

End-point calculation of solvation free energy of amino-acid analogs by molecular theories of solution

Yasuhito Karino^a, Maxim V. Fedorov^b, Nobuyuki Matubayasi^{a,b,c,*}

^a*Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan*

^b*Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, Leipzig, 04103, Germany*

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

Abstract

The computational efficiency is improved for the solvation free energy when the calculation is restricted to the initial and final states of the solute insertion process (pure solvent and solution systems of interest). We explore the possibility of such “end-point” calculations and assess the performance of several approximate free-energy functionals against benchmarks for amino-acid analogs in water. The performance is the best and the second, respectively, for the method of energy representation and the RISM (reference interaction site model)/partial-wave expansion supplemented by semi-empirical corrections for the excluded-volume and hydrogen-bonding effects, while the chemical accuracy is not achieved for the others.

*Corresponding author

Email address: nobuyuki@scl.kyoto-u.ac.jp (Nobuyuki Matubayasi)

1. Introduction

A realistic prediction of the solvation free energy in molecular systems is one of the central tasks for modern computational physical chemistry. During the last decade, there has been much progress in the area, especially for small organic molecules in water [1, 2, 3, 4, 5, 6, 7, 8]. Indeed, the necessity of a reliable estimation of the solvation free energy is increasing since the role of solvent is now well recognized for biophysical, interfacial, and electrochemical processes in solution.

The method to calculate the solvation free energy is categorized as either implicit- or explicit-solvent model. In the implicit model, the nonpolar part of the solvation free energy is typically parametrized by cavity-formation term and the electrostatic part is described by continuum electrostatics, usually with the Poisson-Boltzmann or Generalized-Born equation [9, 10, 11, 12, 13, 14]. The implicit-solvent method does not require extensive computations and is amenable to large solutes. It is often difficult, however, to extend the method beyond the range of model parametrization: for example, to systems with added cosolvents or salts. Furthermore, it is common to ignore the molecular structure of the solvent as well as the specific effect of solute-solvent interaction such as the hydrogen bonding.

In the explicit-solvent model, the molecular details of the solvent are taken into account and molecular-simulation methods are most often employed to sample the system configuration at atomic resolution. There have been developed detailed molecular models for both the solute and solvent which provide realistic pictures of solvation [2, 3]. With the use of well-tuned atomistic force field, the solvation free energy of small to medium organic so-

lute in water can be calculated with a deviation of ~ 1 kcal/mol from the experimental value [2], though an accuracy better than ~ 3 kcal/mol is still a challenge for an arbitrary chosen set of compounds as it has been shown in a recent blind test [15]. The drawback in the free-energy computation using the explicit-solvent model is the computational demand. The standard approach is the free-energy perturbation method and the thermodynamics integration method, and these methods can be conducted in practice only by introducing a number of intermediate states of the solute insertion process connecting the pure solvent and solution systems of interest.

To circumvent the heavy computational load involved in the standard approaches to the solvation free energy, alternative schemes are formulated which refer only to the initial and final states of the process of solute insertion (the pure solvent and solution systems) and use approximate functionals for the solvation free energy. These “end-point” free-energy calculations are computationally more efficient, and are shown to be promising even for large and/or complex solute species when an appropriate assessment is done for the accuracy [5, 6, 7, 8, 11]. The purpose of the present work is to assess the performance of “end-point” methods introduced below.

When the solute and solvent are both treated with an explicit model, the end-point method is classified into two categories. The first category consists of integral equation theories for molecular liquids [7, 8, 16, 17, 18, 19, 20, 21, 22]. The method of reference interaction site model (RISM) and its 3-dimensional extension (3D-RISM) are commonly known integral-equation approaches, and operate with the solute-solvent and solvent-solvent radial (or spatial, 3-dimensional) distribution functions. This approach was introduced

by Chandler and Andersen [16] and further developments have been made towards improving the accuracy [7, 17, 18, 22, 23] and the computational efficiency [20, 24, 25]. In this molecular theory, there is a family of functionals with which the solvation free energy is expressed in closed form in terms of the distribution functions at the end points. The best documented example is the hypernetted-chain (HNC) functional [26], and there are a few others beyond HNC. The earliest non-HNC model was developed by assuming the Gaussian fluctuations (GF) of solvent [27], and an improvement was later proposed on the basis of the repulsive bridge (RB) correction [28] and the distributed partial wave (PW) expansion [23]. The accuracy of the PW approach was further improved with semi-empirical corrections for the excluded-volume and hydrogen-bonding contributions [7].

The second category of the end-point calculation of free energy refers to combined approaches of molecular simulation and a theory of solutions. The simulations are carried out only at the end points (the pure solvent and solution systems), and the distribution functions obtained are substituted into a free-energy functional formulated in the theory of solutions [5, 6, 29, 30]. Within this category, the method of energy representation (ER) is a unique candidate with its accuracy and wide applicability [4, 5, 6, 31, 32, 33]. In this method, the solute-solvent pair interaction energy is taken as the functional variable, and the solvation free energy is expressed as a functional of a set of distribution functions of the pair interaction energy. The method was so far developed in combination with molecular simulation, and the currently used version of functional is exact to low orders in (solvent) density and is applicable to inhomogeneous system and to flexible molecules involving

intramolecular degrees of freedom [5, 6, 32, 33].

In this paper, we assess the performance of the end-point methods described above. We focus on the method of energy representation and the RISM-based schemes, and compare the calculated results against benchmarks for amino-acid side-chain analogs in water solvent [2, 34]. The set of solute molecules examined are taken from Ref. [2] and are listed in Table 1.

2. Computational Procedures

The potential function employed for the molecular dynamics (MD) simulations was the standard TIP3P for water [35] and the OPLS-AA [36, 37] for the solute molecules with modifications adopted in the benchmark work [2, 38]. As shown in Refs. [2, 38] from an extensive set of free-energy calculations, this set of potential functions provides the best agreement with the experimental values among the commonly used force fields. The Lennard-Jones parameters between different kinds of atomic sites were combined by the geometric mean in the MD simulations.

The free-energy calculation with the method of energy representation (ER) was conducted in combination with MD simulations. In the ER method, an MD snapshot is sampled to construct the instantaneous histogram $\widehat{\rho}^e$ of the solute-solvent pair interaction energy ϵ defined as

$$\widehat{\rho}^e(\epsilon) = \sum_i \delta(v(\psi, \mathbf{x}_i) - \epsilon), \quad (1)$$

where ψ is the solute coordinate, \mathbf{x}_i is the coordinate of the i th solvent molecule, v is the potential function for the solute-solvent pair interaction, and the sum is taken over all the solvent molecules. A superscript e is

attached to emphasize that a function is represented over the energy coordinate ϵ . For each solute, three MD simulations were performed. One is the solution system containing a single solute molecule of interest and 900 water molecules. The other two are the pure solvent system consisting only of 900 water molecules and the single-molecule system of the solute; the single-molecule MD was employed as a sampling scheme of the intramolecular configuration of the solute at isolation (in the absence of solvent). For the solution and pure solvent systems, the periodic boundary condition was adopted and the unit cell was taken to be cubic. The configurations were generated in the isothermal-isobaric ensemble at 298 K and 1 bar using the Nosé-Hoover thermostat and the Parrinello-Rahman barostat at time constants of 1.0 ps [39, 40]. The electrostatic interaction was handled by the particle-mesh Ewald (PME) method [41] with a real-space cutoff of 7.5 Å, a spline order of 6, a relative tolerance of 10^{-5} , and a reciprocal-space mesh size of 32 for each of the x , y , and z directions. The Lennard-Jones interaction was truncated by applying the switching function in the form of Ref. [42], where the switching range is 7.0-7.5 Å. Although this cutoff is rather short compared to the common choice, it does not affect the solvation free energies of the present set of solutes when the long-range correction is appropriately made for the Lennard-Jones term [2, 38]. The single-molecule MD of the solute was carried out in the constant-temperature ensemble at 298 K with the Nosé-Hoover thermostat at a time constant of 1.0 ps. The equation of motion was integrated with the velocity Verlet algorithm at a time step of 2 fs [43]. The bond lengths in the solutes were fixed with RATTLE [44], and the water molecule was kept rigid with SETTLE [45]. The above setups

for MD and the choice of potential functions are the same as those in the benchmark work [2]. All the MD calculations were done using GROMACS version 4.0.4 [46].

The MD of the solution system was performed for 100 ps. The solution system was sampled every 10 fs, and $\widehat{\rho}^e$ of Eq. 1 was constructed at each configuration sampled. $\widehat{\rho}^e(\epsilon)$ was then averaged over 10^4 configurations to obtain the average distribution $\rho^e(\epsilon)$. The length of the simulation was 10 ps for the pure solvent system and 10 ns for the single-molecule MD of the solute. The sampling frequency was 100 fs for both the pure solvent MD and the single-solute MD. In the ER scheme, the solute is inserted into the pure solvent system as a test particle, and $\widehat{\rho}^e$ in this case refers to the histogram of the pair interaction between the solute inserted and the solvent. The solute is flexible in the present work, and the single-solute MD serves to sample the intramolecular flexibility of the solute. The test-particle insertion of the solute was carried out at random position and orientation in the pure solvent with the intramolecular configuration sampled from the single-molecule MD. The insertion was performed 1000 times for each pure solvent configuration sampled, and $\widehat{\rho}^e(\epsilon)$ was calculated 10^5 times in total to provide the average distribution $\rho_0^e(\epsilon)$ and the correlation matrix $\chi_0^e(\epsilon, \eta)$. The three energy distribution functions ρ^e , ρ_0^e , and χ_0^e are then the inputs to the functional for the solvation free energy. The form of the functional is given in previous papers [5, 33]. The procedure to fix the potential zero was adopted from Appendix B of Ref. [6], and the long-range correction for the Lennard-Jones term was added through the scheme in Ref. [2] to obtain the ER value of the solvation free energy in Table 1.

Since the solute insertion into the pure solvent as a test particle does not affect the solvent trajectory, the MD of the pure solvent system needs to be performed only once throughout the work and a single trajectory of pure solvent can be used for the insertions of all the solutes. In addition, the insertion at random position and orientation averages out the structural inhomogeneity of the pure solvent system, which might be transiently present in a rather short MD of 10 ps. As pointed out in Ref. [33], the spatial inhomogeneity is exploited to reduce the computation time, which should otherwise be set larger than the relaxation time of the local structure (in space). In this sense, the averaging over time is replaced by the averaging over space.

In the RISM calculations, the original TIP3P water model [35] cannot be used due to the absence of the Lennard-Jones interaction for water hydrogens. The absence of the Lennard-Jones term at hydrogen leads to divergence of RISM calculations [18], while it does not cause any problem in MD and ER because the repulsive core of oxygen prohibits the unphysical configuration of the hydrogen site. To avoid the numerical difficulty in the RISM method, we used the Lue-Blankschtein model of water [18], as done in Ref. [7]. This water model is well adapted to the integral-equation methods and contains non-zero repulsive Lennard-Jones term at the hydrogen site. Similarly, the same non-zero repulsive Lennard-Jones terms were assigned in the RISM calculations to those solute hydrogens which have zero Lennard-Jones parameters in the original OPLS-AA. Although the Lennard-Jones parameters between different kinds of atomic sites were combined with the geometric mean in the MD and ER calculations, the Lorentz-Berthelot rule was em-

ployed in the RISM solver in order to assure the numerical convergence for all the solutes. The bulk solvent-solvent correlation functions for the Lue-Blankschtein model were taken from Ref. [24].

In the RISM calculations in the present work, furthermore, the molecular structures of the solute and solvent are rigid and fixed. The presence of intramolecular flexibility is not allowed, unlike the MD and ER schemes described above. Before performing the RISM calculations, we optimized the solute geometry at isolation (in the absence of solvent) using quantum-chemical calculation with the Gaussian 03 software [47] at the MP2 level and the 6-311G(d,p) basis set. The optimized structure thus obtained was used as the input to the integral equation; only the solute structure was taken from the quantum-chemical calculation and the OPLS-AA force field was adopted for the partial charges and the Lennard-Jones parameters of the solute.

We examined five RISM-based approaches. The first one is the original RISM integral equation with the HNC closure [17]. Three of the non-HNC approaches are the Gaussian fluctuation (GF) model [27], the repulsive bridge (RB) correction [28], and the partial wave (PW) expansion [23]. The last one, called the PW-corrected (PWC) model, expresses the solvation free energy as a sum of the PW value and the informatics-based correction term [7]. The correction takes into account the effects of the excluded volume and the hydrogen bonding, and has two fitting parameters correspondingly. In the present work, a PWC model from Ref. [7] adapted to the OPLS-AA force field was employed as described below. It should be noted that even for the non-HNC schemes, the set of site-site radial distribution functions obtained as the solution to the original RISM/HNC integral equation was used as

the input to the functional for the solvation free energy. The RISM/HNC integral equation was solved using a fast solver [25] at the solvent density of 0.997 g/cm³ and the temperature of 298 K.

The PWC free energy $\Delta\mu_{\text{PWC}}$ is expressed as

$$\Delta\mu_{\text{PWC}} = \Delta\mu_{\text{PW}} + a\rho V_u + bN_{\text{OH}}, \quad (2)$$

where $\Delta\mu_{\text{PW}}$ is the solvation free energy calculated by the RISM/HNC integral equation and the PW free-energy expression, ρ is the number density of solvent (water), V_u is the excluded volume of the solute given by Eq. (8) of Ref. [7], and N_{OH} is the number of OH groups in the solute molecule. a and b are the correction coefficients in the PWC model, and are determined by the least-square fit of Eq. 2 to the experimental values of the solvation free energies of a “training” set of solute molecules. In the present work, the training set consists of the compounds in Table 1 of Ref. [7] except for those contained also in Table 1 of this paper; the amino-acid analog solutes in Table 1 of this paper constitute the “test” set and there is no overlap between the training and test sets. From the least-square fit, a and b of Eq. 2 were determined to be -2.7 and 1.6 kcal/mol, respectively. These values of a and b are different from the ones in the original work [7] since the OPLS-UA force field was used in Ref. [7]. It is necessary in the present work to adopt the model coefficients for the OPLS-AA force field. The HNC, GF, RB, PW, and PWC values of the solvation free energy are listed in Table 1.

3. Results and Discussion

The calculated values of the solvation free energy $\Delta\mu$ are shown in Table 1, together with the experimental and the exactly calculated ones denoted

as Bennett. The values in exact calculation correspond to the potential functions employed for the MD in the present work (TIP3P for water and OPLS-AA for the solutes) and were obtained with the Bennett acceptance ratio method [2, 48]. The comparison is also made graphically in Figure 1 among the experimental, Bennett, and ER values of $\Delta\mu$. Among the $\Delta\mu$'s from the approximate functionals, the ER and PWC are in good agreement with the experimental and Bennett. To assess the performance of approximate scheme, we introduce the mean absolute deviation (MAD) by

$$\frac{1}{n} \sum_i |\Delta\mu_{i,1} - \Delta\mu_{i,2}|, \quad (3)$$

where $\Delta\mu_{i,1}$ and $\Delta\mu_{i,2}$ are the solvation free energies of the i th solute in the first and second sets of the data, respectively, and n is the number of solutes examined. The MAD between the experimental and Bennett values of $\Delta\mu$ is an indicator for the performance of the force field. It is 0.7 kcal/mol with TIP3P for water and OPLS-AA for solute, and is better than with the other commonly used sets of potential functions [2, 38]; $\Delta\mu$ calculated with the prevalent force fields deviates typically by 1-2 kcal/mol from the experimental.

Table 2 lists the MAD of the $\Delta\mu$ values against the experimental and Bennett. An exactly same set of potential functions were used in the Bennett (computationally exact) calculations and the ER (approximate) calculations. In this case, the MAD between the Bennett and approximate $\Delta\mu$ represents the performance of the approximate functional for the solvation free energy. It is a measure of the approximation level in the computational scheme. The accuracies of the force field and the approximate functional are mixed together in the MAD between the experimental and approximate values. This

MAD is an indicator of the usefulness of the approximate scheme in treating realistic systems. It should be noted that the potential functions employed are different between the Bennett and the RISM-based calculations and that the MAD between them is less informative.

The MAD between the Bennett and ER $\Delta\mu$ is found to be 0.5 kcal/mol and is smaller than between the experimental and Bennett. Since the MAD from the use of an approximate functional is expected to depend weakly on the choice of force field and the present choice of force field is the best among the common ones [2, 38], the error from the ER approximation will be smaller than the error from the use of force field. It is further seen in Table 2 that the MAD between the experimental and ER values is 0.7 kcal/mol. This is not different from the MAD between the experimental and Bennett. Thus, the ER method will be as useful as the computationally exact free-energy calculation in terms of the accuracy with respect to experiment.

Tables 1 and 2 show that among the RISM-based methods, the HNC scheme provides too repulsive (too positive) values of $\Delta\mu$, while the RB results are mostly too attractive (too negative). The performance of the free-energy functional is greatly improved with the introduction of the GF or PW approximation. The chemical accuracy (~ 1 kcal/mol) is then achieved with the informatics-based correction (the PWC scheme). Our calculations indicate that the integral-equation method will be quantitative when combined with the informatics which captures structural characteristics of the solute.

As described in Sec. 2, the solvation free energies $\Delta\mu$ in the present work were calculated in the ER method from a 100-ps MD of the solution system

of interest, a 10-ps MD of the pure solvent, and a 10-ns MD of the single solute molecule at isolation (in the absence of solvent). Actually, when the MD of the pure solvent was extended to 50 ps, the deviation of $\Delta\mu$ from the 10-ps result was found to be less than 0.1 kcal/mol. It was observed, on the other hand, that a sampling over ns timescale is necessary for dihedral degrees of freedom of the solute [38]. We thus examined the convergence behavior of $\Delta\mu$ by focusing on the Ile, Met, and Gln analog-solutes which involve dihedral degrees of freedom.

At isolation, the distributions of the dihedral angles of these solutes were seen to be converged in 10 ns. The convergence needs to be examined for the solution system, and we extended the MD to 10 ns for each of the Ile, Met, and Gln analogs in water. Three values of $\Delta\mu$ were then calculated for each solute from the samplings over 100 ps, 1 ns, and 10 ns. It was found for all the solutes that the three $\Delta\mu$ values agree among themselves within 0.1 kcal/mol. In addition, we examined the average sum $\langle u \rangle$ of the solute-solvent interaction energy ($\langle u \rangle = -7.7 \pm 0.1$, -12.9 ± 0.1 , and -29.3 ± 0.3 kcal/mol for the Ile, Met, and Gln analogs, respectively). This was done because the computational error of $\langle u \rangle$ is only statistical, while the error of $\Delta\mu$ is not only statistical but also reflects the use of an approximate functional. It was observed that the three $\langle u \rangle$ values from the 100 ps, 1 ns, and 10 ns samplings are not different within the statistical error. Actually, it was pointed out in Appendix C of Ref. [33] that $\Delta\mu$ is expected to converge faster than $\langle u \rangle$ due to the variational principle in the ER method. The convergence behavior described above is consistent with this expectation, indeed.

According to Sec. 2, the solvation free energy $\Delta\mu$ is calculated in the ER method from the three energy distribution functions of ρ^e in the solution system and ρ_0^e and χ_0^e in the pure solvent system with test-particle insertions. Note also that $\langle u \rangle$ is given exactly by

$$\langle u \rangle = \int d\epsilon \rho^e(\epsilon), \quad (4)$$

where ϵ is the solute-solvent pair interaction energy and serves as the coordinate for the distribution functions in the ER method through Eq. 1. Roughly speaking, the effect of the attractive part of the solute-solvent interaction is taken into account by ρ^e and the effect of the repulsive part is by ρ_0^e and χ_0^e . For the Ile, Met, and Gln analog-solutes examined, our observation described above indicates that the attractive-interaction effect depends rather weakly on the solute intramolecular structure. The effect of the solute structure is mainly carried by the repulsive-interaction part.

An improvement of the RISM-based approach may be possible through combination with the molecular simulation. In the combined scheme, the site-site radial distribution functions are calculated from molecular simulation and are substituted into a RISM-based functional for the solvation free energy. It should be noted that the radial distribution functions are exact (within statistical error) under the used set of potential functions when they are obtained from the molecular simulation, instead of the integral equation. The computational efficiency for the solvation free energy was then seen to be achieved compared to the free-energy perturbation and thermodynamic integration methods [29, 30]. The procedure suffers from numerical instability of several tens of kcal/mol, however. The source of the numerical trouble is the intramolecular correlation matrices in the range of small recip-

rocal vector (large distance). When the reciprocal vector approaches zero, the intramolecular correlation matrices become ill-conditioned and the direct correlation function exhibits a long-range decay of Coulombic form [30, 49].

4. Conclusions

In the present work, we assessed the performance of end-point functionals for the solvation free energy against a benchmark dataset for amino-acid side-chain analogs in water. Among the functionals examined, the method of energy representation (ER) and the partial-wave corrected (PWC) model are the best and the second in terms of accuracy, respectively. The mean absolute deviation (MAD) of (less than) ~ 1 kcal/mol is achieved only with these two methods. In addition to the speed compared to the exact free-energy methods, the ER method is advantageous for its wide applicability; the intramolecular flexibility and/or the system inhomogeneity is not the limitation of the method [6, 32, 33]. The RISM-based schemes are more restricted in this sense, although the computational efficiency is the highest. With the informatics-based correction, the PWC model can further achieve the accuracy.

Acknowledgments

This work is supported by the Grant-in-Aid for Scientific Research (No. 21300111) from the Japan Society for the Promotion of Science, and by the Grant-in-Aid for Scientific Research on Priority Areas (No. 20038034), the Grant-in-Aid for Scientific Research on Innovative Areas (No. 20118002), and the Next-Generation Supercomputing Project, Nanoscience Program from

the Ministry of Education, Culture, Sports, Science, and Technology. N. M. is also grateful to the grant from the Association for the Progress of New Chemistry, the grant from the Suntory Institute for Bioorganic Research, and the Supercomputer Laboratory of Institute for Chemical Research. We further acknowledge Ekaterina Ratkova for help with RISM and quantum-chemical calculations.

References

- [1] C. Chipot, D. A. Pearlman, *Molecular Simulation* 28 (2002) 1.
- [2] M. R. Shirts, V. S. Pande, *J. Chem. Phys.* 122 (2005) 134508.
- [3] J. Westergren, L. Lindfors, T. Höglund, K. Lüder, S. Nordholm, R. Kjellander, *J. Phys. Chem. B* 111 (2007) 1872.
- [4] N. Matubayasi, M. Nakahara, *J. Chem. Phys.* 113 (2000) 6070.
- [5] N. Matubayasi, M. Nakahara, *J. Chem. Phys.* 117 (2002) 3605, erratum: 118 (2003) 2446.
- [6] N. Matubayasi, M. Nakahara, *J. Chem. Phys.* 119 (2003) 9686.
- [7] G. N. Chuev, M. V. Fedorov, J. Crain, *Chem. Phys. Lett.* 448 (2007) 198.
- [8] G. N. Chuev, M. V. Fedorov, S. Chiodo, N. Russo, E. Sicilia, *J. Comp. Chem.* 29 (2008) 2406.
- [9] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* 105 (2005) 2999.

- [10] B. Roux, T. Simonson, *Biophys. Chem.* 78 (1999) 1.
- [11] J. M. J. Swanson, R. H. Henchman, J. A. McCammon, *Biophys. J.* 86 (2004) 67.
- [12] P. Koehl, *Curr. Opin. Str. Bio.* 16 (2006) 142.
- [13] M. S. Lee, M. A. Olson, *J. Phys. Chem. B* 109 (2005) 5223.
- [14] M. B. Ulmschneider, J. P. Ulmschneider, M. S. P. Sansom, A. Di Nola, *Biophys. J.* 92 (2007) 2338.
- [15] J. P. Guthrie, *J. Phys. Chem. B* 113 (2009) 4501.
- [16] D. Chandler, H. C. Andersen, *J. Chem. Phys.* 57 (1972) 1930.
- [17] F. Hirata, P. J. Rossky, *Chem. Phys. Lett.* 83 (1981) 329.
- [18] L. Lue, D. Blankschtein, *J. Phys. Chem.* 96 (1992) 8582.
- [19] M. Ikeguchi, J. Doi, *J. Chem. Phys.* 103 (1995) 5011.
- [20] M. Kinoshita, F. Hirata, *J. Chem. Phys.* 104 (1996) 8807.
- [21] D. Beglov, B. Roux, *J. Phys. Chem. B* 101 (1997) 7821.
- [22] F. Hirata (Ed.), *Molecular theory of solvation*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.
- [23] S. Ten-no, S. Iwata, *J. Chem. Phys.* 111 (1999) 4865.
- [24] G. N. Chuev, M. V. Fedorov, *J. Comp. Chem.* 25 (2004) 1369.

- [25] M. V. Fedorov, H.-J. Flad, G. N. Chuev, L. Grasedyck, B. N. Khoromskij, *Computing* 80 (2007) 47.
- [26] S. J. Singer, D. Chandler, *Mol. Phys.* 55 (1985) 621.
- [27] D. Chandler, Y. Singh, D. M. Richardson, *J. Chem. Phys.* 81 (1984) 1975.
- [28] A. Kovalenko, F. Hirata, *J. Chem. Phys.* 113 (2000) 2793.
- [29] S. M. Kast, *Phys. Chem. Chem. Phys.* 3 (2001) 5087.
- [30] H. Freedman, T. N. Truong, *J. Chem. Phys.* 121 (2004) 2187.
- [31] H. Takahashi, N. Matubayasi, M. Nakahara, T. Nitta, *J. Chem. Phys.* 121 (2004) 3989.
- [32] N. Matubayasi, K. K. Liang, M. Nakahara, *J. Chem. Phys.* 124 (2006) 154908.
- [33] N. Matubayasi, W. Shinoda, M. Nakahara, *J. Chem. Phys.* 128 (2008) 195107.
- [34] R. Wolfenden, L. Andersson, P. M. Cullis, C. C. B. Southgate, *Biochemistry* 20 (1981) 849.
- [35] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 79 (1983) 926.
- [36] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* 118 (1996) 11225.

- [37] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, J. Phys. Chem. B 105 (2001) 6474.
- [38] M. R. Shirts, J. W. Pitner, W. C. Swope, V. S. Pande, J. Chem. Phys. 119 (2003) 5740.
- [39] M. E. Tuckerman, Y. Liu, G. Ciccotti, G. J. Martyna, J. Chem. Phys. 115 (2001) 1678.
- [40] M. Parrinello, A. Rahman, Phys. Rev. Lett. 45 (1980) 1196.
- [41] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, J. Chem. Phys. 103 (1995) 8577.
- [42] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, J. Comp. Chem. 4 (1983) 187.
- [43] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, J. Chem. Phys. 76 (1982) 637.
- [44] H. C. Andersen, J. Comp. Phys. 52 (1983) 24.
- [45] S. Miyamoto, P. A. Kollman, J. Comp. Chem. 13 (1992) 952.
- [46] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theory Comput. 4 (2008) 435.
- [47] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada,

M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.

[48] C. H. Bennett, *J. Comp. Phys.* 22 (1976) 245.

[49] P. T. Cummings, G. Stell, *Mol. Phys.* 46 (1982) 383.

Amino acid	Analog solute	Experimental ^{a)}	Bennett ^{b)}	ER ^{c)}	HNC	GF	RB	PW	PWC
Ala	methane	1.9	2.2	1.9	13.2	4.3	-1.1	6.9	2.7
Val	propane	2.0	2.3	2.8	25.9	5.4	-2.4	10.2	2.4
Leu	<i>iso</i> -butane	2.3	2.3	3.2	33.6	4.3	1.1	10.1	1.4
Ile	<i>n</i> -butane	2.2	2.4	3.0	33.2	5.1	-2.1	10.9	1.9
Phe	toluene	-0.8	-0.9	-0.1	33.0	1.7	-11.0	8.9	-2.0
Ser	methanol	-5.1	-4.5	-5.2	4.4	-6.9	-12.4	-0.9	-3.9
Thr	ethanol	-4.9	-4.2	-4.8	7.9	-9.1	-15.9	-2.0	-6.7
Tyr	<i>p</i> -cresol	-6.1	-5.5	-5.2	25.5	-9.0	-20.5	1.0	-8.3
Cys	methanethiol	-1.2	-0.6	-0.7	12.3	0.9	-5.5	5.2	-0.7
Met	methyl ethyl sulfide	-1.5	-0.4	-0.2	30.3	3.7	1.8	10.1	-0.6
Asn	acetamide	-9.7	-8.5	-9.2	7.4	-10.6	-18.3	-2.7	-8.6
Gln	propionamide	-9.4	-8.6	-9.2	13.3	-11.1	-19.2	-2.2	-9.7
Hid ^{d)}	4-methylimidazole	-10.3	-8.9	-9.2	15.0	-9.7	-22.0	-1.0	-9.3
Hie ^{d)}	4-methylimidazole	-10.3	-9.1	-9.1	15.0	-9.7	-22.0	-1.0	-9.3
Trp	3-methylindole	-5.9	-4.9	-4.2	36.6	-5.2	-19.4	5.2	-7.3

^{a)} Taken from Ref. [34].

^{b)} This column shows the exact values calculated using the Bennett acceptance ratio method in Ref. [2].

^{c)} The standard error is 0.1 kcal/mol or less in the ER calculation of the solvation free energy.

^{d)} Hid and Hie refer to the neutral, protonated states of His, in which the proton is attached to the δ and ϵ nitrogens, respectively. See Ref. [2] for the details. The distinction of Hid and Hie is made only in the MD and ER calculations (the columns of Bennett and ER). No distinction is made in the other cases, and the values for Hid and Hie are identical in the columns of Experimental, HNC, GF, RB, PW, and PWC.

Table 1: Solvation free energy in water in the unit of kcal/mol.

	Bennett	ER	HNC	GF	RB	PW	PWC
Experimental	0.7	0.7	24.2	2.3	7.9	7.7	1.0
Bennett	—	0.5	23.5	2.3	8.4	7.0	0.9

Table 2: Mean absolute deviation (MAD) in the unit of kcal/mol.

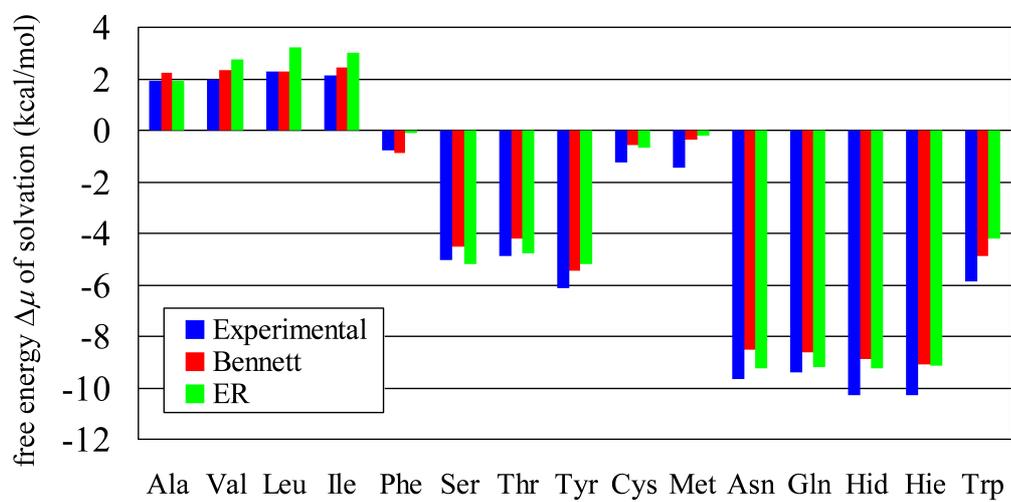


Figure 1: Solvation free energy $\Delta\mu$ in the unit of kcal/mol. Shown are the experimental, Bennett (computationally exact), and ER values.