3605–3616.

(45). Matubayasi N; Nakahara M Erratum: “Theory of Solutions in the Energy Representation. II. Functional for the Chemical Potential” [J. Chem. Phys. 117, 3605 (2002)]. J. Chem. Phys. 2003, 118, 2446.

(46). Matubayasi N; Nakahara M Theory of Solutions in the Energy Representation. III. Treatment of the Molecular Flexibility. J. Chem. Phys. 2003, 119, 9686–9702.

(47). Sakuraba S; Matubayasi N ERmod: Fast and Versatile Computation Software for Solvation Free Energy with Approximate Theory of Solutions. J. Comput. Chem. 2014, 35, 1592–1608, Footnote 1 on page 1598 of this paper was incorrectly stated. The TIP3P model used for the calculations was the CHARMM-modified one with a Lennard-Jones term on the hydrogen atom. [PubMed: 24923817]

(48). Matubayasi N; Liang KK; Nakahara M Free-Energy Analysis of Solubilization in Micelle. J. Chem. Phys. 2006, 124, 154908. [PubMed: 16674266]

(49). Matubayasi N; Shinoda W; Nakahara M Free-Energy Analysis of the Molecular Binding into Lipid Membrane with the Method of Energy Representation. J. Chem. Phys. 2008, 128, 195107. [PubMed: 18500905]

(50). Karino Y; Matubayasi N Free-Energy Analysis of Hydration Effect on Protein with Explicit Solvent: Equilibrium Fluctuation of Cytochrome c. J. Chem. Phys. 2011, 134, 041105. [PubMed: 21280680]

(51). Yamamori Y; Ishizuka R; Karino Y; Sakuraba S; Matubayasi N Interaction-Component Analysis of the Hydration and Urea Effects on Cytochrome c. J. Chem. Phys. 2016, 144, 085102. [PubMed: 26931726]

(52). Date A; Ishizuka R; Matubayasi N Energetics of Nonpolar and Polar Compounds in Cationic, Anionic, and Nonionic Micelles Studied by All-Atom Molecular Dynamics Simulation Combined with a Theory of Solutions. Phys. Chem. Chem. Phys. 2016, 18, 13223–13231. [PubMed: 27117093]

(53). Mizuguchi T; Matubayasi N Free-Energy Analysis of Peptide Binding in Lipid Membrane Using All-Atom Molecular Dynamics Simulation Combined with Theory of Solutions. J. Phys. Chem. B 2018, 122, 3219–3229. [PubMed: 29320189]

(54). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983, 79, 926–935.

(55). Jorgensen WL; Maxwell DS; Tirado-Rives J Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc. 1996, 118, 11225–11236.

(56). Kaminski GA; Friesner RA; Tirado-Rives J; Jorgensen WL Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B 2001, 105, 6474–6487.

(57). Pronk S; Páll S; Schulz R; Larsson P; Bjelkmar P; Apostolov R; Shirts MR; Smith JC; Kasson PM; van der Spoel D; Hess B; Lindahl E GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. Bioinformatics 2013, 29, 845–854. [PubMed: 23407358]

(58). Abraham MJ; Murtola T; Schulz R; Páll S; Smith JC; Hess B; Lindahl E GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX 2015, 1, 19–25.

(59). Yoda T; Sugita Y; Okamoto Y Comparisons of Force Fields for Proteins by Generalized-Ensemble Simulations. Chem. Phys. Lett. 2004, 386, 460–467.

(60). Chodera JD; Swope WC; Pitera JW; Dill KA Long-Time Protein Folding Dynamics from Short Time Molecular Dynamics Simulations. Multiscale Model. Simul. 2006, 5, 1214–1226.

(61). Strodel B; Wales DJ Free Energy Surfaces from an Extended Harmonic Superposition Approach and Kinetics for Alanine Dipeptide. Chem. Phys. Lett. 2008, 466, 105–115.

(62). Feig M Is Alanine Dipeptide a Good Model for Representing the Torsional Preferences of Protein Backbones? J. Chem. Theory Comput. 2008, 4, 1555–1564. [PubMed: 26621437]

(63). Okumura H Optimization of Partial Multicanonical Molecular Dynamics Simulations Applied to an Alanine Dipeptide in Explicit Water Solvent. Phys. Chem. Chem. Phys. 2011, 13, 114–126. [PubMed: 21038036]

(64). Jaillet L; Corcho FJ; Pérez JJ; Cortes J Randomized Tree Construction Algorithm to Explore Energy Landscapes. J. Comput. Chem. 2011, 32, 3464–3474. [PubMed: 21919017]

(65). Wang H; Schütte C; Ciccotti G; Delle Site L Exploring the Conformational Dynamics of Alanine Dipeptide in Solution Subjected to an External Electric Field: A Nonequilibrium Molecular Dynamics Simulation. J. Chem. Theory Comput. 2014, 10, 1376–1386. [PubMed: 26580357]

(66). Deng N; Zhang BW; Levy RM Connecting Free Energy Surfaces in Implicit and Explicit Solvent: An Efficient Method to Compute Conformational and Solvation Free Energies. J. Chem. Theory Comput. 2015, 11, 2868–2878. [PubMed: 26236174]

(67). Almeida GG; Cordeiro JMM; Martín ME; Aguilar MA Conformational Changes of the Alanine Dipeptide in Water-Ethanol Binary Mixtures. J. Chem. Theory Comput. 2016, 12, 1514–1524. [PubMed: 26910305]

(68). Poblete H; Miranda-Carvajal I; Comer J Determinants of Alanine Dipeptide Conformational Equilibria on Graphene and Hydroxylated Derivatives. J. Phys. Chem. B 2017, 121, 3895–3907. [PubMed: 28291356]

(69). Harris RC; Deng N; Levy RM; Ishizuka R; Matubayasi N Computing Conformational Free Energy Differences in Explicit Solvent: An Efficient Thermodynamic Cycle Using an Auxiliary Potential and a Free Energy Functional Constructed from the End Points. J. Comput. Chem. 2017, 38, 1198–1208. [PubMed: 28008630]

(70). Morishita T; Yonezawa Y; Ito AM Free Energy Reconstruction from Logarithmic Mean-Force Dynamics Using Multiple Nonequilibrium Trajectories. J. Chem. Theory Comput. 2017, 13, 3106–3119. [PubMed: 28602083]

(71). Frenkel D; Smit B Understanding Molecular Simulation: From Algorithms to Applications; Academic Press: London, UK, 1996.

(72). Bennett CH Efficient Estimation of Free Energy Differences from Monte Carlo Data. J. Comput. Phys. 1976, 22, 245–268.

(73). van Gunsteren WF; Berendsen HJC A Leap-Frog Algorithm for Stochastic Dynamics. Mol. Simul. 1988, 1, 173–185.

(74). Parrinello M; Rahman A Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. J. Appl. Phys. 1981, 52, 7182–7190.

(75). Miyamoto S; Kollman PA SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. J. Comput. Chem. 1992, 13, 952–962.

(76). Hess B; Bekker H; Berendsen HJC; Fraaije JGEM LINCS: A Linear Constraint Solver for Molecular Simulations. J. Comput. Chem. 1997, 18, 1463–1472.

(77). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 1995, 103, 8577–8593.

(78). van der Spoel D; van Maaren PJ The Origin of Layer Structure Artifacts in Simulations of Liquid Water. J. Chem. Theory Comput. 2006, 2, 1–11. [PubMed: 26626372]

(79). Sugita Y; Okamoto Y Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem. Phys. Lett. 1999, 314, 141–151.

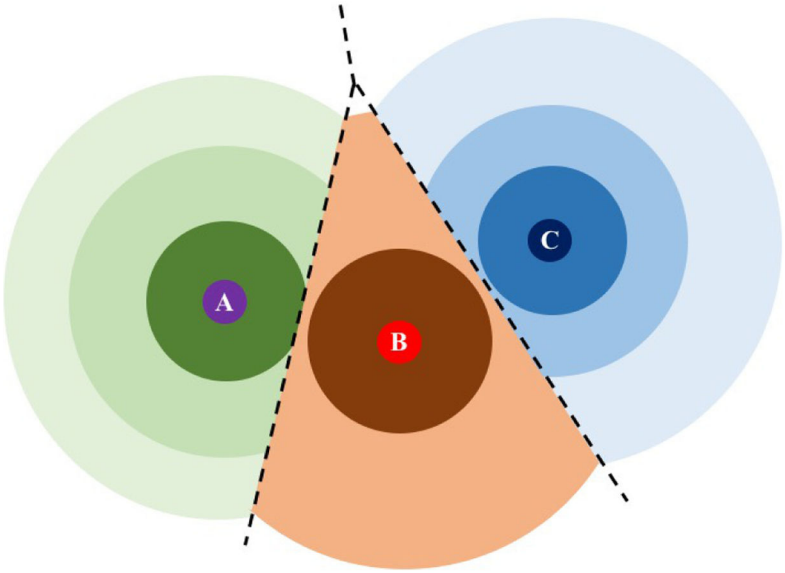


Figure 1:
An example of division of the space. The three regions with the boundaries written as dashed lines are closest to the centers A, B, and C, respectively, from left to right, and are further divided into the regions shown with distinct colors. The white part corresponds to the "far-separated" region in RESULTS AND DISCUSSION section.

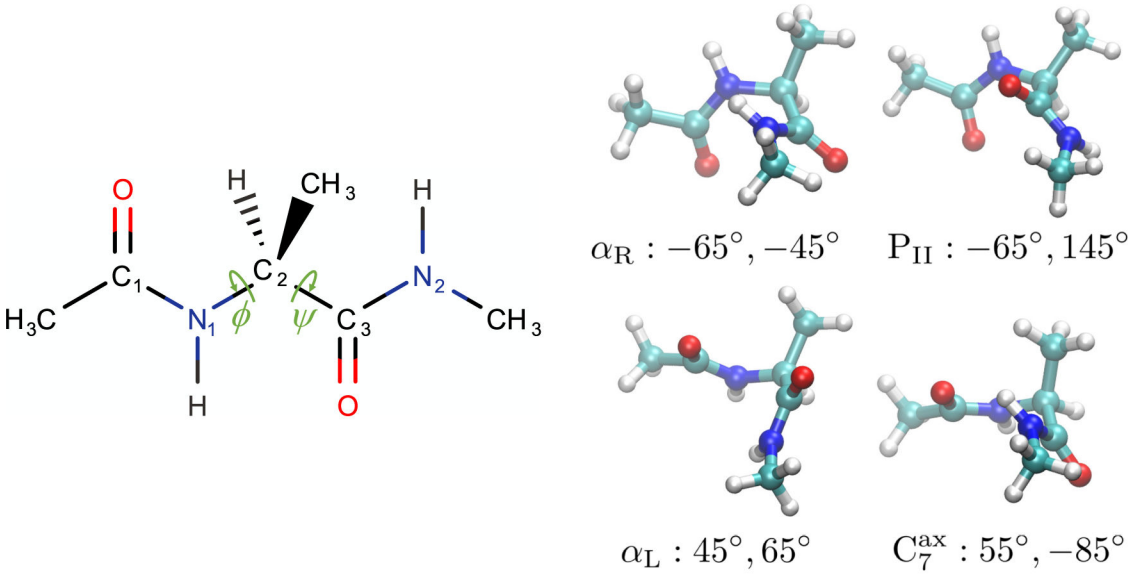


Figure 2:

Structure of alanine dipeptide, with 4 conformational states of α_R, P_{II}, α_L, and C_7^{ax}. The C_1, N_1, C_2, C_3, and N_2 atoms refer to carbon and nitrogen atoms on the main chain (and the termini), and in the subsection of "Alanine Dipeptide", the space is divided in terms of the distances from C_1, N_1, C_2, C_3, and N_2. The dihedral angles ϕ and ψ correspond to C_1—N_1—C_2—C_3 and N_1—C_2—C_3—N_2, respectively, and are 180° at the trans conformations.

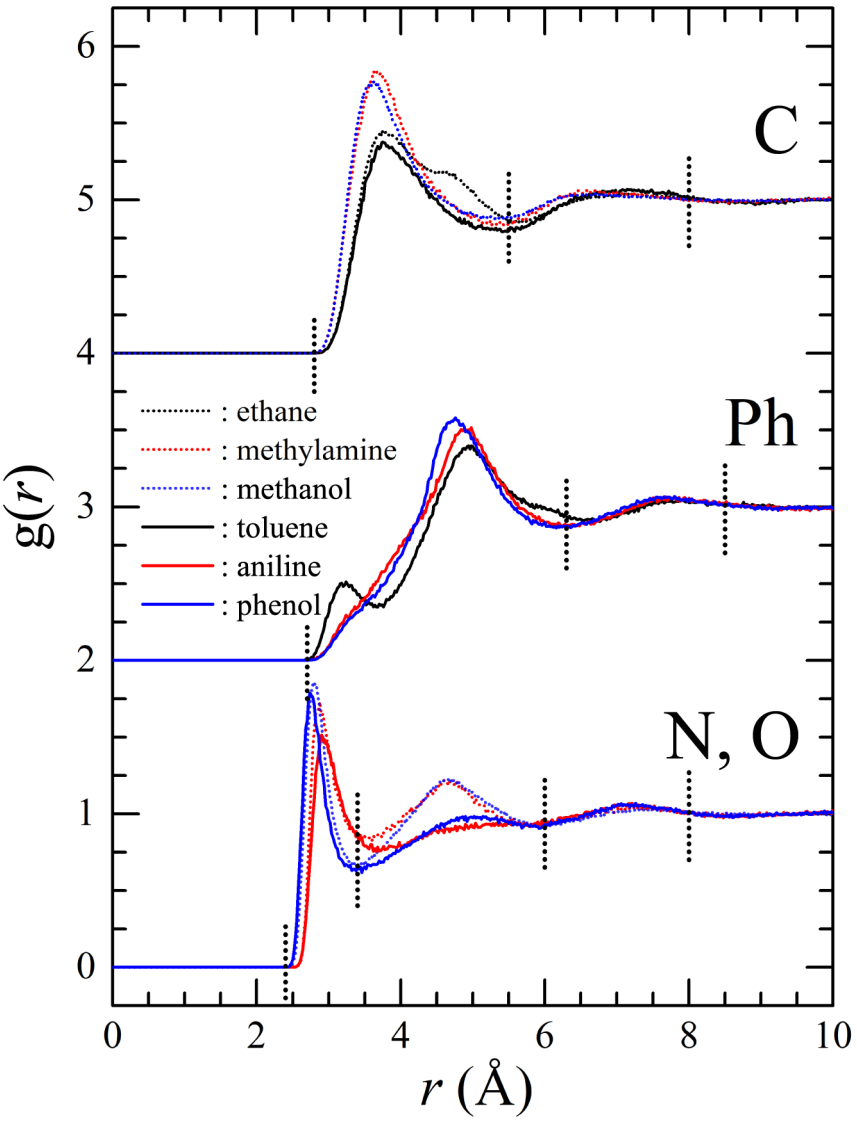


Figure 3:

Radial distribution functions $g(r)$ of water around ethane, methylamine, methanol, toluene, aniline, and phenol, where r is the distance of the oxygen site of a solvent water molecule from the methyl carbon, amine nitrogen, hydroxyl oxygen, or the center of the phenyl ring. C denotes the carbon atom in the methyl group for ethane, methylamine, methanol, and toluene, and $g(r)$ from the two C sites for ethane were averaged on the basis of symmetry. Ph refers to the center-of-mass position of the 6 carbon atoms in the phenyl ring for toluene, aniline, and phenol, and N and O are the amine nitrogen for methylamine and aniline and the hydroxyl oxygen for methanol and phenol, respectively. $g(r)$ in the plot is shifted upward by 4 and 2 for C and Ph, respectively. The dotted lines drawn vertically in the figure correspond to the boundaries of the regions listed in Table 1.

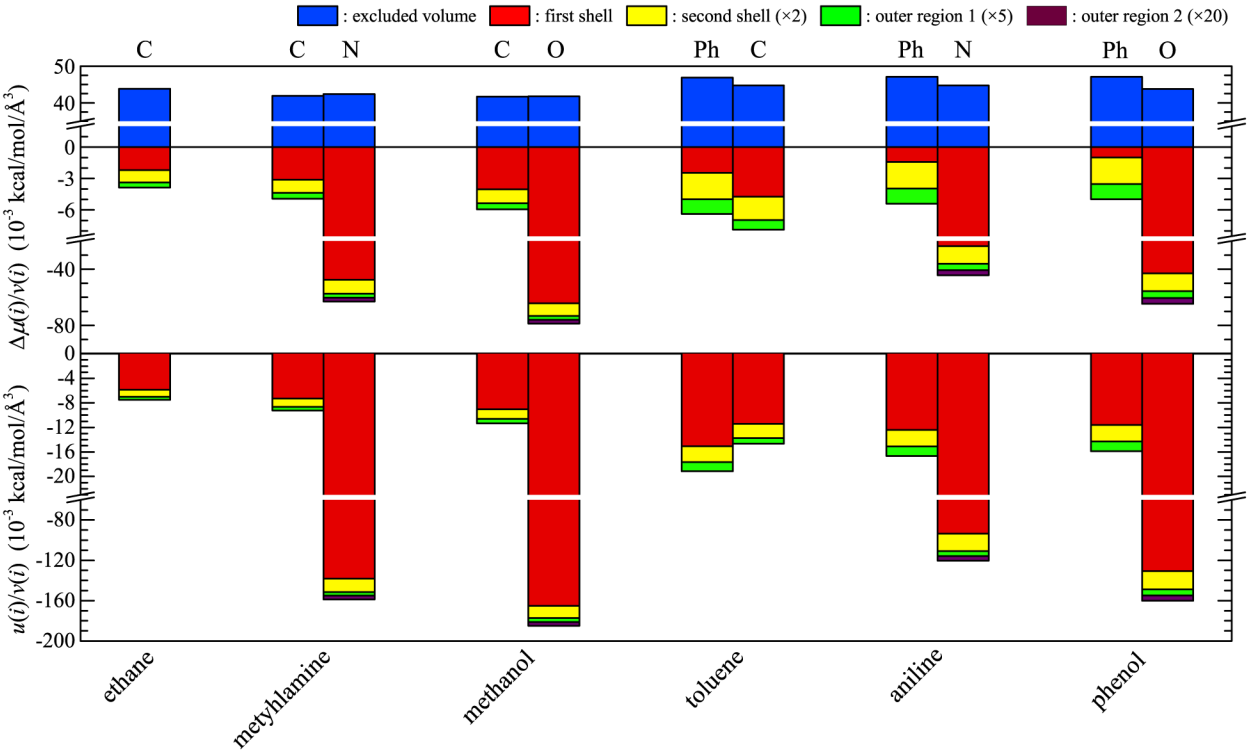


Figure 4:

Densities of the solvation free energy $A\mu(i)/v(i)$ and of the solute-solvent interaction energy $u(i)/v(i)$ in the regions around the solute introduced in Table 1. The LJ long-range correction is added to $A\mu(i)$ and $u(i)$ in the far-separated region in Table 1, and $A\mu(i)/v(i)$ and $u(i)/v(i)$ in the far-separated region are not shown since that region is unbounded and the densities were not determined. The bars representing the values of $A\mu(i)/v(i)$ and $u(i)/v(i)$ are stacked for each solute, and the green bars refer to the outer regions for the methyl carbon (C) and the phenyl center (Ph) and the outer region 1 for the amine nitrogen (N) and the hydroxyl oxygen (O). The values are multiplied by a factor of 2 in the second shell, by a factor of 5 in the outer regions of C and Ph, and by factors of 5 and 20 in the outer regions 1 and 2 of N and O, respectively. $u(i)/v(i)$ in the excluded volume is not plotted because it is simply zero. The numerical values of $A\mu(i)/v(i)$ and $u(i)/v(i)$ are listed in Supporting Information.

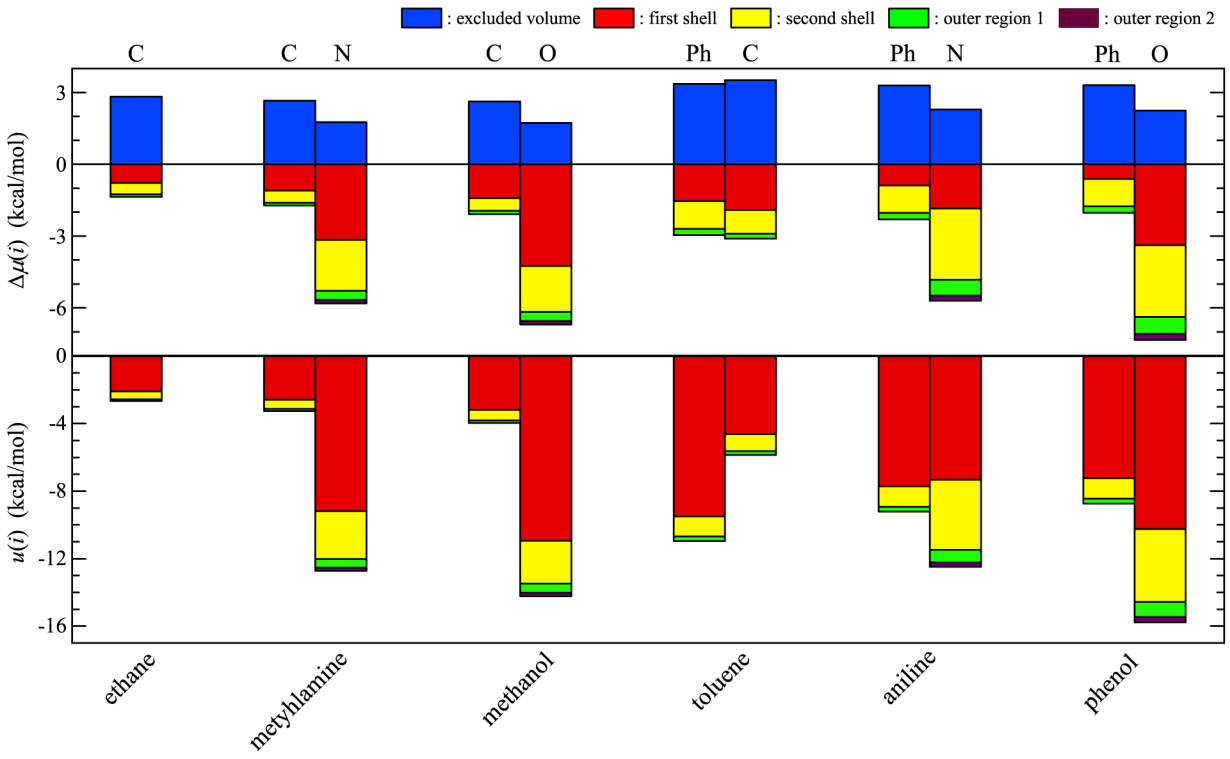


Figure 5:

Solvation free energy $\Delta\mu(i)$ and the solute-solvent interaction energy $u(i)$ in the regions around the solute introduced in Table 1. The bars representing the values of $\Delta\mu(i)$ and $u(i)$ are stacked for each solute, and $u(i)$ in the excluded volume is not plotted because it is simply zero. The LJ long-range correction is added to $\Delta\mu(i)$ and $u(i)$ in the far-separated region in Table 1, and is not shown in this figure. The numerical values of $\Delta\mu(i)$ and $u(i)$ are listed in Supporting Information.

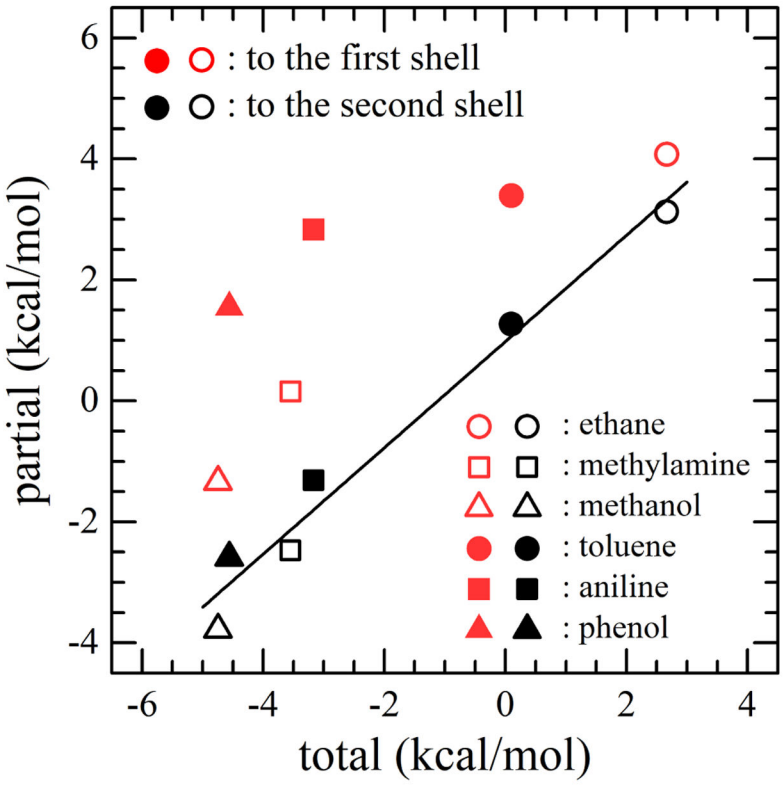


Figure 6:

Correlations of the sum of $\Delta\mu(i)$ over all the regions around the solute molecule against the sums over the excluded-volume and first-shell regions of all the sites and over the excluded-volume to second-shell regions. A single point in the figure corresponds to a single solute species, and with the linear regression of the partial sum to the second-shell region against the total sum, the slope is 0.88 with a correlation coefficient of 0.99.

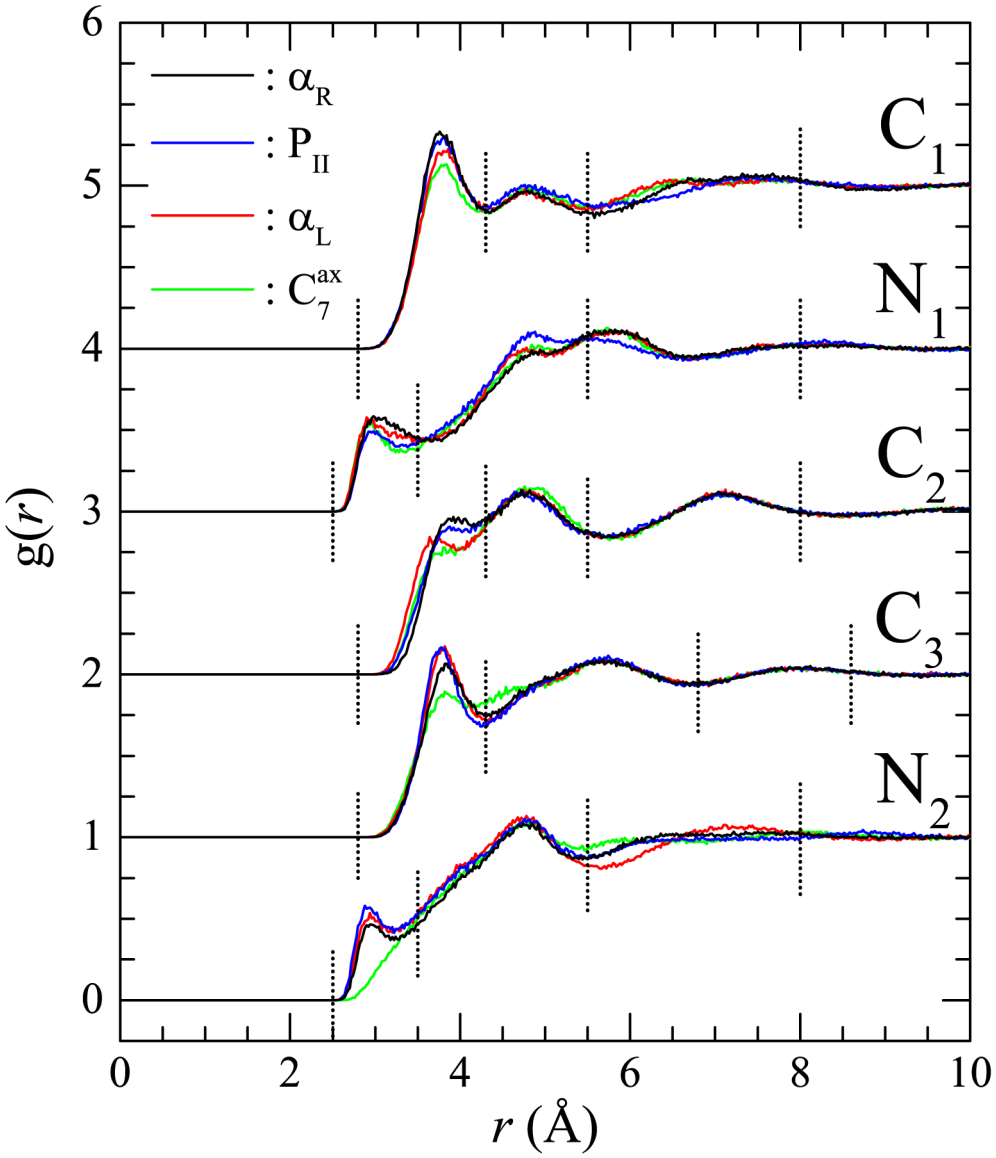


Figure 7:

Radial distribution functions $g(r)$ of water around the C_1, N_1, C_2, C_3, and N_2 atoms labelled in Fig. 2, where r is the distance from the oxygen site of a solvent water molecule. In the plot, $g(r)$ is shifted upward by 4, 3, 2, and 1 for C_1, N_1, C_2, and C_3, respectively, and the dotted lines drawn vertically in the figure correspond to the boundaries of the regions listed in Table 2.

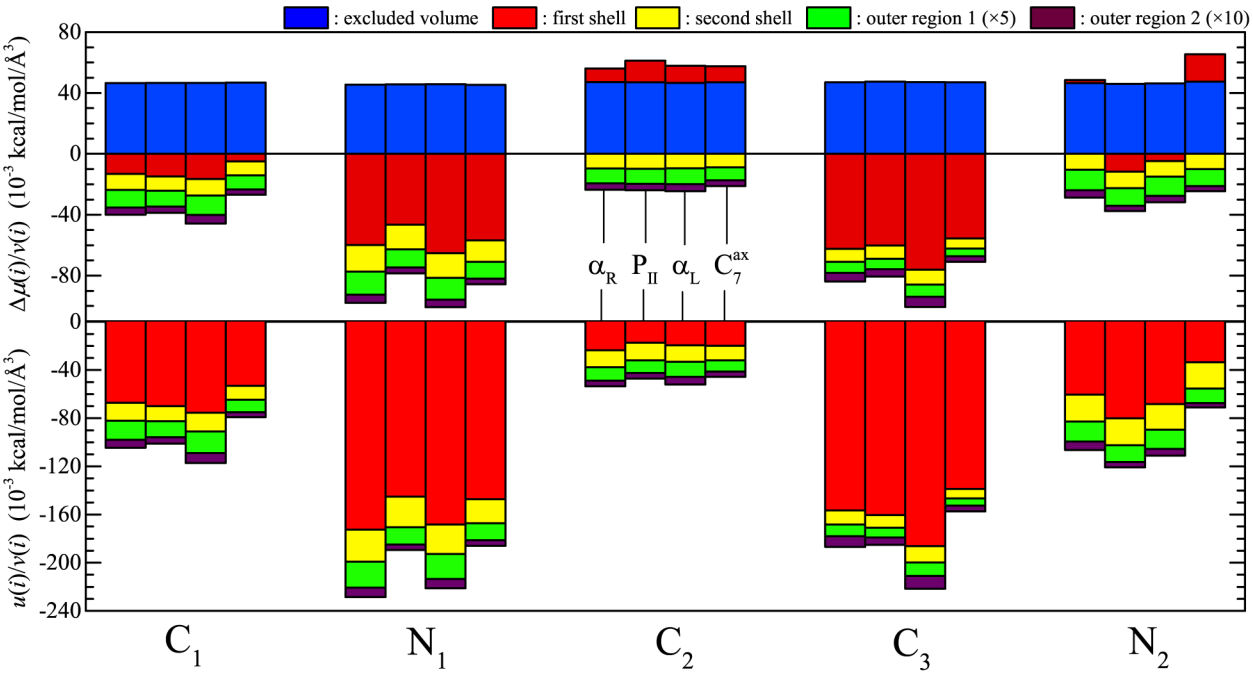


Figure 8:

Densities of the solvation free energy $A\mu(i)/v(i)$ and of the solute-solvent interaction energy $u(i)/v(i)$ in the regions around alanine dipeptide introduced in Table 2. The LJ long-range correction is added to $A\mu(i)$ and $u(i)$ in the far-separated region in Table 2, and $A\mu(i)/v(i)$ and $u(i)/v(i)$ in the far-separated region are not shown since that region is unbounded and the densities were not determined. The bars representing the values of $A\mu(i)/v(i)$ and $u(i)/v(i)$ are stacked for each of C_1, N_1, C_2, C_3, and N_2, and are shown for the conformations of α_R, P_{II}, α_L, and C_7^{ax} from left to right. The values in the outer regions 1 and 2 are multiplied by factors of 5 and 10, respectively, and $u(i)/v(i)$ in the excluded volume is not plotted because it is simply zero. The molecular structure of alanine dipeptide is provided with the ϕ and ψ values in Fig. 2, and the numerical values of $A\mu(i)/v(i)$ and $u(i)/v(i)$ are listed in Supporting Information.

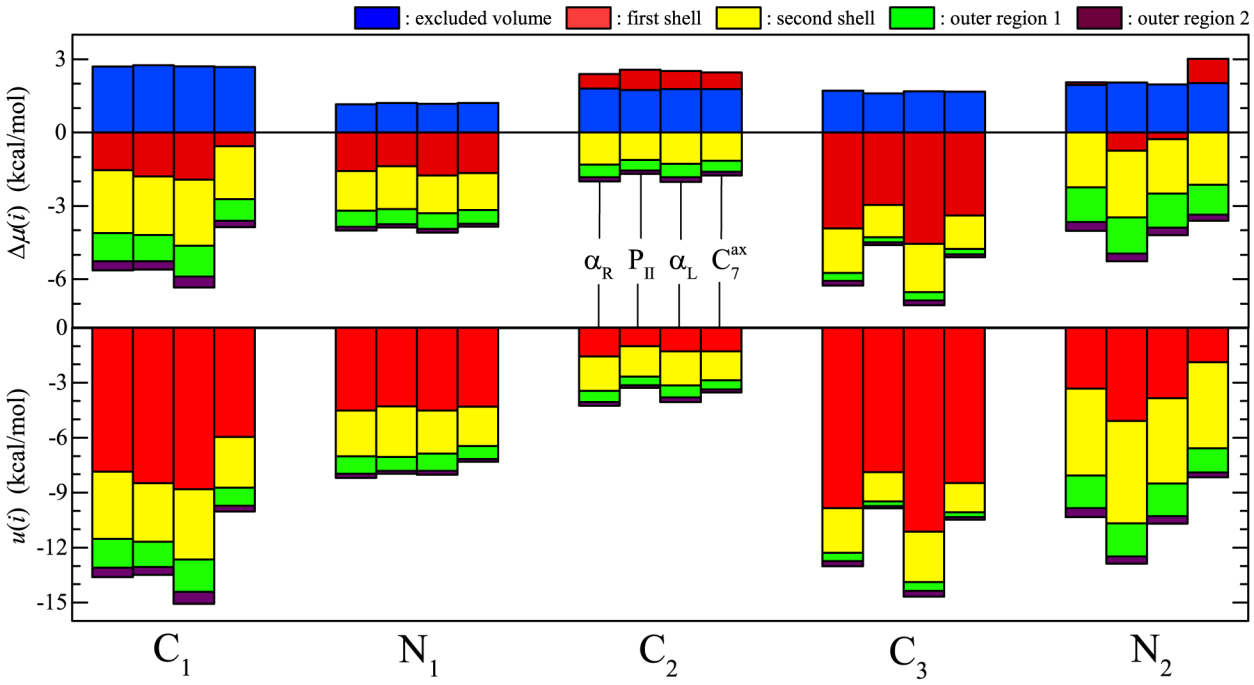


Figure 9:

Solvation free energy $\Delta\mu(i)$ and the solute-solvent interaction energy $u(i)$ in the regions around alanine dipeptide introduced in Table 2. The bars representing the values of $\Delta\mu(i)$ and $u(i)$ are stacked for each of C_1, N_1, C_2, C_3, and N_2, and are shown for the conformations of α_R, P_{II}, α_L, and C_7^{ax} from left to right. $u(i)$ in the excluded volume is not plotted because it is simply zero. The LJ long-range correction is added to $\Delta\mu(i)$ and $u(i)$ in the far-separated region in Table 2, and is not shown in this figure. The molecular structure of alanine dipeptide is provided with the ϕ and ψ values in Fig. 2, and the numerical values of $\Delta\mu(i)$ and $u(i)$ are listed in Supporting Information.

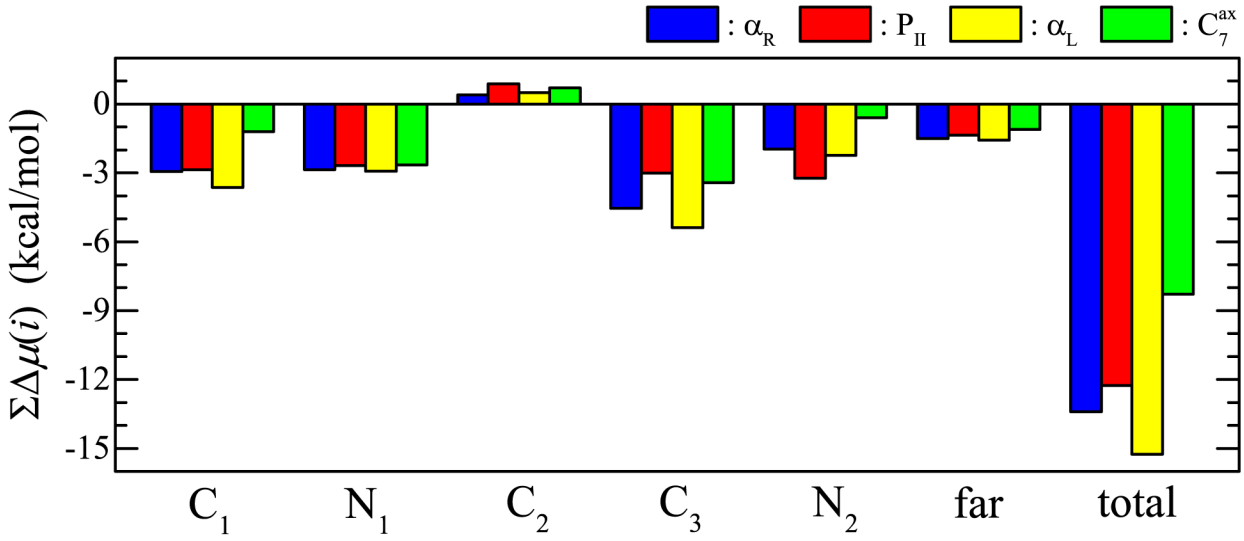


Figure 10:

Sum of $\Delta\mu(i)$ over the excluded-volume to outer regions for each of the C_1, N_1, C_2, C_3, and N_2 atoms at the α_R, P_{II}, α_L, and C_7^{ax} conformations of alanine dipeptide, with the contribution from the far-separated region in Table 2 and the total sum value of $\Delta\mu(i)$. The LJ long-range correction is added to $\Delta\mu(i)$ in the far-separated region.

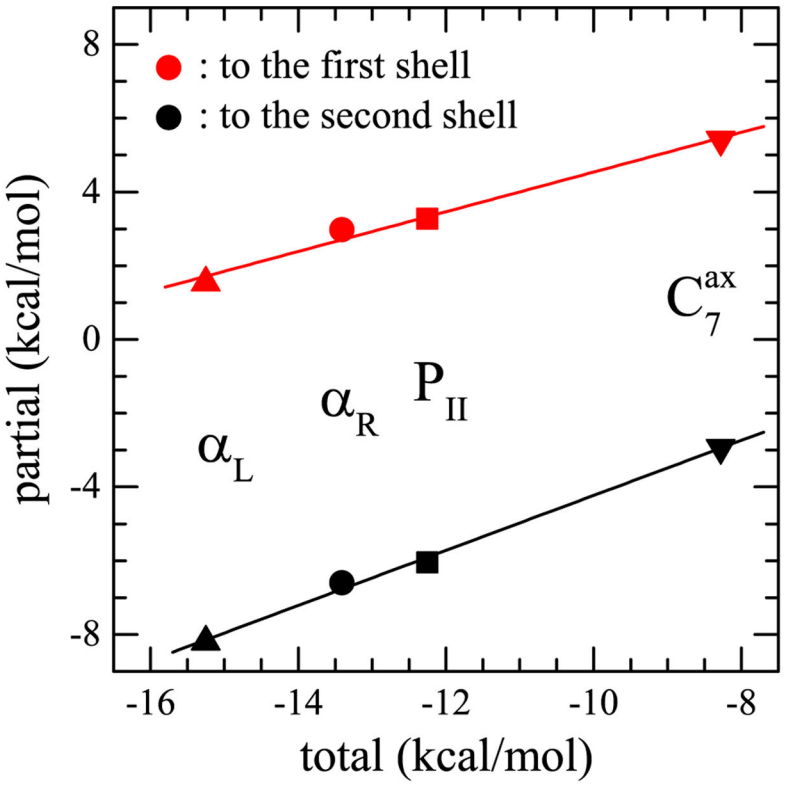


Figure 11:

Correlations of the sum of $\Delta\mu(i)$ over all the regions around alanine dipeptide against the partial sums over the excluded-volume and first-shell regions of all the atomic sites and over the excluded-volume to second-shell regions. A single point in the figure corresponds to a single conformation of alanine dipeptide. With the linear regression against the total sum, the slope is 0.54 and 0.75 for the sums to the first and second shells, respectively, with correlation coefficients of 0.99 and 1.00

Table 1:

Division of the space into regions for the small solutes

atom in the solute	region	distance (Å)[a]
C (ethane, methylamine, methanol, and toluene)[b]	excluded volume	$R < 2.8$
	first shell	$2.8 < R < 5.5$
	second shell	$5.5 < R < 8.0$
	outer region	$8.0 < R < 10.0$
Ph (toluene, aniline, and phenol)[c]	excluded volume	$R < 2.7$
	first shell second shell	$2.7 < R < 6.3$
	second shell	$6.3 < R < 8.5$
	outer region	$8.5 < R < 10.0$
N and O (methylamine, methanol, aniline, and phenol)[d]	excluded volume	$R < 2.4$
	first shell	$2.4 < R < 3.4$
	second shell	$3.4 < R < 6.0$
	outer region 1	$6.0 < R < 8.0$
	outer region 2	$8.0 < R < 10.0$
	far-separated[e]	$R > 10.0$

[a] R is the minimum of the distances of the C, Ph, N and O sites in the solute with the oxygen site of the solvent water molecule.

[b] C is the carbon atom in the methyl group of ethane, methylamine, methanol, and toluene.

[c] Ph is the center of mass of the 6 carbon atoms in the phenyl ring of toluene, aniline, and phenol.

[d] N is the amine nitrogen of methylamine and aniline, and O is the hydroxyl oxygen of methanol and phenol.

[e] The solvent water is separated by more than 10 Å from all of the C, Ph, N, and O sites in the solute.

Table 2:

Division of the space into regions around alanine dipeptide

atom[a]	region	distance (Å)[b]
	excluded volume	$R < 2.8$
	first shell	$2.8 < R < 4.3$
C_1 and C_2	second shell	$4.3 < R < 6.0$
	outer region 1	$6.0 < R < 8.0$
	outer region 2	$8.0 < R < 10.0$
	excluded volume	$R < 2.8$
	first shell	$2.8 < R < 4.3$
C_3	second shell	$4.3 < R < 6.8$
	outer region 1	$6.8 < R < 8.6$
	outer region 2	$8.6 < R < 10.0$
	excluded volume	$R < 2.5$
	first shell	$2.5 < R < 3.5$
N_1 and N_2	second shell	$3.5 < R < 5.5$
	outer region 1	$5.5 < R < 8.0$
	outer region 2	$8.0 < R < 10.0$
	far-separated[c]	$R > 10.0$

[a] The C_1, N_1, C_2, C_3, and N_2 atoms are labelled in Fig. 2.

[b] R is the minimum of the distances of the C_1, N_1, C_2, C_3, and N_2 atoms in Fig. 2 with the oxygen site of the solvent water molecule.

[c] The solvent water is separated by more than 10 Å from all of the C_1, N_1, C_2, C_3, and N_2 atoms in alanine dipeptide.