JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

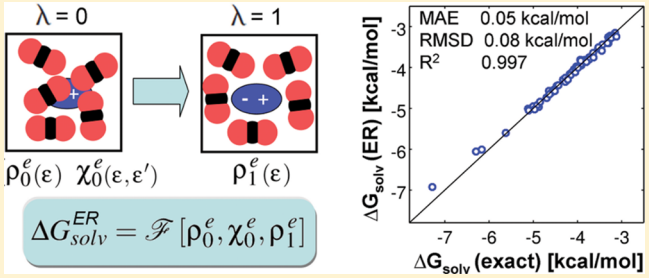
Accurate Calculation of Solvation Free Energies in Supercritical Fluids by Fully Atomistic Simulations: Probing the Theory of Solutions in Energy Representation

Andrey I. Frolov*,†

Institute of Solution Chemistry, Russian Academy of Sciences, Akademicheskaya St. 1, 153045 Ivanovo, Russia

Supporting Information

ABSTRACT: Accurate calculation of solvation free energies (SFEs) is a fundamental problem of theoretical chemistry. In this work we perform a careful validation of the theory of solutions in energy representation (ER method) developed by Matubayasi et al. [*J. Chem. Phys.* **2000**, *113*, 6070−6081] for SFE calculations in supercritical solvents. This method can be seen as a bridge between the molecular simulations and the classical (not quantum) density functional theory (DFT) formulated in energy representation. We performed extensive calculations of SFEs of organic molecules of different chemical natures in pure supercritical CO_2 (sc-CO_2) and in sc-CO_2 with addition of 6 mol % of ethanol, acetone, and *n*-hexane as cosolvents. We show that the ER method reproduces SFE data calculated by a method free of *theoretical* approximations (the Bennett's acceptance ratio) with the mean absolute error of only 0.05 kcal/mol. However, the ER method requires by an order less computational resources. Also, we show that the quality of ER calculations should be carefully monitored since the lack of sampling can result into a considerable bias in predictions. The present calculations reproduce the trends in the cosolvent-induced solubility enhancement factors observed in experimental data. Thus, we think that molecular simulations coupled with the ER method can be used for quick calculations of the effect of variation of temperature, pressure, and cosolvent concentration on SFE and hence solubility of bioactive compounds in supercritical fluids. This should dramatically reduce the burden of experimental work on optimizing solvency of supercritical solvents.



INTRODUCTION

Prediction of solvation free energy (SFE) or equivalently the excess chemical potential (ECP) of a given compound in a multicomponent solution is of crucial importance for thermodynamics and all related disciplines: e.g. chemical engineering, chemistry, and physics. Knowledge of chemical potentials provides a route to all thermodynamic parameters of the system (see, e.g., refs 1 and 2). Different equilibrium partitioning coefficients (solubilities, logP, Henry law's constants, vapor pressures, and so forth) can be calculated from SFE data.[1]

Experimental determination of SFE are laborious.[3] The data can significantly vary for different experimental techniques and independent research groups.[4] All these makes the development of a predictive method for SFE one of the most important task of computational and theoretical chemistry. One can highlight a success of the models where theoretical predictions are complemented by a cheminformatics approach with a certain number of free parameters trained against experimental data. One of the most successful models is the COSMOtherm, where data of single molecule quantum mechanical (QM) calculations in a continuum solvent modeled by the COSMO approach is plugged into a chemical thermodynamics model with a number of adjustable coefficients.[2] This model uses ∼60 free adjustable parameters and is able to predict chemical potentials with an average accuracy of 0.5 kcal/mol for small neutral solutes in different solvents.[5,6] Recently, the group of Fedorov showed that 3D and 1D reference interaction site model (RISM) predictions of hydration free energies of different classes of compounds can be extended by a set of structural descriptor corrections (SDCs) to get the prediction accuracy of 0.5−1.0 kcal/mol with less than 10 free coefficients.[7−10] Moreover, a single correction for the partial molar volume (PMV) allows obtaining an accuracy of 1.0 kcal/mol for 3D-RISM calculations.[8,9,11]

Despite a certain success there is still a big demand for truly "bottom-up" methods which would allow accurate free energy calculations without empirical parameters. It is still too demanding to describe solvation shells on the level of high quality quantum mechanical calculations. Classical molecular simulations based on classical force field (FF) representation of molecular interactions are widely used for SFE calculations. However, due to the limited accuracy of FFs the predictions suffer from large deviations from experimental data of ∼1−3 kcal/mol.[12] It seems to be possible to reduce these errors by

Received: February 20, 2015
Published: April 14, 2015

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

development of the next generation force fields, e.g. a FF where the molecular polarization is modeled explicitly.[13] Also the current FF can be parametrized better to reproduce SFE without spoiling other characteristics, see, e.g., the work of Shirts et al.[14]

Supercritical fluids are a "green" alternative to hazardous solvents in many technologies. They are used in production and processing of bioactive compounds.[16−18] The solvency of supercritical fluids can be widely regulated by variation of the state parameters or by an addition of cosolvents.[17,19,20] However, the advantages of supercritical solvents come at a cost of the need to scan multiparameter space (temperature, pressure, cosolvent(s) concentration(s)) to find optimal state parameters and cosolvents for a given process. Prediction of cosolvent effect on the solubility of organic compounds in supercritical fluids is an important area of application of bottom-up simulation approaches. Indeed, in this field there are not so much experimental data available as in the case of aqueous solutions which challenges the development of semiempirical equations and hybrid models. Su and Maroncelli showed that molecular simulations with FF models are able to predict SFE of organic compounds in supercritical media with a reasonable accuracy of ∼0.5 kcal/mol.[21] Andersen and Siepmann used grand canonical Monte Carlo simulations to calculate partitioning of organic compounds to supercritical CO_2 with addition of cosolvents, which reasonably correlates with the experimental observations.[22]

Recently we performed extensive calculations of SFE of monofunctional organic molecules at infinite dilution in supercritical media based on fully atomistic MD simulations.[23] A series of monofunctional organic molecules were solvated in pure sc-CO_2 and sc-CO_2 with addition of 6 mol % of cosolvents of different chemical nature. The data was helpful to rationalize how different types of solute−solvent interactions contribute to the cosolvent induced solubility enhancement (CISE) factors.[23] However, the calculations are computationally demanding and it would be highly desirable to reduce the computational consumptions of free energy calculations for future applications.

For a single SFE calculation the conventional the free energy perturbation methods require performing several simulations with a continuous coupling of a solute molecule to the rest of the system. However, this is undesirable not only from the point of view of considerable computational consumptions but also for rationalizing the simulations, because no such process can take place in reality. That is why such calculations are usually referred to as *alchemical* free energy calculations.[15]

Matubayasi et al. developed a theory of solutions in energy representation (ER method) which allows to skip all intermediate simulations in the process of solute coupling[24−27] and thus requires by an order less computational resources compared to conventional free energy perturbation methods. However, the gain in performance comes on a cost of introducing theoretical approximations. In this work we benchmark the free energy calculations by ER method for different monofunctional solutes in supercritical carbon dioxide with addition of cosolvents of different polarity against the data of a theoretically rigorous Bennett's acceptance ratio method. Also, we analyze the origins of the theoretical errors in ER calculations and the sensitivity of the predictions to the length of MD simulations.

CHOICE OF SYSTEMS

In the present work we performed SFE calculations for the same systems which we studied by the Bennett's acceptance ratio (BAR) method in the previous work.[23] Briefly, we chose one point on the sc-CO_2 phase diagram at a relatively high fluid density and low temperature, where one observes high solvency of the fluid: $P = 3.39P_{crit} = 250$ bar and temperature $T = 1.01T_{crit} = 308.15$ K (the corresponding density is $\sim 1.93\rho_{crit} = 900.0$ kg/m^3). These state parameters are often used in experimental studies.[19,28] We consider a set of monofunctional hydrocarbons. The list of solutes studied in this work is present in Table S1 of the Supporting Information (SI). In the present study we consider three cosolvents: ethanol, acetone, and *n*-hexane. Ethanol molecule is of moderate polarity (dipole moment ∼1.7 D) and is able to act as an acceptor and donor of protons in hydrogen bonding. Acetone molecule is highly polar (dipole moment ∼2.9 D) and is able to act only as proton acceptor in hydrogen bonding. *n*-Hexane is nonpolar, does not form hydrogen bonds, and interacts with surrounding molecules mostly via van-der-Waals interactions. The cosolvent concentration was set to 6 mol %.

METHODS

Definition of the Solvation Free Energy (SFE). For the sake of simplicity let us consider an infinite diluted solution of a solute *s* in a one component solvent *w*. We consider a system with $N_s = 1$ solute and N_w solvent molecules in isothermo−isobaric ensemble (NPT-ensemble) at pressure *P*. We model molecular interaction on the level of the classical force field approximation, which implies that a molecule is represented as a set of atomic sites interacting via spherically symmetric potentials. In the present study we define the SFE according to Ben-Naim[1] as the free energy associated with switching on the interactions between a *selected* solute molecule and the rest of the system:

$$\Delta G_{solv} = -k_B T \ln \frac{\Delta(N_s, N_w, P, T, \lambda = 1)}{\Delta(N_s, N_w, P, T, \lambda = 0)} \quad (1)$$

where k_B is the Boltzmann constant, T is temperature, $\Delta(N_s, N_w, P, T, \lambda)$ is the partition function in isothermo−isobaric ensemble, λ indicates that the partition functions are written with the λ-parametrized Hamiltonian (see eq 21 in the SI), where λ is a parameter which controls coupling of the interaction between a selected solute molecule and the rest of the solution.

Since the ideal gas contributions of the solute molecule cancel out (see ref 29), this free energy is equal to the free energy of an imaginary process of transferring one solute molecule from a fixed position of center of mass in the gas phase to a fixed position of center of mass in the solution. Moreover, by accident the free energy difference in this process is equal to the excess chemical potential of solute molecules in the system $\Delta G_{solv} \equiv \mu_{ex}$ (please, see the proof in ref 29). This identity preserves also for finite solute concentrations and when the solute's internal molecule degrees of freedom are affected by solvation, in contrast to what is sometimes wrongly believed[1] (see ref. 29 for more details).

Here one should note that in NPT ensemble introducing a single solute molecule to solution results in an increase of the system's volume. Therefore, an additional contribution to the excess chemical potential appear: the thermodynamic work performed by an ideal gas particle to expand the volume by the

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

partial molar volume of solute (see formula 19 of ref 29). However, this contribution is system size dependent and vanishes in the thermodynamic limit. Shirts et al.[14] showed that it is small even compared to the *sampling error* of simulations for regular box sizes and thus we neglect it here.

Bennett's Acceptance Ratio (BAR) Method. BAR method[30] was found to be one of the most efficient methods for SFE calculations using classical force fields.[31,32] In this method the free energy (or free enthalpy) differences with an optimal *sampling error* between two states characterized by different solute−solvent interaction coupling parameters can be found by self-consistent solution of eqs 12a and 12b of ref 30. When the energy differences are sampled in the canonical ensemble the result is the free energy difference, in the case of the isothermo−isobaric ensemble it is the Gibbs free energy (free enthalpy).[31] For one SFE calculation this method usually requires simulation of more than 10 independent intermediate systems (stages) with different solute−solvent coupling λ parameters, which makes this method highly demanding in terms of computational resources. The SFE is calculated as the sum of the free energy changes between all intermediate stages from zero to full coupling of the solute−solvent interactions.

Theory of Solutions in Energy Representation (ER Method). Matubayasi et al. developed ER method, which allows to avoid the simulation of intermediate stages of the free energy perturbation path.[24−27,33] The ER method can be seen as a bridge between the molecular simulations and the *classical* (not quantum) density functional theory (DFT).[34] In ER method one defines a new collective coordinate—ϵ—the interaction energy of a solute and a solvent molecule. The configuration of solvent molecules at the solute can be characterized by the density distribution function over this coordinate. SFE is expressed as a functional of the solute−solvent density distribution defined over the collective variable ϵ. The latter can be approximated by the Percus's method of functional expansion, which leads to the end-point (not dependent on the λ-coupling path) free energy expression. Importantly, the definition of the collective coordinate should be λ-independent, such that this coordinate is calculated with the solute−solvent potential at full coupling $v_{sw}(\mathbf{x}_s, \mathbf{x}_w)$, irrespective of the ensemble and Hamiltonian for which this configuration was generated:

$$\epsilon = v_{sw}(\mathbf{x}_s, \mathbf{x}_w) \equiv u_{sw,\lambda=1}(\mathbf{x}_s, \mathbf{x}_w) \tag{2}$$

where $u_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w)$ denotes the λ-dependent solute−solvent interaction potential, s stands for solute, w denotes solvent, $\mathbf{x}_s$ and $\mathbf{x}_w$ collectively define spatial position and orientation of the solute and solvent molecules (see eq 19 of the SI).

For a single configuration of the system the microscopic density in energy representation can be written as (here and after superscript e denotes that the function is defined over the energy coordinate):

$$\hat{\rho}^e_{sw}(\epsilon) = \sum_{i=1}^{N_w} \delta(v_{sw}(\mathbf{x}_s, \mathbf{x}_w) - \epsilon) = \int_{-\infty}^{+\infty} d\mathbf{x}_s\, d\mathbf{x}_w \delta(v_{sw}(\mathbf{x}_s, \mathbf{x}_w) - \epsilon)\hat{\rho}_{sw}(\mathbf{x}_s, \mathbf{x}_w) \tag{3}$$

where $\hat{\rho}_{sw}(\mathbf{x}_s, \mathbf{x}_w)$ is the microscopic density in the full coordinate representation.

The solute−solvent interaction potential in energy representation can be written as

$$u^e_{sw,\lambda}(\epsilon) = \int_{-\infty}^{+\infty} d\mathbf{x}_s\, d\mathbf{x}_w \delta(v_{sw}(\mathbf{x}_s, \mathbf{x}_w) - \epsilon) u_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w) \tag{4}$$

It is important for the following derivation that we choose the λ path in such a way that $u_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w)$ is constant on each equi-energy surface of $v_{sw}(\mathbf{x}_s, \mathbf{x}_w)$. This can be achieved, for instance, when $u_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w) = \lambda \cdot v_{sw}(\mathbf{x}_s, \mathbf{x}_w)$. With this restriction, one can write the following identity:

$$u_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w) = \int_{-\infty}^{+\infty} d\epsilon \delta(v_{sw}(\mathbf{x}_s, \mathbf{x}_w) - \epsilon) u^e_{sw,\lambda}(\epsilon) \tag{5}$$

The solute−solvent density distribution in the NPT ensemble is written as

$$\rho^e_{sw,\lambda}(\epsilon) = \langle \hat{\rho}_{sw}(\epsilon) \rangle_\lambda = \int_{-\infty}^{+\infty} d\mathbf{x}_s\, d\mathbf{x}_w \delta(v_{sw}(\mathbf{x}_s, \mathbf{x}_w) - \epsilon) \rho_{sw,\lambda}(\mathbf{x}_s, \mathbf{x}_w) \tag{6}$$

where $\langle ... \rangle_\lambda$ denotes ensemble average with λ-parametrized Hamiltonian at given λ.

In the distribution function formalism SFE is expressed via the density distribution using the Kirkwood's charging formula (see eq 28 of the SI). One can prove[33] that the same relation holds in energy representation (see derivation in the SI):

$$\Delta G_{solv} = \int_0^1 d\lambda \int_{-\infty}^{+\infty} d\epsilon \frac{\partial u^e_{sw,\lambda}(\epsilon)}{\partial \lambda} \rho^e_{sw,\lambda}(\epsilon) \tag{7}$$

SFE in eq 7 can be considered as a functional of the solute−solvent density distribution. This implies that there should be only one $u^e_{sw,\lambda}(\epsilon)$ to which a given $\rho^e_{sw,\lambda}(\epsilon)$ corresponds. However, both in coordinate and energy representation it is not the case if one considers ensembles where the number of particles is fixed.[26,35−37] Matubayasi proposed a way how to retain the one-to-one $\rho-u$ correspondence by introducing additional condition based on the physical sense. First, he showed that the potentials which gives different density distributions can differ only by an additive constant (see the Appendix of ref 35 and ref 26). Second, the additive constant is fixed to ensure that the solute−solvent pair potential reaches zero when the particle separation tends to infinity. With this approach the one-to-one correspondence between $u_{sw,\lambda}(\epsilon)$ and $\rho_{sw,\lambda}(\epsilon)$ is retained. This allows us to consider the potential $u^e_{sw,\lambda}(\epsilon)$ as a functional of $\rho^e_{sw,\lambda}(\epsilon)$ in a fixed-N ensemble and use the functional calculus to obtain approximate free energy functionals. Therefore, SFE (eq 7) can be written as a density functional of the solute−solvent density distribution (please see more details in the SI):

$$\Delta G_{solv}[\rho^e_{sw,\lambda}(\epsilon)] = \int_{-\infty}^{+\infty} d\epsilon\, \rho^e_{sw,\lambda=1}(\epsilon)\epsilon - \mathcal{F}[\rho^e_{sw,\lambda}(\epsilon)] \tag{8}$$

where the first term is the average interaction energy between a solute and the solvent molecules and the all the higher-order correlation terms are contained in the $\mathcal{F}[\rho^e_{sw,\lambda}(\epsilon)]$functional.

We can introduce an auxiliary function $w^e_{sw,\lambda}(\epsilon)$, which is an analogue of the indirect part of potential of mean force (IPMF) in coordinate representation:

$$\rho^e_{sw,\lambda}(\epsilon) = \rho^e_{sw,\lambda=0}(\epsilon) \exp[-\beta(u^e_{sw,\lambda}(\epsilon) + w^e_{sw,\lambda}(\epsilon))] \tag{9}$$

The functional in eq 8 can be written via IPMF. Here and after, we use the following simplified notations:

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

$$\rho^{e}_{sw,0} \equiv \rho^{e}_{sw,\lambda=0}$$

$$\rho^{e}_{sw} \equiv \rho^{e}_{sw,\lambda=1}$$

Similar notations are adopted for other functions.

Equation 8 can be further simplified if one chooses the λ-dependence of the potential such that the density distribution is a linear function of λ:

$$\rho^{e}_{sw,\lambda}(\epsilon) = \lambda\rho^{e}_{sw}(\epsilon) + (1-\lambda)\rho^{e}_{sw,0}(\epsilon) \tag{10}$$

The functional in eq 8 then reads:

$$\mathcal{F}[\rho^{e}_{sw,\lambda}(\epsilon)] = k_B T \int_{-\infty}^{+\infty} d\epsilon \Bigg[(\rho^{e}_{sw}(\epsilon) - \rho^{e}_{sw,0}(\epsilon)) - \rho^{e}_{sw}(\epsilon) \ln \frac{\rho^{e}_{sw}(\epsilon)}{\rho^{e}_{sw,0}(\epsilon)} - \beta(\rho^{e}_{sw}(\epsilon) - \rho^{e}_{sw,0}(\epsilon)) \int_0^1 d\lambda\, w^{e}_{sw,\lambda}(\epsilon) \Bigg] \tag{11}$$

The exact free energy functional (eq 11) contains the term which depends on λ. To eliminate the λ-dependence we apply the Percus method of functional expansion to formulate approximate functionals. Following the work of Percus[38] we obtain the HNC-like approximation by expanding $\ln \rho^{e}_{sw}(\epsilon) + \beta u_{sw}(\epsilon)$ functional in powers of density fluctuations $\rho^{e}_{sw}(\epsilon) - \rho^{e}_{sw,0}(\epsilon)$ (see the SI). The HNC-like approximation of IPMF is written as

$$w^{e,HNC}_{sw}(\epsilon) = -k_B T \Bigg[\frac{\rho^{e}_{sw}(\epsilon) - \rho^{e}_{sw,0}(\epsilon)}{\rho^{e}_{sw,0}(\epsilon)} - \int_{-\infty}^{+\infty} d\epsilon' \cdot [\rho^{e}_{sw}(\epsilon') - \rho^{e}_{sw,0}(\epsilon')] (\chi^{e}_{sww,0})^{-1}(\epsilon, \epsilon') \Bigg] \tag{12}$$

where $(\chi^{e}_{sww,0})^{-1}(\epsilon, \epsilon')$ is the functional inverse of the density–density correlation function at zero solute coupling.

With the linear dependence of $\rho^{e}_{sw,\lambda}$ on λ (eq 10) the λ-dependent IPMF in HNC-like approximation can be written via $w^{e,HNC}_{sw}$ at full solute coupling:

$$w^{e,HNC}_{sw,\lambda}(\epsilon) = \lambda \cdot w^{e,HNC}_{sw}(\epsilon) \tag{13}$$

Thus, the λ-integral in eq 11 can be written in HNC-like approximation as

$$\beta \int_0^1 d\lambda\, w^{e,HNC}_{sw,\lambda}(\epsilon) = \frac{1}{2}\beta w^{e,HNC}_{sw}(\epsilon) \tag{14}$$

Following the work of Percus,[38] we obtain the Percus–Yevick-like (PY-like) approximation by expanding the $\rho^{e}_{sw}(\epsilon)\, e^{\beta u_{sw}(\epsilon)}$ functional. The PY-like approximation can be written via the HNC-like IPMF (see the SI for more details):

$$w^{e,PY}_{sw}(\epsilon) = -k_B T \ln(1 - \beta w^{e,HNC}_{sw}(\epsilon)) \tag{15}$$

The λ-integral in eq 11 can be written in PY-like approximation as

$$\beta \int_0^1 d\lambda\, w^{e,PY}_{sw,\lambda}(\epsilon) = -\ln[1 - \beta w^{e,PY}_{sw}(\epsilon)] + 1 + \frac{\ln[1 - \beta w^{e,PY}_{sw}(\epsilon)]}{\beta w^{e,PY}_{sw}(\epsilon)} \tag{16}$$

There is a general knowledge in the field that in the case of simple liquids the PY approximation works better for short-range repulsive potentials, while HNC approximation performs better for long-range attractive potentials.[34] Matubayasi et al.[25,33] decided to use the PY-like expression for the λ-integral in the unfavorable region of solvation ($w^{e}_{sw} \geq 0$) and HNC-like expression for the λ-integral in the favorable region of solvation ($w^{e}_{sw} < 0$). IPMF w^{e}_{sw} can be determined from unbiased molecular simulations (regular MD or Monte Carlo) only outside of the solute-core region (region of very large solute–solvent interaction energies ϵ). Matubayasi et al.[33] proposed to use HNC-like approximation of the potential of mean force when w^{e}_{sw} is not resolved by molecular simulations. In HNC-like approximation $w^{e,HNC}_{sw}$ is determined by the solute–solvent density distribution $\rho^{e}_{sw,0}$ and the inverse of the density–density correlation function $(\chi^{e}_{sww,0})^{-1}$ in the case of the zero solute–solvent coupling (see eq 12, note this function is used when $\rho^{e}_{sw} \ll \rho^{e}_{sw,0}$). The latter can be in the most convenient way calculated by the insertion of the solute molecule configurations into the ensemble of pregenerated pure solvent configurations.

Combination of the two different expressions for λ-integral and two different input w^{e}_{sw} functions results into a hybrid functional which consists of four parts. The final expression for SFE comprises several expressions, eqs 8 and 11, where the λ-integral is approximated by the following expression:

$$\beta \int_0^1 d\lambda\, w^{e}_{sw,\lambda}(\epsilon) = \alpha(\epsilon) \cdot F_w(\epsilon) + [1 - \alpha(\epsilon)] \cdot F_{wHNC}(\epsilon) \tag{17}$$

where the functions F_w and F_{wHNC} are written as the combination of PY and HNC-like expressions for the λ-integral. The parameter $\alpha(\epsilon)$, which is responsible for merging parts with different w^{e}_{sw} functions, is set heuristically.[33] The explicit forms of the functions are provided in the SI. The extension of the ER method to the case of multicomponent solvent is presented in ref 35.

Sources of Errors in Calculations. Total error of computer calculations with respect to accurate experimental data is the sum of several independent contributions: *modeling error*—the errors associated with the accuracy of the potential model including truncation of the potentials, use of Ewald sum, etc.; *theoretical error*—the errors resulting from theoretical approximations (e.g., mean field, HNC, PY, etc.) which are usually difficult to estimate; *finite sampling error*—the error associated with finite sampling; and *numerical error*—the errors resulting from the use of different numerical schemes (e.g., second-order MD integrators, function discretization, etc.). The later error is usually assumed to be small when one uses recommended values of the corresponding parameters and we disregard this in the forthcoming analysis.

Both BAR and ER calculations based on MD sampling rely on the classical "force field" description of interactions and therefore suffer from the *modeling error*, which can be estimated only by comparison with experimental data. BAR calculations are free of theoretical approximations and therefore apart from *modeling error* suffer only from *finite sampling error*. The later can be controlled in the course of simulations by changing the simulation time and starting from different initial configurations. Additionally, the ER method also introduces a *theoretical error*, because it uses a combination of HNC-type and PY-type approximations in energy representation. In this study we estimate *finite sampling* and *theoretical errors*. The

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245–2256

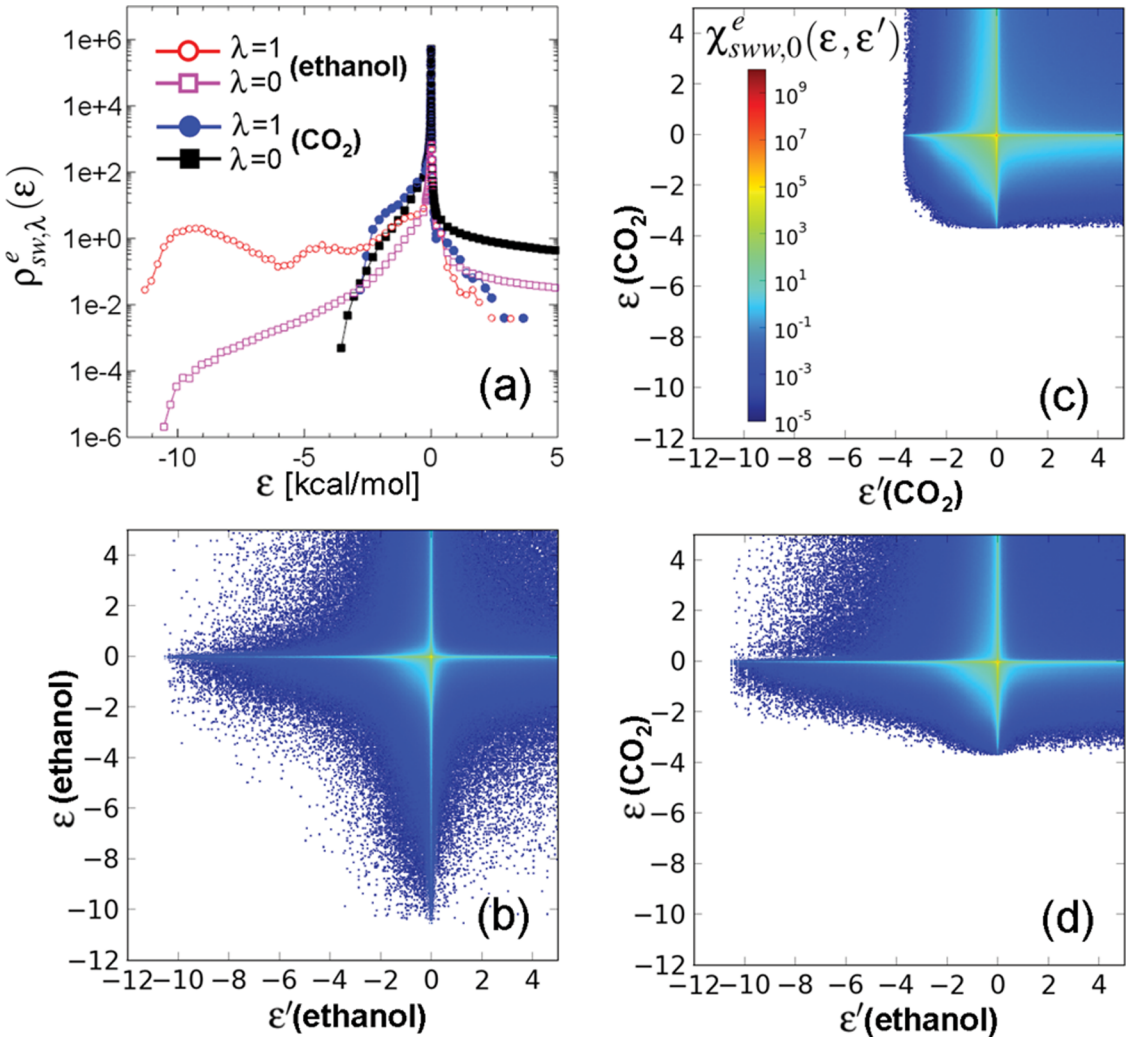


Figure 1. Example of the distribution functions in energy representation which enter the approximate free energy functional (see eqs 11, 12, 17). (a) Solvent density distributions ρ^{e}_{sw} ($\lambda = 1$) and $\rho^{e}_{sw,0}$ ($\lambda = 0$) in energy representation for butyric-acid in sc-CO_2 with addition of 6 mol % of ethanol. (b) Ethanol−ethanol, (c) CO_2−ethanol, and (d) CO_2−CO_2 density−density correlation functions $\chi^{e}_{sww,0}$ at $\lambda = 0$. The 2D distributions are normalized by a factor proportional to the product of the number of molecules of the corresponding species.

modeling error is discussed in section "Comparison with Experiment" and in the previous study.[23]

Simulation Details. *MD Simulations.* MD simulations were performed in Gromacs 4.6.5 program package compiled in double precision.[39] We used stochastic leapfrog integrator with 2 fs time step, friction coefficient of 0.5 ps^{-1} (see Gromacs manual[40]), and temperature of 308.15 K. A constant pressure of 250 bar was maintained by a Parrinello−Rahman barostat with characteristic oscillation time (τ_p) of 10.0 ps (the compressibility of the fluid was set to 7.8×10^{-4} bar^{-1}). All bonds with hydrogen atoms were constrained by the SHAKE method with the tolerance of 10^{-8}. The periodic boundary conditions and a cubic simulation box were used. The neighbor list updated every ten steps was used for the short-ranged interactions. The interactions via Lennard-Jones potential were calculated with an atomic-based cutoff switched off in the interval 1.0−1.2 nm. The mean-field long-range correction was added to energy and pressure to eliminate finite-range cutoff effects. The smooth particle mesh Ewald method was used to account for electrostatic interactions[41] with the real-space cutoff (R_c) of 1.2 nm, relative accuracy ("ewald_rtol" parameter in Gromacs notation[40]) of 10^{-6}, B-spines of the 6th order, and mesh spacing close to 0.1 nm.

We used the same potential models as in our previous study.[23] Zhang and Duan model of carbon dioxide was employed[42] (please see the implementation details in ref 23). For other molecules we used the OPLS-AA force field (OPLSAA 2005 in Shrödinger's notation)[43] as assigned by the "ffld_server" tool of Shrödinger's Maestro package.[44] The topology was converted to the Gromacs format using the "ffconv.py" tool.[23] The combined Lennard-Jones parameters between atoms of CO_2 and the atoms of other molecules were estimated using the geometric mean rule.

System Preparation and Collection of Statistics. The initial solvent configurations were prepared by placing 401 CO_2 and 26 cosolvent molecules (where applicable) in the 3.2 nm × 3.2 nm × 3.2 nm box with the help of the Packmol program.[45] For each simulation of the BAR series we performed 10 ps equilibration in NVT ensemble, 500 ps equilibration in NPT ensemble, and 10 ns of production run simulations in NPT ensemble. The configurations were stored each 0.2 ps for further analysis. For three systems, butyric acid, phenol, and 1-pentanol in sc-CO_2 with ethanol, the production simulation

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

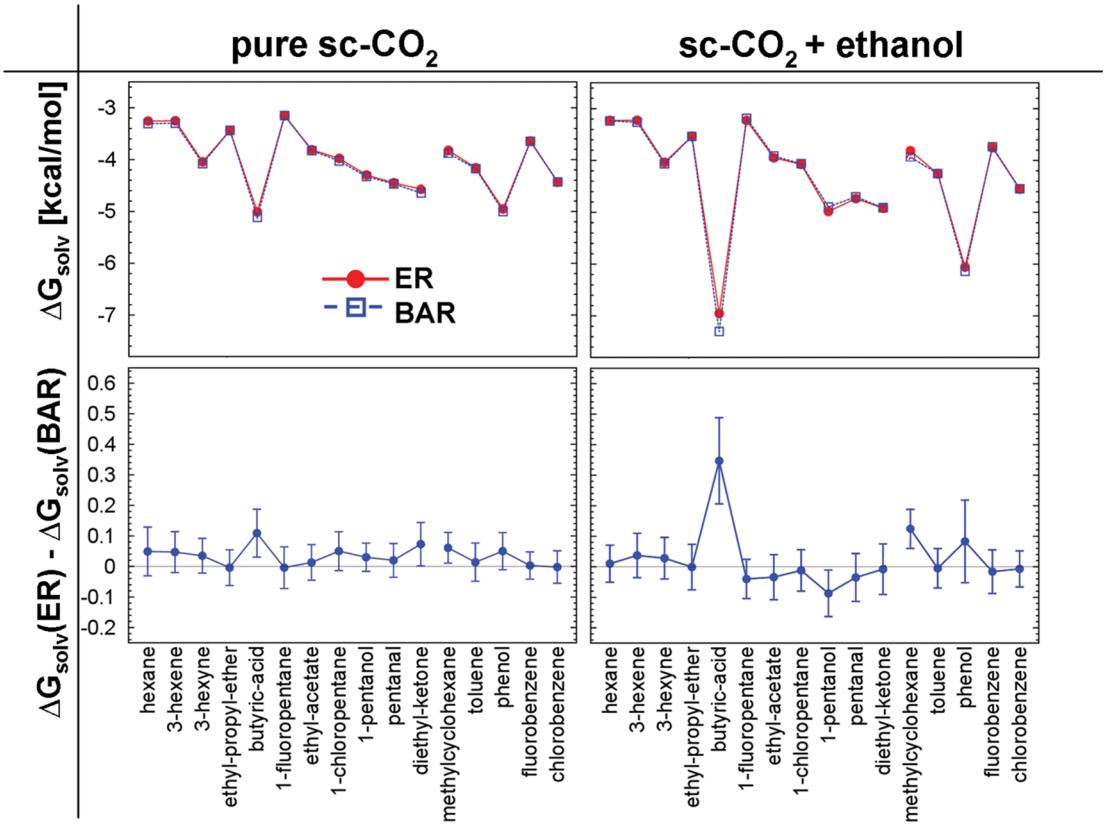


Figure 2. (top) SFE of studied organic compounds calculated with ER (red circles) and BAR (blue squares) methods for pure sc-CO_2 (first column) and sc-CO_2 with 6 mol % of ethanol (second column). Analogous plots for other solvents are presented in Figure S1 of the SI. Here and after the lines are drawn to guide the eye. The molecules are split into two groups: solutes with linear backbone and solutes with cyclic backbone, to facilitate the comparison. In each group the solutes are sorted in the order of increase of the molecule dipole moment (see Table S2). (bottom) Difference between the SFE calculated with ER and BAR methods. The error bars represent the 99% confidence intervals.

time was extended by 10 ns to decrease the statistical uncertainty. The single solute molecule MD simulations were performed for 100 ns and configurations were sampled each 2 ps.

Bennett's Acceptance Ratio Calculations. For each SFE calculation, we performed 12 independent MD simulations differed from each other by the values of the scaling coefficients λ for potential functions describing solute−solvent interactions (λ equals 0 and 1 correspond to fully decoupled and coupled solute−solvent interactions, correspondingly). The λ dependence of the nonbonded interaction potentials was adopted from ref 31 (see eq 4 therein), with two scaling coefficients: one is for scaling the *soft-core* Lennard-Jones (λ_{LJ}) and another is for scaling the Coulomb interactions (λ_C). Therefore, the potential energy expression for an intermediate stage can be characterized by the corresponding pair of values $\{\lambda_{LJ}, \lambda_C\}$.

In our previous work we calculated the total SFE in the stated systems (ΔG_{solv}) as the free energy difference between the stages {1, 1} and {0, 0}. Also, we calculated the contribution of excluded volume and van-der-Waals interactions to the total SFE ΔG^{vdW}_{solv}, which were defined as the free energy differences between systems {1, 0} and {0, 0}.

In the present work, we additionally calculated the excluded volume contribution to SFE (ΔG^{WCA}_{solv}) for some systems by performing BAR calculations for solute interacting with the repulsive part of the Weeks−Chandler−Andersen (WCA) potential.[46] This functionality was additionally implemented in the Gromacs code and this code successfully reproduced ΔG^{WCA}_{solv} of cyclohexane in water as reported in ref 47. For ΔG^{WCA}_{solv} calculations we used the following sets of scaling parameters: {0.0, 0.0}, {0.2, 0.0}, {0.3, 0.0}, {0.4, 0.0}, {0.5, 0.0}, {0.6, 0.0}, {0.7, 0.0}, {0.8, 0.0}, and {1.0, 0.0}. The energy differences were stored each 0.2 ps and the calculation of the Gibbs free energy differences between neighboring stages were later performed with the "g_bar" tool of Gromacs.[39] The resulting value was calculated as a sum of the Gibbs free energy changes between all neighboring stages in the BAR series.

ER Calculations. SFE calculations with ER method require three independent simulations: (1) simulation of the solution system at full solute coupling, (2) simulation of the pure solvent system (one simulation for all studied solutes), (3) simulation of a single solute molecule in vacuum. Energy distribution functions were calculated by "ermod" module of the ERmod program package (version 0.3.8-α8) based on the trajectories of the MD simulations.[35] The solute−solvent density distributions ρ^e_{sw} at full solute coupling were calculated based on MD simulation of solution. The solute−solvent density distribution $\rho^e_{sw,0}$ and the density−density correlation functions $\chi^e_{sww,0}$ at zero solute coupling were calculated by inserting single molecule configurations from the single molecule MD trajectory at random position and random orientation to each configuration of the pure solvent MD trajectory.

The default ERmod resolution options were employed for constructing the distribution functions. Six regions with different resolutions are defined on the energy coordinate

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

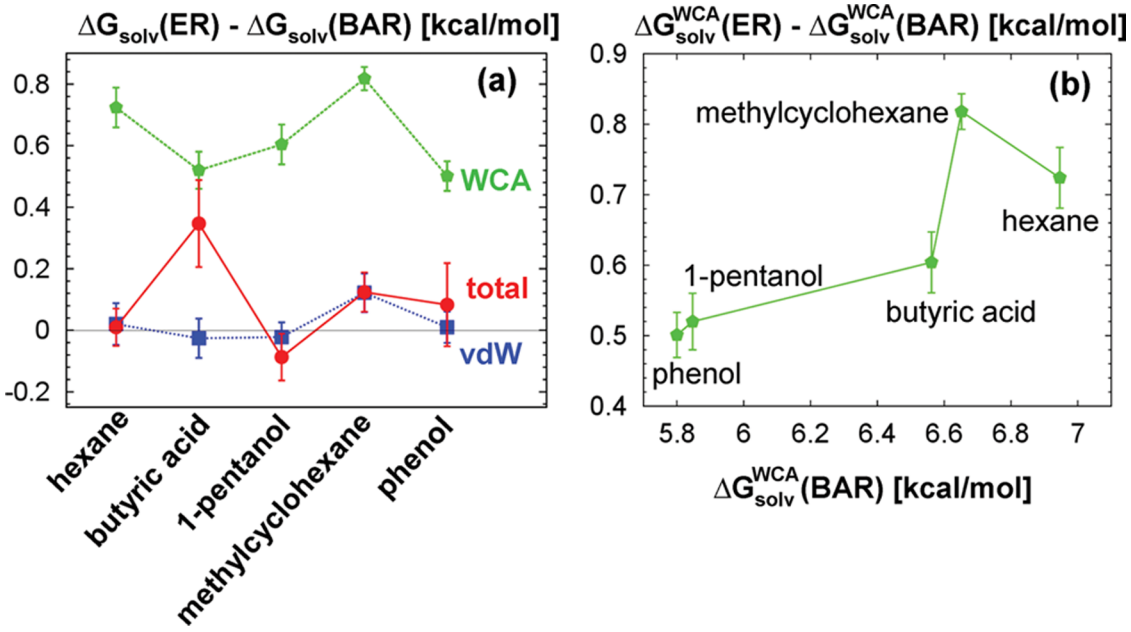


Figure 3. (a) Difference between ΔG_{solv}, ΔG^{vdW}_{solv}, and ΔG^{WCA}_{solv} calculated with ER and BAR methods for some solutes in sc-CO_2 with 6 mol % of ethanol. The errorbars represent the 99% confidence intervals. (b) Difference of ΔG^{WCA}_{solv} plotted versus the absolute value of free energy of cavity formation calculated by BAR method. The difference for methylcyclohexane outlays from the systematic deviation observed for the rest of the solutes. The examples of the distribution functions calculated for different types of potentials are shown in Figure S6 of the SI.

(values are in kilocalories per mole): (1) coarse part with the bin width of 5×10^{-2} starting from −20.0 up to −0.2, (2) fine part with the bin width of 2×10^{-3} up to −0.02, (3) very fine part with the bin width of 2×10^{-4} up to 0.02, (4) fine part with the bin width of 2×10^{-3} up to 0.2, (5) coarse part with the bin width of 5×10^{-2} up to 20, and (6) logarithmic part up to 10^{11} with the bin width increasing exponentially.

For the SFE calculations the resolution of the distribution functions is decreased by merging each 3 subsequent bins into a single bin and the effect of the discretization level on the *numerical error* of SFE calculations is controlled by the "slvfe" program.[35] The SFE was calculated based on the energy distribution functions with the help of the slvfe program. The example of distribution functions in energy representation are shown in Figure 1.

Finite Sampling Error Estimate. The finite sampling errors of the BAR and ER method calculations were estimated by the block averaging technique with 10 blocks of 1 ns each, except of three systems: butyric acid, phenol, and 1-pentanol in sc-CO_2 with ethanol at full solute coupling, where each block was 2 ns. Throughout the manuscript we report the error bar of $3\sigma_M$, where $\sigma_M = \sigma/\sqrt{10}$ with σ being the standard deviation of the values of different blocks. Throughout the text we assume that the 99% confidence interval corresponds to $3\sigma_M$; however, we note that a more rigorous way to estimate the 99% interval should be based on the Student's *t*-distributions.

RESULTS AND DISCUSSION

Accuracy of ER Method. *Theoretical Accuracy.* We analyzed the theoretical accuracy of the ER method for calculation of SFEs of organic molecules in pure sc-CO_2 and sc-CO_2 with cosolvents (see Figure 2 and Figure S1 of SI). Results show that the confidence intervals of the SFE calculated by the ER and BAR methods overlap in most cases. The maximum gap between the zero line and the 99% confidence intervals of SFE difference in pure sc-CO_2 is 0.04 kcal/mol for butyric acid. The discrepancy between ER and BAR calculations remains small even in the case of mixed solvents, where the maximum gap is 0.20 kcal/mol for butyric acid in sc-CO_2 with ethanol. Among all systems the mean absolute error (MAE) is 0.05 and the maximum deviation is 0.35 ± 0.14 kcal/mol. Therefore, we may conclude that SFE calculations are not significantly affected by the approximations behind the ER method. One can see a systematic *theoretical error* of ER calculations only in the case of butyric acid and methylcyclohexane.

Analysis of Contributions to the Theoretical Error of the ER Method. In order to rationalize why the theoretical error appears in the case of particular solutes, we analyzed how accurately the ER method can calculate different contributions to SFE associated with different types of interactions. We calculated the free energy of switching on all but electrostatic interactions between a solute and the solvent molecules ΔG^{vdW}_{solv}, and the excluded volume contribution ΔG^{WCA}_{solv} (see Figure 3 and Figure S2 of the SI). First, the theoretical error for butyric-acid vanishes when electrostatic interactions are switched off. This shows that the theoretical error has an electrostatic origin. However, in the present models hydrogen bonding is also modeled by interaction of partial charges. Therefore, we can not distinguish which interactions: purely electrostatic multipole−multipole (which include dipole−dipole, dipole−quadrupole, etc.), hydrogen-bonding, or both contribute to the error.

Second, the error in ΔG^{vdW}_{solv} for the case of methylcyclohexane remains the same as in the case of the total SFE, which shows a nonelectrostatic origin of this theoretical error. The calculated ΔG^{WCA}_{solv} show that the present functional of ER method significantly overestimates the excluded volume contribution to SFE for all solutes. This also suggests that the current free energy functional benefits from the error cancellation in SFE calculations. However, one should note that due to the hybrid nature of the free energy functional (eq 17), it is very likely that different approximations (PY-like and HNC-like) are used in different proportion for solutes with the

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

purely repulsive WCA and usual Lennard-Jones potentials. This complicates direct comparison of the theoretical errors for different types of potentials.

Figure 3b shows the theoretical error of the ΔG_{solv}^{WCA} calculations of ER method versus the absolute value of ΔG_{solv}^{WCA}. Overall, the error increases with the increase of ΔG_{solv}^{WCA}. However, one can see that the error for methylcyclohexane deviates from the general trend. This allows us to suggest that the systematic deviations of the methylcyclohexane ΔG_{solv}^{vdW} are related to the inaccuracies of calculation of the excluded volume contribution to SFE. However, this hypothesis should be further verified in the forthcoming work.

Benchmarking and Timing. *Comparison of Statistical Uncertainties of the ER and BAR Calculations.* SFE calculation with ER method requires the distribution functions of the system at $\lambda = 1$ and $\lambda = 0$. Additionally, the solvent−solvent density correlation functions have to be estimated at $\lambda = 0$. However, in the current implementation[35] the functions at $\lambda = 0$ are calculated by insertion of solute configurations from a single solute molecule MD trajectory to configurations of pure solvent generated by MD simulation. Single molecule MD is very fast even on a PC and the pure solvent MD can be performed only once for all solvents under investigation. Moreover, since solute configurations can be inserted in each solvent configuration at different random positions and orientations many times, the solvent MD trajectory should not be very long (though, the configurations should sufficiently represent the thermodynamic ensemble).

For the sake of comparison in this study we collected statistics for BAR and ER calculations from identical simulations. The data were stored with the same interval, therefore the number of data points for constructing the energy distributions at $\lambda = 1$ for ER calculations and the average energy differences for BAR calculations were the same. The pure solvent simulation was of the same length as the simulation with $\lambda = 1$; however, a solute configuration was inserted 1000 times to each pure solvent configuration to construct the functions at $\lambda = 0$. With this setup the statistical uncertainties of ΔG_{solv} calculated by the two methods are comparable. The 99% confidence intervals in SFE calculations do not exceed 0.05 kcal/mol for the most of the solutes for both methods (see Figure S3 of the SI). However, when solute molecules form a significant number of hydrogen bonds with cosolvent molecules (see Table S2 of SI), there is a considerable increase of the confidence interval.

Sensitivity of the BAR and ER Calculations to the Simulation Length. We investigated how sensitive is the statistical error to the number of trajectory configurations used for SFE calculations (see Figure 4 and Figures S4 and S5 of the SI). There are three important observations:

1. One can see that ∼2.5−5 × 10^3 configurations from the solution trajectory are sufficient to get a prediction with the ER method fairly close to the reference value. The same number of configurations per λ-stage simulation is required for confidence interval of BAR calculations to reach a plateau.
2. The predictions of the ER method based on small number of frames are sufficiently biased, however, the confidence interval does not increase much with the decrease of the trajectory size. In turn, the BAR predictions are not biased. However, their confidence interval increases more significantly with the decrease of the trajectory compared to the ER method calculations.
3. For both methods the confidence interval has a tendency to decrease with the increase of the number of trajectory configurations. However, after some threshold value (∼2.5−5 × 10^3 per λ-window simulation) the confidence interval does not change for both methods. This illustrates that the configurations were sampled too frequently, such that the subsequent data points were not statistically independent. This shows that we can safely increase the sampling interval of trajectory configurations by ∼1 order without losing statistical accuracy. Overall, the confidence intervals of the ER predictions are smaller compared to the BAR predictions (see Figure S3 of the SI).

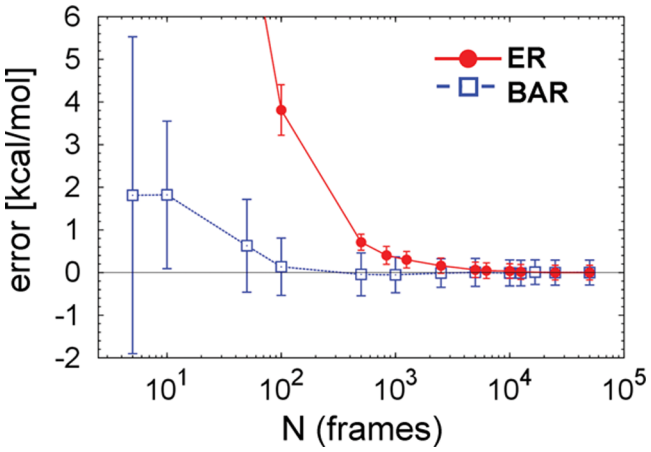


Figure 4. Deviation from the reference value of the SFE calculated with the different number of MD trajectory configurations per λ-stage. The reference values are the SFE calculated with the largest number of configurations. For each method there is its own reference value. The 99% confidence intervals are shown. The distributions at full solute coupling $\rho_{sw}^{e}(\epsilon)$ were constructed from trajectories of different length. The trajectories with the reduced number of configurations were obtained by collecting configurations with a certain interval from the initial trajectory. The distribution functions at zero solute coupling ($\rho_{sw,0}^{e}(\epsilon)$ and $\chi_{sww,0}^{e}(\epsilon, \epsilon')$) were calculated from the initial pure solvent trajectory and 1000 solute insertions for each configuration. The data is shown for the case of butyric acid in sc-CO_2 with ethanol. More data is shown in the Figure S4 of the SI.

These data show that both methods require similar simulation length per λ-stage to achieve comparable statistical errors. Also, one has to carefully validate the ER calculations because a lack of statistics can result into considerable bias of the predictions far beyond the confidence interval, which can be monitored only by increasing the simulation length. The origin of the bias is not clear at the moment. This should be investigated in the forthcoming studies.

One should note that in the cases of poor overlap between accessible phase spaces of neighboring stage simulations the free energy estimates of BAR method also can be biased. The bias depends on the degree of the phase space overlap and on the number of statistical data[48] and in practice can be minimized by choosing sufficient number of intermediate stages. Therefore, absence of the bias in Figure 4 does not guaranty that it would not reveal itself if we change the number of stages. However, in the present study we used a relatively high number of intermediate stages and thus assume that the bias of the BAR calculations is negligible.

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

One should note that formally we used 1000 times more sampling for the system with $\lambda = 0$ in the case of ER method compared to the BAR calculations. However, in reality the length of the pure solvent simulation was the same. Indeed, sampling of 2D correlation functions (see Figure 1) requires more data compared to 1D solute−solvent density distributions in energy representation. The question is how computationally demanding is to calculate the 2D distributions with a sufficient accuracy? We investigated how many data points of $\lambda = 0$ system are required to converge the SFE calculations with ER method (see Figure S5 of the SI). One can see that for the case of butyric acid in sc-CO_2 with ethanol it is required to perform ∼100 solute insertions to each pure solvent trajectory configuration to get the same prediction as with more statistics.

Timing. Calculation of the distribution and the density−density correlation functions with 100 solute insertions per solvent configuration takes ∼4 min on 12 CPUs. The calculation of distribution functions at $\lambda = 1$ takes less than 1 min on 12 CPUs. For comparison an MD simulation to collect the trajectory configurations requires ∼10 h per λ-stage on 12 CPUs on the same machine. Therefore, we may conclude that computational consumptions required for calculation of the functions in energy representations can be neglected. Therefore, in the present work the ER method reduces computational consumptions by 12 times compared to the BAR calculations. However, one should note that we did not perform any optimization of the number of stages for the BAR calculations in the present study. Probably with an optimal choice of the stages less computational resources will be required to achieve similar statistical uncertainty. However, this is out of the scope of the present study.

Comparison with Experiment. In this section we asses *modeling* accuracy of the ER calculations. In the previous section we showed that the theoretical accuracy of ER method calculations is very high. Therefore, the discrepancy between experimental and simulation data is largely determined by the accuracy of potential models (force field) and differences in the systems studied in simulations and experiment.

The modeling accuracy was discussed in detail in our previous publication.[23] In this section we summarize the main observations deduced in ref 23. The quantity which we compare with experimental data is the cosolvent-induced solubility enhancement factor (CISE), which is defined as a ratio between the equilibrium solubilities of solute in sc-CO_2 with cosolvent and in pure sc-CO_2. CISE can be calculated based on SFE data:

$$\mathrm{CISE} = \exp(-\beta[\Delta G_{\mathrm{solv}}^{\text{binary solvent}} - \Delta G_{\mathrm{solv}}^{\text{pure sc-CO}_2}]) \qquad (18)$$

where $\beta = (k_B T)^{-1}$, T is temperature, and k_B is the Boltzmann constant.

Since we consider only monofunctional organic solutes, we may deduce how different functional groups in the solute structure affect CISE. Comparison between calculated and experimental data for different organic molecule classes is presented in Figure 5. One should note that the experimental data comprises data for monofuctional solute molecules of the same chemical classes as the molecules studied here but with bigger molecular weight.

Overall, one can see that the present calculations reproduce the trend observed in the experimental data. All three classes of cosolvents have little effect on solubility of arenes and chloroarenes. In the case of carboxylic acids, simulations correctly show that the CISE decreases in the following

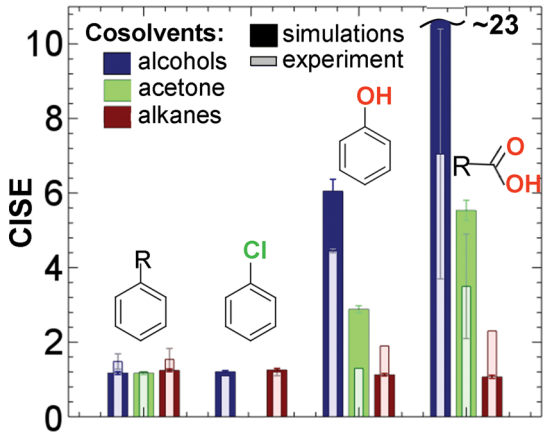


Figure 5. Calculated by ER method (thin light bars) and experimental (wide dark bars) data of the cosolvent-induced solubility enhancement factors (CISE) for different solute classes and three types of cosolvents: alcohols (blue), acetone (green), and alkanes (brown color). Comparison for solutes of four chemical classes is given: alkylarenes, chloroarenes, phenols, and carboxylic acids. Experimental data is summarized from Table 3 of ref 23. The data for stearic acid was excluded from the analysis, because of its high molecular weight. The bars represent the mean values for different molecules of the same chemical class, and the error bars represent the 66% confidence intervals. For experimental data it was calculated as standard deviation of the data within the given chemical class. Absence of the errorbars on experimental data indicates that only a single data point was available. The experimental data was measured at similar state parameters as those used in the present study.

sequence of cosolvents: alcohols, acetone, and alkanes. However, there is a severe overestimation of the CISE for carboxylic acids in the case of alcohols as cosolvents. This can be explained by the fact that a large fraction of carboxylic acid molecules form dimers in sc-CO_2 and the calculations at infinite dillution performed in the present study are not valid (see more discussion in ref 23). We note that the effect of solute association on CISE can be taken into account if we know association free energies (dimerization in the simplest case), which in turn can be calculated from dimerization free energy profiles,[49] obtained, e.g., by umbrella sampling simulations.[50] In contrast to our data, in the case of phenols the experimental data shows almost no effect of acetone on their solubility in sc-CO_2. This result is counterintuitive since the phenol is able to form hydrogen bonds with acetone's carbonyl group. The lack of the statistical uncertainty of experimental data does not allow to make a decisive conclusion on the source of discrepancy.

CONCLUSIONS

The goal of the present work was to reveal the theoretical accuracy of the theory of solutions in energy representation for calculation of solvation free energies in supercritical fluids. We have performed extensive calculations of solvation free energies of organic molecules of different chemical natures in pure supercritical CO_2 and in supercritical CO_2 with addition of 6 mol % of ethanol, acetone, and *n*-hexane as cosolvents. The conclusions are as follows:

1. The theory of solutions in energy representations is a powerful method for solvation free energy calculations in supercritical CO_2. It reproduces the SFE values predicted by the Bennett's acceptance ratio method (the method free of *theoretical error*) with a very low mean absolute

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

deviation of 0.05 kcal/mol. However, ER method required by an order less computational resources compared to the BAR calculations.

2. We showed that the ER method significantly overestimates (by ∼0.6−0.8 kcal/mol) the free energy of cavity formation in the studied solvents. This suggests that the ER method calculations for total solvation free energies can benefit from the theoretical error cancellation. However, the "empirical" compilation of the current free energy functional from different expressions associated with different approximations makes it complicated to draw a definite conclusion.
3. We showed that the little theoretical errors of ER method, which persistently appear for butyric acid and methylcyclohexane across different solvents, have different origins. In the case of butyric acid the error arises because of the lack of the theoretical accuracy in calculation of electrostatic contribution to SFE. For the employed models this contribution apart from the electrostatic multipole−multipole interactions (e.g., dipole−dipole, etc.) also includes the hydrogen bonding interactions. In the case of methylcyclohexane the error seems to appear due to inaccuracies in calculation of the excluded volume contribution to SFE.
4. We showed that the lack of sampling can result into a considerable bias in the ER calculations. However, the statistical uncertainty is less sensitive to the sampling compared to the bias. For the BAR calculations, the observations are opposite. This highlights the fact that the small confidence interval of the ER calculations does not guaranty accurate predictions. The bias has to be controlled as well.
5. Overall, the simulation results reproduce trends in the cosolvent-induced solubility enhancement factors observed in experimental data. Thus, molecular simulations coupled with the ER method can be used for in-silico prediction of the cosolvent effect on solubility of organic molecules in sc-CO_2.

Perspectives. Calculation of SFE in supercritical solvents has important practical applications. Relative solubilities are determined by the differences between SFE.[1,23] This can be used to rationally select cosolvents for a given solute in supercritical media[23] or to estimate the effect of variation of temperature, pressure, and cosolvent concentration(s) on the solubility of bioactive compounds in supercritical fluids. This in turn should reduce the amount of experimental work on optimizing solvency of supercritical solvents.

The present study showed that ER method is an efficient and accurate method for these purposes. There are several suggestions for future research. First, does the success of the ER methods extend to drug-like solutes? The theoretical accuracy of ER method should be studied for bigger molecules with more functional groups. Indeed, drug-like molecules contain many functional groups in their structure and there is no guaranties that ER method will be as successful in this case. Second, how sensitive is the theoretical error for variation of temperature and pressure of the supercritical solvent? We studied only one (but relevant for applications) point on the $P-T$ phase diagram. To answer the above stated question, other supercritical fluids and other regions on the phase diagram should be investigated. Third, will the observed error cancellation hold in the case of purely HNC-like or PY-like approximations? The use of the hybrid functional complicates the analysis of the origins of the theoretical errors in the ER method. However, to the best of our knowledge performance of different functions of ER methods were not reported so far.

ER method proved to be useful for many tasks of computational chemistry, e.g. study of partitioning of molecules to micelles,[51] study of solvent effects on chemical reactions,[52−55] study of protein solvation,[56] calculation of solute-polymer affinities,[57,58] etc. However, as can be seen from publications only the group of the developers and collaborators use this method as a research tool in the everyday practice. We think that one of the reasons why the ER method has not gained wide recognition which it deserves is that there are only few works which quantitatively benchmark the ER calculations against predictions of approximation-free methods or experimental data.[25,59] The present study is an attempt to fill this gap and should encourage more investigations in this direction.

■ ASSOCIATED CONTENT

Supporting Information

Chemical structures of solute molecules. Dipole moments of solutes and number of hydrogen bonds between solute and cosolvents. SFE for sc-CO_2 with addition of acetone and hexane. Data ΔG_{solv}^{vdW} for all solutes and two solvents: pure sc-CO_2 and sc-CO_2 with ethanol. Comparison of the distribution functions in energy representation for different solute−solvent potentials. Details on the derivation of the ER method. This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: andrey.i.frolov@mail.ru.

Present Address

[†]SANOFI R&D, Lead Generation to Candidate Realization, Structure Design & Informatics, 13 quai Jules Guesde, 94403 Vitry-sur-Seine, France.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.I.F. is thankful to Prof. Michael Kiselev for useful discussions, to Prof. David Mobley who kindly provided the data of ΔG_{hydr}^{WCA} for methylcyclohexane, and Prof. Michael Shirts for useful discussion regarding the implementation of the WCA potential in the Gromacs code. A.I.F. is thankful for the support of the Russian Foundation for Basic Research (Grant No. 12-03-31354). The author acknowledges the REA-Research Executive Agency, Grant No. 247500 "BioSol", Programme FP7-PEOPLE-2009-IRSES.

■ REFERENCES

(1) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press, 2006.

(2) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101−122.

(3) Guthrie, J. P. A. Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.

(4) Sander, R. Compilation of Henry's law constants, version 3.99. *Atmos. Chem. Phys. Discuss.* **2014**, *14*, 29615−30521.

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

(5) Reinisch, J.; Klamt, A. Prediction of free energies of hydration with COSMO-RS on the SAMPL4 data set. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 169−173.
(6) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. On the Performance of Continuum Solvation Methods. A Comment on Universal Approaches to Solvation Modeling. *Acc. Chem. Res.* **2009**, *42*, 489−492.
(7) Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. An Accurate Prediction of Hydration Free Energies by Combination of Molecular Integral Equations Theory with Structural Descriptors. *J. Phys. Chem. B* **2010**, *114*, 12068−12079.
(8) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J. Phys.: Condens. Matter* **2010**, *22*, 492101.
(9) Frolov, A. I.; Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Hydration Thermodynamics Using the Reference Interaction Site Model: Speed or Accuracy? *J. Phys. Chem. B* **2011**, *115*, 6011−6022.
(10) Ratkova, E. L.; Fedorov, M. V. Combination of RISM and Cheminformatics for Efficient Predictions of Hydration Free Energy of Polyfragment Molecules: Application to a Set of Organic Pollutants. *J. Chem. Theory Comput.* **2011**, *7*, 1450−1457.
(11) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Toward a Universal Model To Calculate the Solvation Thermodynamics of Druglike Molecules: The Importance of New Experimental Databases. *Mol. Pharmaceutics* **2011**, *8*, 1423−1429.
(12) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259−279.
(13) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(14) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.
(15) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical prediction of hydration free energies for SAMPL. *J. Comput.-Aided Mol. Des.* **2011**, *26*, 551−562.
(16) Pasquali, I.; Bettini, R.; Giordano, F. Supercritical fluid technologies: An innovative approach for manipulating the solid-state of pharmaceuticals. *Adv. Drug Delivery Rev.* **2008**, *60*, 399−410.
(17) Jung, J.; Perrut, M. Particle design using supercritical fluids: Literature and patent survey. *J. Supercrit. Fluids* **2001**, *20*, 179−219.
(18) Herrero, M.; Mendiola, J. A.; Cifuentes, A.; Ibáñez, E. Supercritical fluid extraction: Recent advances and applications. *J. Chromatogr. A* **2010**, *1217*, 2495−2511.
(19) Dobbs, J. M.; Wong, J. M.; Johnston, K. P. Nonpolar co-solvents for solubility enhancement in supercritical fluid carbon dioxide. *J. Chem. Eng. Data* **1986**, *31*, 303−308.
(20) Oparin, R. D.; Idrissi, A.; Fedorov, M. V.; Kiselev, M. G. Dynamic and Static Characteristics of Drug Dissolution in Supercritical CO2 by Infrared Spectroscopy: Measurements of Acetaminophen Solubility in a Wide Range of State Parameters. *J. Chem. Eng. Data* **2014**, *59*, 3517−3523.
(21) Su, Z.; Maroncelli, M. Simulations of solvation free energies and solubilities in supercritical solvents. *J. Chem. Phys.* **2006**, *124*, 164506−164506−15.
(22) Anderson, K. E.; Siepmann, J. I. Solubility in Supercritical Carbon Dioxide: Importance of the Poynting Correction and Entrainer Effects. *J. Phys. Chem. B* **2008**, *112*, 11374−11380.
(23) Frolov, A. I.; Kiselev, M. G. Prediction of Cosolvent Effect on Solvation Free Energies and Solubilities of Organic Compounds in Supercritical Carbon Dioxide Based on Fully Atomistic Molecular Simulations. *J. Phys. Chem. B* **2014**, *118*, 11769−11780.
(24) Matubayasi, N.; Nakahara, M. Theory of solutions in the energetic representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070−6081.
(25) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. II. Functional for the chemical potential. *J. Chem. Phys.* **2002**, *117*, 3605−3616.
(26) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J. Chem. Phys.* **2003**, *119*, 9686−9702.
(27) Matubayasi, N. Free-energy analysis of solvation with the method of energy representation. *Front. Biosci., Landmark Ed.* **2009**, *14*, 3536−3549.
(28) Dobbs, J. M.; Wong, J. M.; Lahiere, R. J.; Johnston, K. P. Modification of supercritical fluid phase behavior using polar cosolvents. *Ind. Eng. Chem. Res.* **1987**, *26*, 56−65.
(29) Frolov, A. I. Theory of Solutions in Energy Representation in NPT-ensemble: Derivation Details. arXiv:1502.04355. arXiv.org e-Print archive. http://arxiv.org/abs/1502.04355 (accessed February 20, 2015).
(30) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(31) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122*, 134508−134508−13.
(32) Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005**, *122*, 144107.
(33) Matubayasi, N. Lecture notes of Prof. N. Matubayasi on the theory of solutions in energy representation. http://sourceforge.net/projects/ermod/files/ (accessed February 7, 2013).
(34) Hansen, J. P., McDonald, I. R. *Theory of Simple Liquids*, 3rd ed.; Academic Press: London, 1991.
(35) 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.
(36) White, J. A.; Velasco, S. The Ornstein-Zernike equation in the canonical ensemble. *Europhys. Lett.* **2001**, *54*, 475.
(37) Hernando, J. A. Density functional theory in the canonical ensemble: I. General formalism. *J. Phys.: Condens. Matter* **2002**, *14*, 303.
(38) Percus, J. K. Approximation Methods in Classical Statistical Mechanics. *Phys. Rev. Lett.* **1962**, *8*, 462−463.
(39) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(40) van der Spoel, D.; Lindahl, E.; Hess, B. *GROMACS User Manual version 4.6.4*; 2013; available at www.gromacs.org.
(41) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(42) Zhang, L.; Siepmann, J. I. Pressure Dependence of the Vapor-Liquid-Liquid Phase Behavior in Ternary Mixtures Consisting of n-Alkanes, n-Perfluoroalkanes, and Carbon Dioxide. *J. Phys. Chem. B* **2005**, *109*, 2911−2919.
(43) Jorgensen, W. L.; Maxwell, D. S.; 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.
(44) *Maestro*, version 9.3; Schrödinger, LLC: New York, NY, 2012.
(45) Martinez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(46) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(47) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256

(48) Wu, D.; Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J. Chem. Phys.* **2005**, *123*, 054103.
(49) de Jong, D. H.; Schäfer, L. V.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C.; Grubmüller, H. Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 1919−1928.
(50) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(51) Matubayasi, N.; Liang, K. K.; Nakahara, M. Free-energy analysis of solubilization in micelle. *J. Chem. Phys.* **2006**, *124*, 154908.
(52) Takahashi, H.; Matubayasi, N.; Nakahara, M.; Nitta, T. A quantum chemical approach to the free energy calculations in condensed systems: the QM/MM method combined with the theory of energy representation. *J. Chem. Phys.* **2004**, *121*, 3989−3999.
(53) Takahashi, H.; Ohno, H.; Kishi, R.; Nakano, M.; Matubayasi, N. Computation of the reduction free energy of coenzyme in aqueous solution by the QM/MM-ER method. *Chem. Phys. Lett.* **2008**, *456*, 176−180.
(54) Takahashi, H.; Ohno, H.; Yamauchi, T.; Kishi, R.; Furukawa, S.-I.; Nakano, M.; Matubayasi, N. Investigation of the dominant hydration structures among the ionic species in aqueous solution: novel quantum mechanics/molecular mechanics simulations combined with the theory of energy representation. *J. Chem. Phys.* **2008**, *128*, 064507.
(55) Matubayasi, N.; Takahashi, H. Free-energy analysis of the electron-density fluctuation in the quantum-mechanical/molecular-mechanical simulation combined with the theory of energy representation. *J. Chem. Phys.* **2012**, *136*, 044505.
(56) Saito, H.; Matubayasi, N.; Nishikawa, K.; Nagao, H. Hydration property of globular proteins: An analysis of solvation free energy by energy representation method. *Chem. Phys. Lett.* **2010**, *497*, 218−222.
(57) Kawakami, T.; Shigemoto, I.; Matubayasi, N. Free-energy analysis of water affinity in polymer studied by atomistic molecular simulation combined with the theory of solutions in the energy representation. *J. Chem. Phys.* **2012**, *137*, 234903.
(58) Kawakami, T.; Shigemoto, I.; Matubayasi, N. Erratum: Free-energy analysis of water affinity in polymer studied by atomistic molecular simulation combined with the theory of solutions in the energy representation. *J. Chem. Phys.* **2014**, *140*, 169903.
(59) Karino, Y.; Fedorov, M. V.; Matubayasi, N. End-point calculation of solvation free energy of amino-acid analogs by molecular theories of solution. *Chem. Phys. Lett.* **2010**, *496*, 351−355.

DOI: 10.1021/acs.jctc.5b00172
J. Chem. Theory Comput. 2015, 11, 2245−2256