



HAL
open science

Hydration free energies and solvation structures with molecular density functional theory in the hypernetted chain approximation

Sohvi Luukkonen, Maximilien Levesque, Luc Belloni, Daniel Borgis

► **To cite this version:**

Sohvi Luukkonen, Maximilien Levesque, Luc Belloni, Daniel Borgis. Hydration free energies and solvation structures with molecular density functional theory in the hypernetted chain approximation. *Journal of Chemical Physics*, 2020, 152 (6), pp.064110. 10.1063/1.5142651 . hal-03094136

HAL Id: hal-03094136

<https://hal.science/hal-03094136>

Submitted on 4 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hydration free energies and solvation structures with molecular density functional theory in the hypernetted chain approximation

Sohvi Luukkonen,¹ Maximilien Levesque,² Luc Belloni,³ and Daniel Borgis^{1,2, a)}

¹⁾ *Maison de la Simulation, USR 3441 CNRS-CEA-Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

²⁾ *PASTEUR, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France*

³⁾ *LIONS, NIMBE, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

(Dated: January 4, 2021)

The capability of molecular density functional theory in its lowest, second-order approximation, equivalent to the hypernetted chain approximation in integral equations, to predict accurately the hydration free-energies and microscopic structure of molecular solutes is explored for a variety of systems: spherical hydrophobic solutes, ions, water as a solute, and the Mobley’s dataset of organic molecules. The successes and the caveats of the approach are carefully pinpointed. Compared to molecular simulations with the same force field and the same fixed solute geometries, the theory describes accurately the solvation of cations, less so that of anions or generally H-bond acceptors. Overall, the electrostatic contribution to solvation free-energies of neutral molecules is correctly reproduced. On the other hand the cavity contribution is poorly described but can be corrected using scaled-particle theory ideas. Addition of a physically-motivated, one-parameter cavity correction accounting for both pressure and surface effects in the nonpolar solvation contribution yields a precision of 0.8 kcal/mol for the overall hydration free energies of the whole Mobley’s dataset. Inclusion of another one-parameter cavity correction for the electrostatics brings it to 0.6 kcal/mol, that is $k_B T$. This is accomplished with a three-orders of magnitude numerical speed-up with respect to molecular simulations.

I. INTRODUCTION

The solvation free energy gives the reversible work associated with the transfer of a solute molecule from ideal gas to a solvent at a certain temperature and pressure. Solvation free energies in general, and especially hydration free energies, are fundamental thermodynamic quantities in themselves. Moreover there is a broad range of interesting physical properties related to solvation free energies such as solubilities, partition coefficients between immiscible solvents, infinite-dilution activity coefficients or binding free energies and potentials of mean force in solutions.

Computing solvation free energies (SFE) or hydration free energies (HFE) is difficult as it requires the sampling of all the possible thermodynamic states that can be visited during the transformation from the initial state (usually, neat solvent) to the final one (solute in solvent). Multiple approaches for solvation free energy calculation have been developed since the dawn of times. These start from ‘simple’ and fast implicit solvent and continuum mean-field approaches where the solvent is treated as a dielectric continuum with a permittivity ϵ , plus a surface area correction for cavitation free-energies. Implicit solvent models involve methods based on the Poisson-Boltzmann equation^{1,2}, the simpler Generalized Born Approximation^{3,4}, or more sophisticated free-energy functionals coupling hydrophobicity, dispersion, and electrostatics⁵. As their name says, the solvent molecules are presented implicitly in these models, not explicitly as individual molecules, therefore they lack any information on the microscopic solvent structure that

can be indeed an important ingredient of the solvation process.

Rigorous and exact simulation-based free-energy methods have been developed to take into account the solvent molecules explicitly. They range from simple exponential averaging (EXP) introduced by Zwanzig⁶ 60 years ago to more sophisticated methods employing non-physical intermediate states such as the thermodynamic integration (TI)⁷, the Bennett acceptance ratio (BAR)⁸, the weighted histogram analysis method (WHAM)⁹ or the multistate Bennett acceptance ratio (MBAR)¹⁰. All these methods require multiple ergodic molecular dynamics (MD) or Monte-Carlo (MC) simulations for a single free energy estimate. In other words, they require multiple simulations to be run, typically few tens, possibly in parallel, with the associated few tens of simulation time. These methods and the protocol associated for producing the free-energy estimate define today’s standard in terms of free energy predictions.

Alternative so called “end-point” explicit methods exist, ie. methods without alchemical intermediates, such as Watermap^{11,12} where the water-density is obtained from an explicit solvent MD simulation and is then injected into a functional that estimates the binding free energy, or the Energy-Representation method^{13,14}, in which an energy-based integral equation theory makes it possible to infer the solvation free energy from the sampling of the solute-solvent interaction energy distribution.

Other approaches accounting explicitly for the molecular nature of the solvent have roots in the liquid state theories. The so-called morphometric approach to solvation elaborates on scaled-particle theory¹⁵ and fundamental measure theory¹⁶ to provide fast estimations of the solvation free-energy of complex molecules^{17–19}. Field theory or classical density functional theory (cDFT) with a simpli-

^{a)}Electronic mail: daniel.borgis@ens.fr

fied dipolar solvent model leads to a local²⁰ or non-local Poisson–Boltzmann–Langevin formalisms^{21,22}, that incorporate solvent and ions sizes. For more realistic molecular models, such as SPC or TIPnP models for water, integral equation or classical density function theory approaches have been developed. A first class of those relies on a reference interaction site model (RISM) that is solved through integral equations, as in 3D-RISM^{23–28} or through a classical DFT (cDFT) formalism using a site-based functional^{29,30}. These methods are having large success and are gaining momentum because of their good balance between precision, simplicity and speed. They have been applied to SFE predictions^{30–37} as well as to a number of structural biology problems^{38–43}. Nevertheless, they rely on site-site correlations only, not on a full molecular description, and hence they require some compromises with phenomenological corrections.

The molecular integral equation theory^{44–46} and molecular density functional theory (MDFT)^{47–50} are the only methods based on liquid state theories that keep the full molecular picture by solving the (angular-dependent) molecular Ornstein-Zernike equation (MOZ)^{51,52}. They are known to be diagrammatically consistent and free of some patches that are necessary in the RISM approach, for example to get the dielectric constant correct. Since Ding et al.⁵³, MDFT can be solved efficiently at the hyper-netted chain (HNC) level of approximation. In a cDFT language, the HNC approximation can be understood as a second-order Taylor expansion of the (unknown) excess functional around a reference homogeneous solvent density ρ_b (We initially qualified it as the homogeneous reference fluid approximation (HRF)^{21,54}). It can thus be considered as exact up to second order in $\Delta\rho = \rho - \rho_b$, and can be used as a sound starting point to be systematically improved. This can be done by developing so-called bridge functionals that bring terms of order $\Delta\rho^3$ and more^{55–61}.

This paper is thus dedicated to MDFT in its “crudest” HNC approximation. Its aim is three-fold. 1) To complement our reference paper, ref⁵³, that was focused mainly on the technical issues, by applications of the MDFT software at its state-of-the-art to relevant solvation issues, namely here its capacity to predict the solvation structure and free-energy of small solutes, going from simple charged and uncharged spherical solutes to a whole dataset of drug-like organic molecules. 2) To assess carefully the accuracy of MDFT at HNC level, acknowledge its successes, and more importantly enlighten where it fails, in order to pinpoint on which aspects the efforts for proper bridge functionals should be put. To this end, the MDFT results will be compared systematically throughout this paper to ‘exact’ results generated by ourselves by Monte-Carlo. Direct comparison of MDFT to experimental results will come in its time 3) To make available a series of results that can be compared point-to-point to those of other ‘simulation-free’ implicit or explicit methods, such as 3D-RISM, site-based DFT, or field theories, so that the different approaches can easily amend themselves and cross-fertilize.

The outline of the paper is as follows. In section II, we recall briefly the MDFT framework and the quantities that are computed. In section III, we apply MDFT to so-

lutes of increasing complexity: first spherical solutes, from hydrophobic spheres to ions, then molecular solutes of arbitrary 3D-shape and charge distribution. In particular, We span the whole Mobley’s database composed of more than 600 molecules for which experimental and simulations solvation free-energies are available. To be fully consistent with the current stage of MDFT, limited so far to solutes with fixed geometries, we have recomputed on our own the reference, simulation HFE values for the whole Mobley’s dataset with a *fixed, rigid geometry* of each molecule. We used for this an original 4D-MC algorithm developed recently by one of us⁶². Conclusions and perspectives are drawn in section IV.

II. MOLECULAR DENSITY FUNCTIONAL THEORY

A. Theory

The molecular density functional theory of classical, molecular fluids computes rigorously and efficiently the solvation free energy and equilibrium solvent density around a solute. MDFT is a cousin of the well-known Kohn-Sham density functional theory for electrons, extended to finite temperature in the grand canonical ensemble by Mermin^{63–65} and further developed for classical fluids by Evans^{66,67}. In the classical density functional theory formalism, the solvation free energy ΔG_{solv} (SFE) is defined as the difference between the grand potential Ω of the solvated system and the grand potential Ω_b of the bulk solvent:

$$\Delta G_{\text{solv}} = \Omega - \Omega_b = \min \{ \mathcal{F}[\rho] \} = \mathcal{F}[\rho_{\text{eq}}], \quad (1)$$

where $\mathcal{F}[\rho]$ is the functional to be minimized, $\rho = \rho(\mathbf{r}, \omega)$ the molecular solvent density function with \mathbf{r} a three dimensional position vector and ω the Euler angles (θ, ϕ, ψ) , characterizing the position and the orientation of the rigid solvent molecule, typically the SPC/E or TIP3P models for water⁶⁸. ρ_{eq} is the equilibrium solvent density

The MDFT functional \mathcal{F} , to be minimized, is made of three parts:

$$\mathcal{F} = \mathcal{F}_{\text{id}} + \mathcal{F}_{\text{ext}} + \mathcal{F}_{\text{exc}}, \quad (2)$$

where \mathcal{F}_{id} is the ideal term of a fluid of non-interacting particles, \mathcal{F}_{ext} is the external term induced by the solute (the molecule, protein, ligand or their complex, ... embedded in water), and \mathcal{F}_{exc} is the excess term that includes structural interaction and correlations between solvent molecules. The ideal term, coming solely from the entropy of mixing of the solvent molecules, reads

$$\mathcal{F}_{\text{id}} = k_{\text{B}}T \int d\mathbf{r}d\omega \left[\rho(\mathbf{r}, \omega) \ln \left(\frac{\rho(\mathbf{r}, \omega)}{\rho_b} \right) - \Delta\rho(\mathbf{r}, \omega) \right], \quad (3)$$

where $k_{\text{B}}T$ is the thermal energy, $d\mathbf{r} \equiv dx dy dz$, $d\omega \equiv d\cos\theta d\phi d\psi$. $\Delta\rho(\mathbf{r}, \omega) = \rho(\mathbf{r}, \omega) - \rho_b$ is the excess density over the bulk homogeneous density $\rho_b = n_b/8\pi^2$.

n_b is typically 0.033 molecule per \AA^3 for water at room conditions (1 kg/l). $8\pi^2$ is the angular normalization constant. The external contribution comes from the interaction potential v_{ext} between the solute molecule and a solvent molecule. It reads

$$\mathcal{F}_{\text{ext}} = \int d\mathbf{r}d\omega \rho(\mathbf{r}, \omega) v_{\text{ext}}(\mathbf{r}, \omega), \quad (4)$$

where v_{ext} is the interaction energy between a solute and a solvent molecule that is made of a van-der-Waals term, typically Lennard-Jones, and electrostatic interactions. Those are the same non-bonded force-field parameters as in a molecular dynamics simulation. For now and in what follows, the MDFT does not use the intramolecular force-field parameters as the solute and the solvent are considered rigid. For the solvent, we will be using SPC/E or the TIP3P water models according to the reference results that are available or generated. TIP3P is the one that is most commonly found in biological applications and which was used in our reference simulations detailed below. The final, excess term describes the effective solvent-solvent interactions. It may be written as a density expansion around the homogeneous bulk density ρ_b :

$$\begin{aligned} \mathcal{F}_{\text{exc}} &= -\frac{k_B T}{2} \int d\mathbf{r}_1 d\omega_1 \int d\mathbf{r}_2 d\omega_2 \Delta\rho(\mathbf{r}_1, \omega_1) \\ &\quad c^{(2)}(\mathbf{r}_{12}, \omega_1, \omega_2) \Delta\rho(\mathbf{r}_2, \omega_2) + \mathcal{F}_B \\ &= -\frac{k_B T}{2} \int d\mathbf{r}_1 d\omega_1 \Delta\rho(\mathbf{r}_1, \omega_1) \gamma(\mathbf{r}_1, \omega_1) + \mathcal{F}_B \\ &= \mathcal{F}_{\text{HNC}} + \mathcal{F}_B, \end{aligned} \quad (5)$$

where $c^{(2)}(\mathbf{r}_{12}, \omega_1, \omega_2)$ is the homogeneous solvent-solvent molecular direct correlation function, \mathcal{F}_B the bridge functional containing all the (unknown) terms of order $\Delta\rho^3$ and beyond. $\gamma \equiv c^{(2)} * \Delta\rho$ is the indirect solute-solvent correlation defined as the spatial and angular convolution of the excess density with $c^{(2)}$. If one cuts the expansion to order two in excess density, that is, if one cancels the bridge functional⁶⁹, one finds that the MDFT functional produces at its variational minimum the well-known HNC relation for the solute-solvent distribution function:

$$\ln(g(\mathbf{r}, \omega)) = -\beta v_{\text{ext}}(\mathbf{r}, \omega) + \gamma(\mathbf{r}, \omega), \quad (6)$$

where $g(\mathbf{r}, \omega) = \rho_{\text{eq}}(\mathbf{r}, \omega)/\rho_b$ and $\beta = 1/k_B T$. For this reason, we call the first term in eq. 5 the HNC excess functional. The function $c^{(2)}$, *i.e.* the second-order direct correlation function of the bulk solvent for a given temperature and pressure, is an input of the present theory and is provided by previous Monte Carlo simulations coupled to integral equations calculations^{70,71}, carefully corrected for finite-size effects⁷², performed for the neat liquid. The rest of the excess term, the so-called bridge functional can be approximated empirically⁵⁵⁻⁵⁸ or rigorously through higher-order direct correlation functions, although this remains a difficult technical challenge. The aim of this paper, however, is to benchmark MDFT at its lowest level of accuracy, the MDFT-HNC, *i.e.* with a vanishing bridge functional,

$\mathcal{F}_B \equiv 0$. This HNC level can only be improved by adding subsequent well-funded, bridge functionals. MDFT-HNC can be considered as a rigorous basis that one can only improve.

This theory and corresponding algorithms are implemented into a in-house high performance code that predicts the solvation free energy of arbitrary solutes in few seconds to minutes, depending on the simulation cell size and spacial and angular resolutions. All technical details are presented in Ref.⁵³. The efficient computation of the excess free-energy term in eq. 5 requires a generalised spherical harmonic expansion up to an order m_{max} . The chosen value for m_{max} fixes the number of discretised orientations to be considered on each spatial grid point. For water, we find in general that $m_{\text{max}} = 3$, corresponding to 84 orientations per grid point, gives sufficient accuracy compared to higher order expansions, *e.g.* $m_{\text{max}} = 5$ corresponding to 330 orientations. Many of the calculations presented below were performed with $m_{\text{max}} = 5$ just for safety; this is completely affordable for the relatively small molecular solutes (up to 40 atoms) that were considered.

B. Solvent equilibrium structure, polarization and beyond

Equation 1 states that at the same time as MDFT produces the solvation free energy of arbitrarily complex molecule (the value of the functional at its minimum), it produces the equilibrium solvent structure around this solute (the density that minimizes the functional) in its full molecular description; the molecular solvent distribution function (reduced density) is given by $g(\mathbf{r}, \omega) = \rho(\mathbf{r}, \omega)/\rho_b$. From this full molecular distribution, one can extract more readable information. For instance, the first moment of $g(\mathbf{r}, \omega)$ is the three dimensional scalar field $g(\mathbf{r}) \equiv \frac{1}{8\pi^2} \int g(\mathbf{r}, \omega) d\omega$ from which one can derive the usual spherically symmetric radial distribution function $g_i(r)$ between solute sites and water oxygens or hydrogens. $n(\mathbf{r}) = n_b g(\mathbf{r})$ is the number density. Another important quantity embedded in $g(\mathbf{r}, \omega)$ is the polarization field $\mathbf{P}(\mathbf{r}) = \frac{\mu n_b}{8\pi^2} \int \hat{\omega} g(\mathbf{r}, \omega) d\omega$ where μ is the solvent molecule dipole value, $\hat{\omega}$ is the unitary vector along the dipole axis depending on (θ, ϕ) only. One can also obtain so called water maps from $g(\mathbf{r}, \omega)$, catching the most probable water molecules position and orientation around the solute.

It should be noted that the equilibrium molecular solvent density $g(\mathbf{r}, \omega)$ is a direct output of MDFT. In the case of molecular simulations, one would have to accumulate such data during a long trajectory, averaging in spatial voxels of typical size 0.25-0.5 \AA for a series of orientations. For just positions, this can be tackled nowadays, especially with the recent approach using an estimator based on forces rather than simple binning to decrease the variance of the estimate of g ⁷³⁻⁷⁵. Nevertheless, accumulating data in the full six-dimensional orientation and position space is a daunting task, even more difficult than computing SFE's. MDFT produces this 6-dimensional map in the same few minutes as it needs to predict the SFE.

III. REFERENCE CALCULATIONS FROM SIMULATION

To assess the quality of the hydration free energies (HFE) and solvation profiles predicted by MDFT-HNC for different systems presented in the next sections, we have calculated rigid-solutes reference data using the same force-field parameters as in MDFT with an original MC/4D-hybrid method and associated code developed by one of us⁶². This code does in parallel a (i) MC simulation of bulk water with N water molecules and (ii) a simulation of the solute in N water molecules. Every i^{th} step the rigid solute is inserted slowly into the simulation box in (i), or removed slowly in (ii) via a fictitious 4th dimension and through a non-equilibrium process. The work ΔH of each insertion/deletion is calculated and the systems are returned their initial state to continue the equilibrium MC simulations. The Jarzynski equation^{76,77} $e^{-\beta\Delta G} = \langle e^{-\beta\Delta H} \rangle$ states the solvation free energy can be obtained from these out-of-equilibrium insertion/deletion process. To improve the statistics the insertion and deletion profiles are combined with the BAR method⁸. One can also obtain the solvents molecular density around the solute from the (ii) simulations. Each hydration free energy calculation takes 60 cpu.h to obtain a well controlled of 0.02-0.03 kcal/mol.

In all our simulations the TIP3P water model was used as solvent. In the case of the FreeSolv database^{78,79} containing small drug-like molecules the force field is GAFF (v1.7)⁸⁰ with AM1-BCC partial charges^{81,82}.

IV. RESULTS FOR SIMPLE SPHERICAL SOLUTES

A. hydrophobic solutes

In Fig. 1 we compare the solvation profile that we obtained either by MC simulations or MDFT for a one-site united-atom representation of methane ($\sigma = 3.73$ Å, $\epsilon = 0.294$ kcal/mol) and of neopentane, $\text{C}(\text{CH}_3)_4$ ($\sigma = 6.15$ Å, $\epsilon = 0.835$ kcal/mol). As already evidenced in the past⁴⁹, the HNC approximation predicts correctly the cavity volume and the rising of the first peak. Its characteristic feature for small hydrophobic solutes is to slightly displace the first peak and overestimate its height. The peak location is better for the larger solute but the height overestimation remains. Overall, however, the approximation is doing fine on the structure. When one goes to solvation free-energies comes a disaster, that is shared with 3D-RISM^{83,84}. For the two preceding example the computed MDFT-HNC free-energies are 6.2 and 18.1 kcal/mol, instead of 2.0 and -0.3 kcal/mol, respectively. The discrepancy worsens with solute size. This deficiency of the HNC approximation is, in part, well understood: for large spheres (or general cavities), the solvation free-energy should tend to PV , where P is the experimental or simulation pressure, thus virtually zero at a normal pressure of 1 atm unless micrometric sizes are reached. Instead the HNC approximation leads to $P_{hnc}V$, with a pressure given by

$$P_{hnc} = k_B T n_b \left(1 - \frac{1}{2} n_b \hat{c}^{000}(0) \right) \quad (7)$$

where $\hat{c}^{000}(0)$ is the $q = 0$ Fourier component of the spherically averaged direct correlation function in eq. 5. P_{hnc} has largely overestimated values: 11260 atm for SPC/E and 9400 atm for TIP3P ! The first patch that we proposed previously was to correct the HNC free energy by a factor $-P_{hnc}V_{pm}$, where V_{pm} is the measured partial molar volume (PMV), i.e. computed as $V_{pm} = -\int_V dr \Delta n(\mathbf{r})/n_b$, where $\Delta n(\mathbf{r}) = n(\mathbf{r}) - n_b$. Correction terms of the type $aV_{pm} + b$ with empirically adjusted constants a and b had been already proposed for 3D-RISM calculations^{84,85}, which we have connected to our pressure correction (PC) formulation^{86,87}.

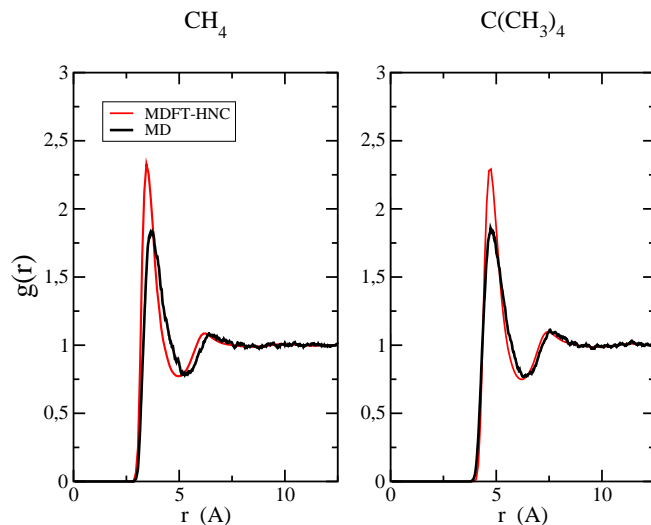


Figure 1. Solvation profiles obtained with MDFT-HNC (red) or MC simulations (black) for two hydrophobic Lennard-Jones solutes.

To go deeper into this problem, we examine here the solvation of a hydrophobic sphere which is the paradigmatic problem for either the standard scaled particle theory^{15,88} or more recent advances in the theory of hydrophobicity and hydrophobic interactions^{89,90}. We already showed in previous works that the HNC approximation suffers from the original sin of being a quadratic theory around the liquid bulk density. The homogeneous free energy as a function of density presents a single instead of a double well; it cannot accommodate for the liquid-gas transition. In particular the density profile around the hydrophobic sphere tends to a constant finite value for large radii instead of turning to depletion^{57,58}, as expected and as simulations show^{90,91}. We focus here on the solvation free-energy and rely on the simulation results of Hummer *et al.*⁹² and Huang and Chandler⁸⁹ generated for either the SPC or SPC/E models. In Fig. 2-a, we plot the hydration free energy of a hard sphere of increasing radius R computed by MDFT and compared to the simulation results. We also plot the analytical limit for cavity volumes that can only accommodate 0 or 1 water molecule, namely⁹²

$$\Delta G = -k_B T \ln(1 - n_b V), \quad (8)$$

where $V = 4\pi R^3/3$ is the hard-sphere volume. It is seen

that MDFT-HNC and simulation fulfil this exact small-radii limit up for $R < 2 \text{ \AA}$; MDFT-HNC even matches the simulation results slightly beyond that radius and diverges from them afterwards. In Fig. 2, we compare the solva-

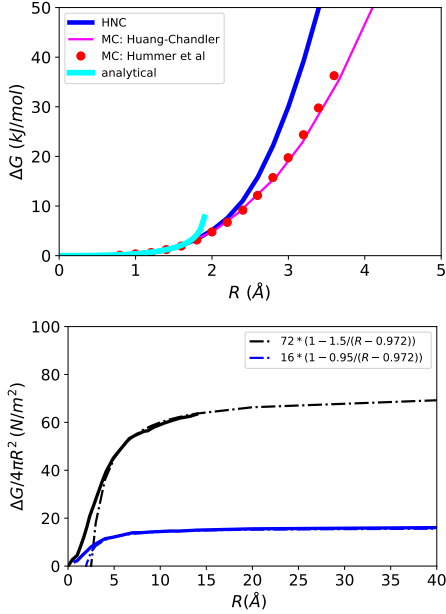


Figure 2. Top: Solvation free energy of a hard sphere of radius R obtained by MDFT-HNC (blue line) or by Monte-Carlo simulations by Hummer *et al* for the SPC model⁹², or Huang and Chandler for SPC/E⁸⁹ (red bullets and red line, respectively). The line in cyan is the analytical result of eq. 8. Bottom: Hydration free energy per unit area, $\Delta G/4\pi R^2$, as function of HS radius computed by MC or by MDFT after pressure correction.

tion free energy per unit area, computed either by MC by Huang-Chandler as $\Delta G(R)/4\pi R^2$, or by MDFT after pressure correction as $(\Delta G(R) - P_{hnc}V_{pm})/4\pi R^2$. Both curves present an horizontal asymptote pointing to the surface tension γ . Simulations yields $\gamma_{sim} = 72 \text{ N/m}^2$, a value close to the experimental one but somewhat larger than the reported gas-liquid surface tension of SPC/E⁹³, whereas MDFT-HNC yields the much smaller values $\gamma_{hnc} = 16 \text{ N/m}^2$: thus not only the HNC pressure has to be corrected but also the surface tension.

If one refers to scaled particle theory, an important question is the definition of the solute volume and surface to be considered, usually derived from either the solute Van-der-Waals surface (VdW), or the solvent-accessible surface (SAS); the two of them differ roughly by the extension of a water molecule radius R_w . Here the question is rather the relationship between the measured partial molar volume V_{pm} or the hard-sphere volume V . In Fig. 3-a we clarify that relationship; we observe that V_{pm} is optimally fitted by $V_{pm} = 4\pi R^{*3}/3$ with a shifted radius $R^* = R - R_w$ and $R_w = 0.972 \text{ \AA}$. R^* can be identified to the Van-der-Waals radius generating a VdW surface of area $S = 4\pi R^{*2}$ rather than the solvent accessible surface of area $S = 4\pi R^2$. Supposing a (truncated) scaled-particle theory expression for the solvation free-energy with the same R_w for both simu-

lation and MDFT-HNC⁸⁸

$$\Delta G = P \frac{4\pi}{3} R^{*3} + \gamma 4\pi R^{*2} \left(1 - \frac{\delta}{R^*}\right) \quad (9)$$

with $R^* = (3V_{pm}/4\pi)^{1/3}$, and accounting for the fact that $P_{sim} \simeq 0$, one can write a correction to the HNC approximation as

$$\Delta G_{hnc} - \Delta G_{sim} = a(R^*) P_{hnc} V_{pm} \quad (10)$$

with

$$a(R^*) = 1 + \frac{3\Delta\gamma}{P_{hnc}R^*} \left(1 - \frac{d}{R^*}\right) \quad (11)$$

and $\Delta\gamma = \gamma_{hnc} - \gamma_{sim} = -56 \text{ N/m}^2$. The first term in $a(R^*)$ yields the pure pressure correction and the next one a surface correction to it; the length parameter d relates to the so-called curvature correction to the surface tension –here in fact the surface tension difference. It can be determined by imposing the condition that $\Delta G_{hnc} = \Delta G_{sim}$ for small radii, *e.g.* for $R^* = 1 \text{ \AA}$ ($R \simeq 2 \text{ \AA}$; see Fig. 2-a). This condition yields $d = 1 + P_{hnc}/3\Delta\gamma = 0.32 \text{ \AA}$. The approximation of eq. 11 is compared to the simulation results in Fig. 3 and fits quite well. Note that it is a parameter-free expression and only $\Delta\gamma$ enters. The simpler approximation with $d = 0$ applies only above $\sim 500 \text{ \AA}^3$.

The conclusion of this section is that a pressure correction proportional to the PMV as we proposed previously is strictly valid for very large solutes of micro-metric size. For microscopic to nanoscale solutes, at least a surface correction $\Delta\gamma$, preferably the next correction term in a scaled-particle theory parametrisation, should be accounted for. On this simple paradigmatic example, one observes that there is no way that a simple correction strictly proportional to the PMV can be applied, unless limited to a small range of PMV. This remark will be extremely useful in Sec. V, when we will discuss the hydration free energy of molecular solutes whose PMV range typically between 50 and 500 \AA^3 .

B. Ions

Here we study the solvation of simple monovalent ions : 4 anions, F^- , Cl^- , Br^- , I^- , and 4 cations, Li^+ , Na^+ , K^+ , Cs^+ , described with force field parameters given by Horinek and Netz⁹⁴. Those are recapitulated in Table I. The MDFT-HNC results, that is, the output of the minimization of the functional given in eq. 1, were obtained within a cubic supercell of side 32 \AA , with periodic boundary conditions, a spacial resolution of 0.25 \AA ($= 128$ grid nodes in each direction) and an angular resolution of 330 orientations per spatial grid node.

1. Solvation profiles

A primary output of MDFT is the full molecular equilibrium solvent structure $g(\mathbf{r}, \omega)$ from which one can derive

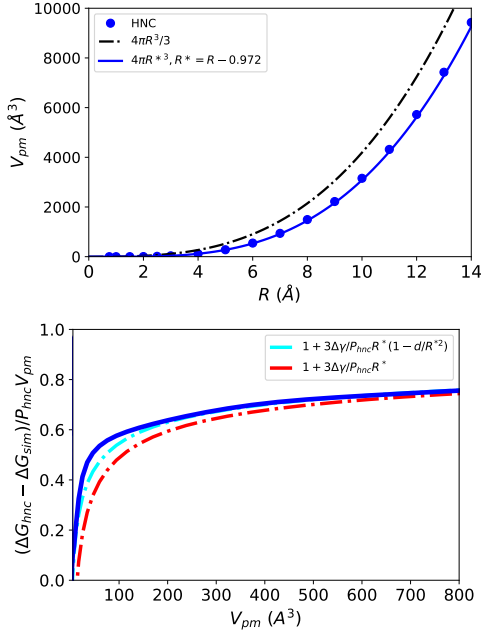


Figure 3. Top: Partial molar volume V_{pm} versus hard-sphere radius R obtained by MDFT-HNC. It is best fitted by $V_{pm} = 4\pi R^{*3}/3$ with $R^* = R - 0.972$. Bottom: Correction to the HNC solvation free energy normalized by the first order pressure correction (thus the quantity $a(R^*)$ defined in eqn. 10) as a function of the partial molar volume: exact computation (solid blue line), or estimation using eq. 11, with the value of d given in the text (cyan dash-dots) or with $d = 0$ (red dash-dots).

easy radial $g(r)$ and polarization $P(r)$ distribution functions or any other angular-dependent density distribution. Fig. 4 presents the radial density and polarization distributions for all the ions. Concerning the $g(r)$'s, MDFT-HNC clearly performs better for the cations than anions. For the cations MDFT correctly predicts the position of the first two maxima and first minimum. For the smallest cation, Li^+ , MDFT slightly overestimates the intensities of the maxima and for the larger cations MDFT slightly underestimates the relative intensities of the maxima and the minimum. This effect increases with the cation size. In the case of the anions the $g(r)$ predicted with MDFT deviates much more from the simulation results. The position of the first peak and its width are correct. As for the second peak, it is displaced to larger distances. Since the position of the second peak in water is a sign of tetrahedral order, the cation here taking the place of one water molecule, we concluded before that the HNC approximation is missing here some tetrahedral order. For the polarization radial distributions $P(r)$ the correspondence between MD and MDFT is much better both for cations and anions, with some differences in the intensities of the minima and maxima, but globally an excellent agreement.

Beyond the traditional computation of the atomic pair distribution functions, MDFT has the great advantage of providing in addition a complete information on the orientations of the water molecules around the solute. Here, in spherical symmetry, this translates to the knowledge of the angular-dependent density maps $g(r, \cos \theta', \psi')$, where

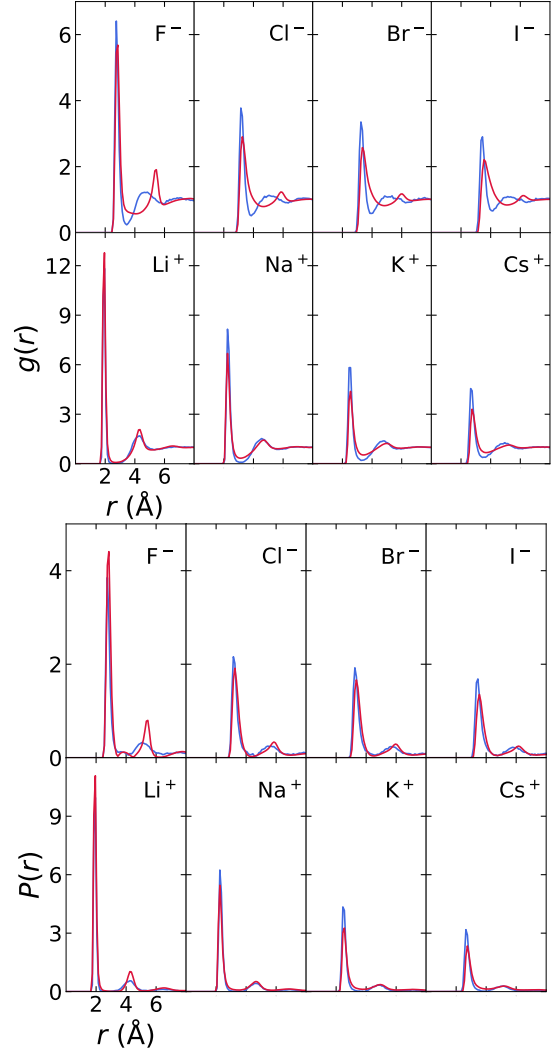


Figure 4. Radial density (top) and polarization density (bottom) for the ions obtained with simulations (blue) and MDFT-HNC (red).

θ' is the angle between the dipole direction of one water molecule at position \mathbf{r} from the ion and \mathbf{r} itself, ψ' is the rotation angle around the dipole direction. $g(r, \cos \theta', \psi')$ is easily deduced from the full distribution in laboratory frame, $g(\mathbf{r}, \omega)$, by spherical average over all \mathbf{r} -orientations. In Fig. 5, we have concatenated all this information into the 2D-plots of (i) $g(r, \cos \theta') = \frac{1}{2\pi} \int d\psi' g(r, \cos \theta', \psi')$, indicating the preferred orientation of the solvent dipoles as function of the radial distance, and (ii) $g(\psi', \cos \theta') \equiv g(r_{max}, \cos \theta', \psi')$ where r_{max} is the distance corresponding to the maximum of the radial density distribution $g(r)$; this indicates the preferred orientation of the hydrogens as a function of the dipole orientation. The plots are given for both Cl^- and Na^+ and compare MDFT to simulations. As can be seen, the agreement is again excellent for the cation. At the peak of $g(r)$ the water dipole is directed radially away from the cation ($\cos \theta' = 1$), *i.e.* with the oxygen closest to the cation and the hydrogens pointing away symmetrically, with no angular dependency in ψ' close to $\cos \theta' = 1$: the hydrogens rotate freely around the dipole axis. For val-

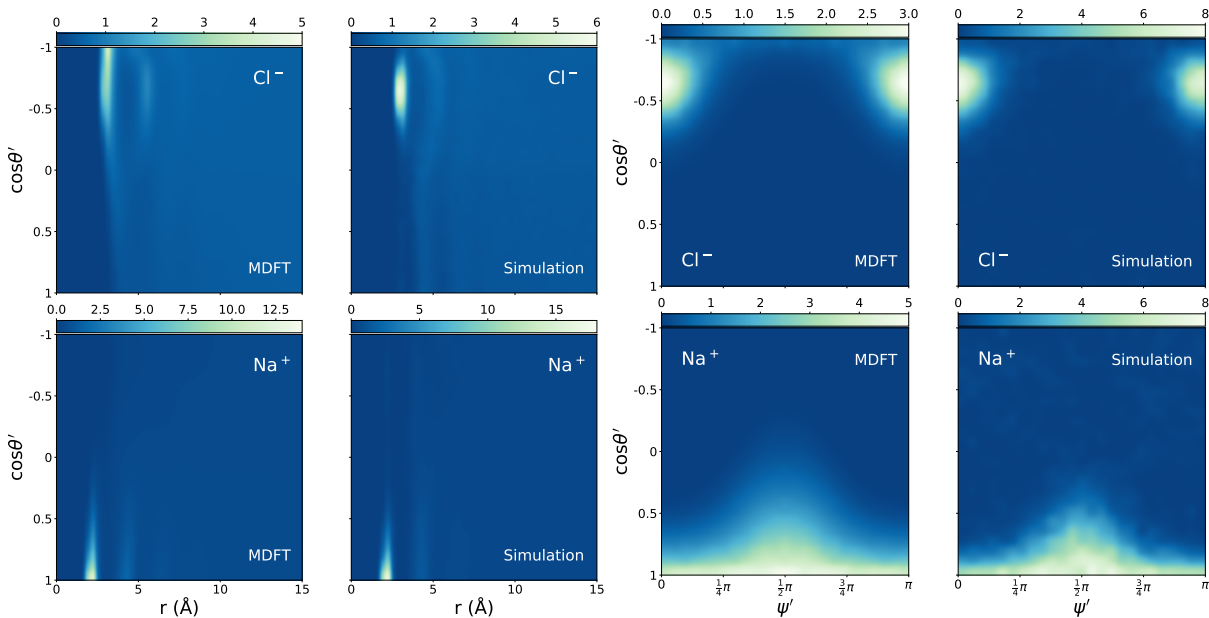


Figure 5. Two-dimensional maps of $g(r, \cos \theta')$ on the left and $g(\psi', \cos \theta')$ on the right (see definitions in the text), computed by simulations or by MDFT for Cl^- (Top) or Na^+ (Bottom).

ues of $\cos \theta'$ departing from 1, a distribution in ψ' appears around $\psi' = \frac{\pi}{2}$. In our conventions, $\psi' = 0$ or π corresponds to the water molecule in the plane formed by the dipole direction and ion-oxygen direction. The value $\pi/2$ corresponds to a configuration in which the two hydrogens become equidistant from the cation, thus maximising the sum of the two distances. As for the anion, the $g(\cos \theta', \psi')$ maps look quite similar for MDFT and MC and display a peak centred around $\cos \theta' \simeq -0.58 = \cos(\pi - \theta_0/2)$, where θ_0 is the HOH angle of the TIP3P model, and $\psi' = 0$ or π in order to have the optimal H-bond to the anion. So far so good but as seen before it cannot be perfect: a difference does appear in the $g(r, \cos \theta')$ map in which the peak in simulation appears consistently at the $\cos \theta' = -0.58$ as before and extends roughly between -0.4 and -0.8 , whereas in MDFT it extends more floppily from -0.4 and -1 , with its maximum at -1 . The strength and directionality of the O-H- X^- bond is clearly underestimated. The second peak is displaced and is somewhat more narrow in angle and more pronounced.

2. Hydration free energies

Table I reports the solvation free energies obtained with the reference MC simulations and with MDFT-HNC. In both cases, since the calculation are done for a periodic system, two type of correction have been applied, of so-called B and C type^{53,95,96}. The first one in q^2/L is a finite size, Madelung-like correction incorporating the contribution of all the periodic images. The second one proportional to q refers to the so-called Galvani potential in periodic boundary conditions.

Fig. 6 shows the correlation between MDFT and simulation for the absolute solvation free energies ΔG and

the relative solvation free energies as defined in Ref.⁹⁴, $\Delta\Delta G = \Delta G + z\Delta G(\text{Cl}^-)$, with $z = \pm 1$ according to the ion valence and $\Delta G(\text{Cl}^-)$ the value obtained for Cl^- . The latter were the reference free energies that Horinek's and Netz's used to fit their ions force-field parameters, since it cancels the somehow uncontrolled surface charge corrections that should be added when comparing to experimental values.

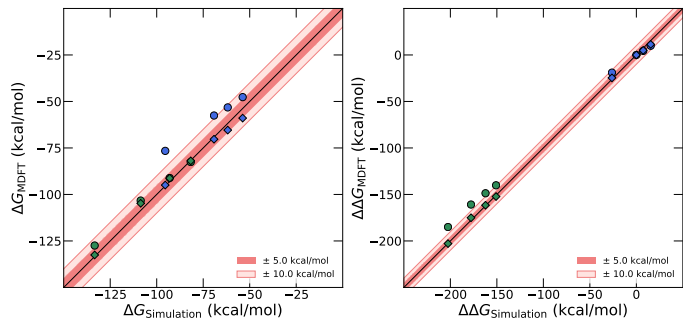


Figure 6. Correlation between **a.** absolute hydration free energies ΔG and **b.** relative hydration free energies $\Delta\Delta G$ defined in the text, obtained with MC simulations or MDFT, for the halide and alkali ions series listed in Table I. The circles correspond to the bare HNC results, whereas the diamonds include the two-parameter correction derived independently in Sec. V B 2.

Not surprisingly, the bare HNC results are quite good for the cations, with a maximum relative error for 3 – 4 % for the smallest ions in the series, i.e. Li^+ and Na^+ , but it is not so for the anions, with an maximum error of ~ 20 % for, again, the smallest ones. This corroborates the caveats in structure that were discussed above. If any effort should be put in finding the appropriate angular dependent bridge functional, it should focus on solute negative charges, giv-

	$\sigma(A)$	ε	$\Delta G_{\text{Simulation}}$	ΔG_{MDFT}	$\Delta G_{\text{MDFT}}^{\text{cor}}$	$\Delta\Delta G_{\text{Simulation}}$	$\Delta\Delta G_{\text{MDFT}}$	$\Delta\Delta G_{\text{MDFT}}^{\text{cor}}$
F ⁻	3.434	0.1110	- 95.40 ± 0.03	- 76.7	-95.0	- 27.13	-19.09	-24.6
Cl ⁻	4.394	0.0994	- 68.27 ± 0.03	- 57.6	-70.3	0.0	0.0	0.0
Br ⁻	4.834	0.0503	- 61.79 ± 0.03	- 53.3	-65.4	6.48	4.32	4.9
I ⁻	5.334	0.0763	- 53.78 ± 0.03	- 47.7	-58.9	14.49	9.86	11.4
Li ⁺	2.874	1.47.10 ⁻⁴	- 133.38 ± 0.03	- 128.5	-132.5	- 201.65	- 186.08	-202.9
Na ⁺	3.874	<i>id.</i>	- 108.63 ± 0.03	- 104.1	- 104.8	- 176.89	- 161.71	-175.1
K ⁺	4.534	<i>id.</i>	- 93.04 ± 0.02	- 91.9	-91.3	- 161.31	- 149.47	-161.6
Cs ⁺	5.174	<i>id.</i>	- 81.71 ± 0.02	- 83.6	-81.9	- 149.99	- 141.15	-152.2

Table I. Force-field parameters and absolute and relative solvation free energies (all energies in kcal/mol) for a series of halide and alkali ions, obtained with reference simulations or with MDFT, either in the bare HNC approximation or including the two-parameter correction derived independently in Sec. VB 2. The results correspond to those displayed in Fig. 6.

ing rise to H-bond acceptor sites. For the time being, we show both in Table I and Fig. 6 that an a-posteriori two-parameter correction developed independently in Sec. VB 2 for neutral molecules considerably improves the HNC results for the charged solutes too.

V. MOLECULAR SOLUTES

A. Water as solute

We begin by the case of a TIP3P water molecule in TIP3P, a paradigm for both a H-bond acceptor and donor. The MDFT-HNC result, that is, the output of the minimization of the functional given in eq. 1, was obtained within a cubic supercell of side 24 Å, with periodic boundary conditions, a spacial resolution of 0.25 Å (= 96 grid nodes in each direction) and $m_{max} = 5$ (330 orientations per spatial grid node).

Fig. 7a. shows the radial site-site pair distribution functions between the solute oxygen and hydrogens sites and the solvent oxygen and hydrogen sites obtained from the MC simulation and MDFT-HNC. Here we recover the equivalent results obtained already 20 years ago by Richardi *et al.*⁴⁴ and Lombardero *et al.*⁴⁵ using 1D-MOZ-HNC integral equations for both TIP3P and SPC/E water. Indeed the same deficiencies of HNC appear: it does miss some of the (subtle) tetrahedral symmetry in water. The first O-O peak is correctly placed but too wide on its right side; the second peak is misplaced and appear at a position pertinent to the second neighbour in a general dipolar fluid, and not at the 4.6 Å value imposed by the tetrahedral symmetry. The first O-H or H-O peak is also at correct position but underestimated. The H-H pair distribution function appears almost structureless in MDFT-HNC. One can extract also the three-dimensional solvent charge densities, easily with MDFT, more painfully by simulation since one needs to explore three-dimensional space with sufficient statistics. This is illustrated in Fig. 7-b and c. There is represented iso-surfaces of the 3D-polarization density $P_y(\mathbf{r})$, where y is the axis perpendicular to the molecular plane. Panel b shows the isosurfaces $P_y(\mathbf{r}) = \pm 0.035$ Debye/Å³ obtained by MDFT with a grid size of 0.25 Å, whereas Panel c shows the same quantity obtained by col-

lecting histograms of identical voxel size along a 50-ns-long MD trajectory (25000 independent configurations). These 3D plots look familiar compared to previous simulations⁷⁵ with a change of sign when crossing the symmetry plane. The two rather loose, upper caps correspond to the solvent donor molecules presenting their hydrogens to the solute oxygen negative partial charge. The two lowest ones represent the solvent water molecule presenting its oxygen to the hydrogen site pointing in the figure, and whose orientation can depart from the average, symmetric one with the two hydrogens pointing away and a vanishing P_y . Beyond the satisfactory agreement between MDFT and simulation, the noise appearing in the MD results illustrates the statistical difficulty of accumulating 3D-densities by simulation, not to speak of position and orientation densities, which are the direct output of MDFT. Finally we present in Fig. 7-e a feature that is not accessible to RISM-based approaches and would require intense statistics in simulations: we plot the probability of finding a water molecule in a fixed orientation at distance z from another, here the most probable orientation for a O-H-O bond on the positive side, $z > 0$, which becomes a H-bond mismatch on the other side of the donating molecule, for $z > 0$.

As for the hydration energy of an additional water molecule in water (namely the chemical potential of TIP3P water), MDFT-HNC predicts -4.69 kcal/mol when including a simple pressure correction, $-P_{hnc}V_{pm}$ whereas MC gives -6.04 ± 0.07 kcal/mol. The more elaborate correction developed in the next section yields -6.1 kcal/mol.

B. Small organic molecules

In this section, we assess the quality of the solvation structures and HFEs of small organic molecules predicted by MDFT-HNC. For that we use the FreeSolv database^{78,79} that contains 642 small neutral organic drug-like molecules, for which the experimental values have been concatenated and theoretical values have been computed systematically by MD using a free-energy-perturbation method with a standard, flexible force field. As mentioned before, we have regenerated ourselves the reference values for the whole dataset at fixed solute geometries, those given in the database.

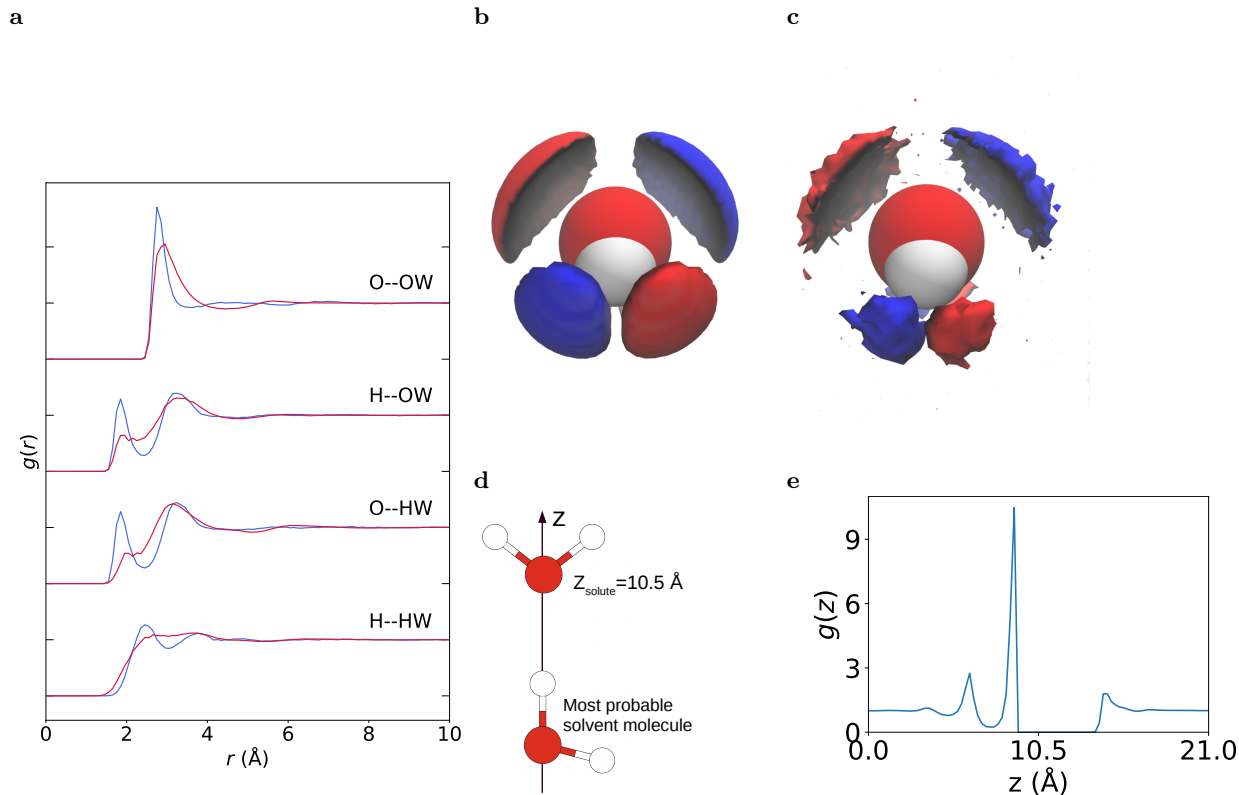


Figure 7. **a.** Site-site radial distribution functions for site-site densities. Simulation and MDFT results in blue and red respectively. **b.** Isosurface for the polarization density $P_y(\mathbf{r})$ (red and blue for positive and negative, respectively) computed by MDFT with a voxel size of 0.25 \AA . **c.** Same computed from a 50 ns long MD simulation with identical voxel size. **d.** Representation of a solute water fixed in the center of the box and a second solvent one at its most probable location ($r_{w-w} = 2.45 \text{ \AA}$) and in its optimal angular configuration for H-bonding. **e.** Probability distribution for a water molecule keeping the same fixed orientation as in **d.** and gliding along the z-axis.

1. Solvation profiles

We begin by showing the capacity of MDFT to predict solvation profiles around a molecule. It is illusory to span the whole database. We have chosen for illustration the case of quinoline (FreeSolv ID: mogley_5857, a typical molecule of the FreeSolv database, whose chemical structure presented in fig. 8c). The 3D-solvation structure obtained by MDFT is represented in Fig. 8. Panel a displays the number density in the plane of the molecule, with the associated alternation of maxima and minima. Panel b concerns another important quantity embedded in $g(\mathbf{r}, \omega)$, that is the polarization field $\mathbf{P}(\mathbf{r})$. There is displayed the norm of the polarization field, i.e., $\|\mathbf{P}(\mathbf{r})\|$, in the plane of the quinoline molecule obtained with MDFT-HNC. As expected we find high polarization close to the sites wearing localized charges, and the expected polarization with OH pointing toward N.

A direct comparison to simulation results is made for the site radial distribution functions in Fig. 8c. The agreement appears very reasonable. For all solute atoms, the rise of the first peak follows exactly that of the simulation: the shape of cavity is perfectly reproduced. The maximum of the first peak, if any, is correctly located, meaning that the first solvation shell lies where it should. For the carbon

sites exposed to the solvent (e.g. C1, C6, C8, C9) one does recover the overestimation of the height that was found in Fig. 1. The $g(r)$ for the nitrogen site misses the important H-bond first peak. Since the nitrogen atom wears a high partial charge of $-0.65e$, one is back to the problem encountered before for strong negative charges, e.g. for anions or the oxygen of water.

At last, from $g(\mathbf{r}, \omega)$ one can also obtain so-called water maps catching the most probable water molecules position and orientation around the solute. Fig. 8d shows the four most probable position and the orientation of water molecules around the quinoline. Expectedly, they are found close to the nitrogen atom.

2. Hydration free energies

Fig. 9 shows different correlations between HFE's of the FreeSolv database, for both the uncharged (pure LJ) and charged molecules, obtained by a MDFT-HNC calculations and reference simulations. The MDFT-HNC results were obtained within a cubic supercell of side 32 \AA , with periodic boundary conditions, a spacial resolution of 0.33 \AA ($= 96$ grid nodes in each direction) and an angular resolution of 84 orientations per spatial grid node ($m_{max} = 3$). The MDFT calculations were performed on the initial configura-

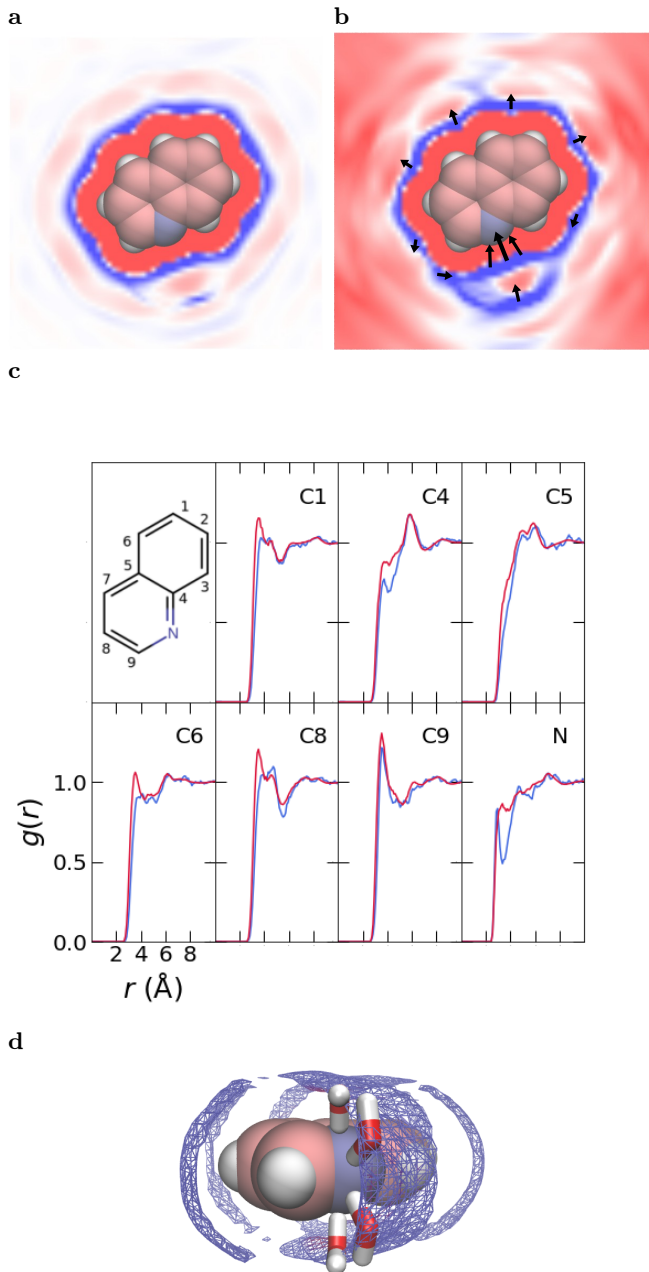


Figure 8. **a.** Water density map (red : $\rho < \rho_b$, white : $\rho = \rho_b$ and blue : $\rho > \rho_b$) and **b.** norm of the polarization vector field (blue : high polarization, black arrows representing the orientation) in the plane of the molecule obtained with MDFT-HNC. **c.** Chemical structure of the quinoline and the radial distribution function $g(r)$ between the heavy atoms of the quinoline and water oxygens (blue lines correspond to MDFT-HNC as obtained in few minutes, and red lines to MD simulations as obtained in few hours). **d.** Representation of the quinoline molecule with four most probable water molecules.

tion given in the FreeSolv database. They did not converge for 22 molecules (3 % of the database). All results presented below are for the 620 molecules that led to convergence. The computational cost, or average computation time per molecule on a single CPU was 8 min –we usually use 8 cpu-threads. Note, we could have done the calculation in a box

of 21 Å, and hence have had only 64 nodes in each direction, for the vast majority of the molecules in the database, decreasing the simulation time to under a minute on average. As we mentioned, the reference MC/4D-hybrid calculations take 60 cpu hours/molecule. We checked that MD simulations with gromacs⁹⁷ and its implemented MBAR free-energy protocol require an equivalent cpu time. This yields a 3 orders of magnitude factor between MDFT and molecular simulations for equal computer resources.

As recognised earlier for both RISM and DFT, and as discussed thoroughly in Sec. IV A, MDFT-HNC fails for what seems the simplest case, i.e. estimating the free-energy cost of creating cavities. This failure extends to non-polar solutes composed of LJ sites with no partial charges. We thus start by discussing that case. In Fig. 9-a, we begin by correlating the MDFT and MC results for the solvation free energies ΔG^{LJ} of the uncharged solutes, when a simple pressure correction is added, $-P_{hnc}V_{pm}$, as we prescribed in previous papers^{86,87}. This correction does improve greatly the bare results which would appear far higher in the figure, but it cannot be considered as satisfactory yet, with a RMSE around 3 kcal/mol and a correlation coefficient around 0.5. Following the discussion of Sec. IV A and Fig. 3, we have plotted in Fig. 10 the "exact" correction factor of eq. 11, $(\Delta G_{hnc} - \Delta G_{sim})/P_{hnc}V_{pm}$ and we have fitted it with the analytical form of eq. 11, with $R^* = (3V_{pm}/4\pi)^{1/3}$, thus as for an hypothetical, equivalent spherical solute. Only the parameter $\Delta\gamma$ has to be adjusted since it stands here for an effective value accounting for a mean Lennard-Jones attraction which was not present in our derivation of Sec IV A. We find an optimal value $\Delta\gamma = -6.9 N/m^2$. For a purely repulsive hard-sphere in TIP3P water, one would expect $\Delta\gamma = -39.2 N/m^2$ with a computed HNC value of $13.1 N/m^2$ and $\gamma_{sim} = 53.2 N/m^2$ from Ref.⁹³; the corresponding curve is presented in Fig. 10. We have also represented there the horizontal line corresponding to a simpler correction of the form $a * PV_{pm}$, compatible with previous suggestions⁸⁴, with an optimal value $a = 0.86$. This type of correction only applies because the range of V_{pm} values that are spanned is relatively small. In contrast to the formula in eq. 11 this correction gives an incorrect limit when V_{pm} becomes larger. In the panels b and c of Fig. 9, we report the new correlations between the MDFT-HNC and simulation HSE's with just a pressure correction renormalized by the constant factor 0.86, or applying the more elaborated analytical form of eq. 11, which gives a better description of the surface effects and yields the correct large volume limit. From panels a to c, one goes initially from a RMSE of 3 kcal/mol and a correlation of 0.56, to 0.55 kcal/mol and 0.8, and finally to 0.5 kcal/mol and 0.9. Note that this agreement is obtained with a very rude, spherical approximation for the Van der Waals surface area, which could certainly be improved. In Particular, the slope of the correlation should be corrected. We note in passing that in Refs.^{86,87} we have proposed a so-called PC+ correction in which the pressure in eq. 7 should be replaced by the excess pressure, i.e., $P_{exc} = P_{hnc} - P_{id} = P_{hnc} - k_B T n_b$. This PC+ correction did improve the numerical results compared to the bare PC correction, and it has been uptaken in the RISM community³⁴. For TIP3P, it turns out that

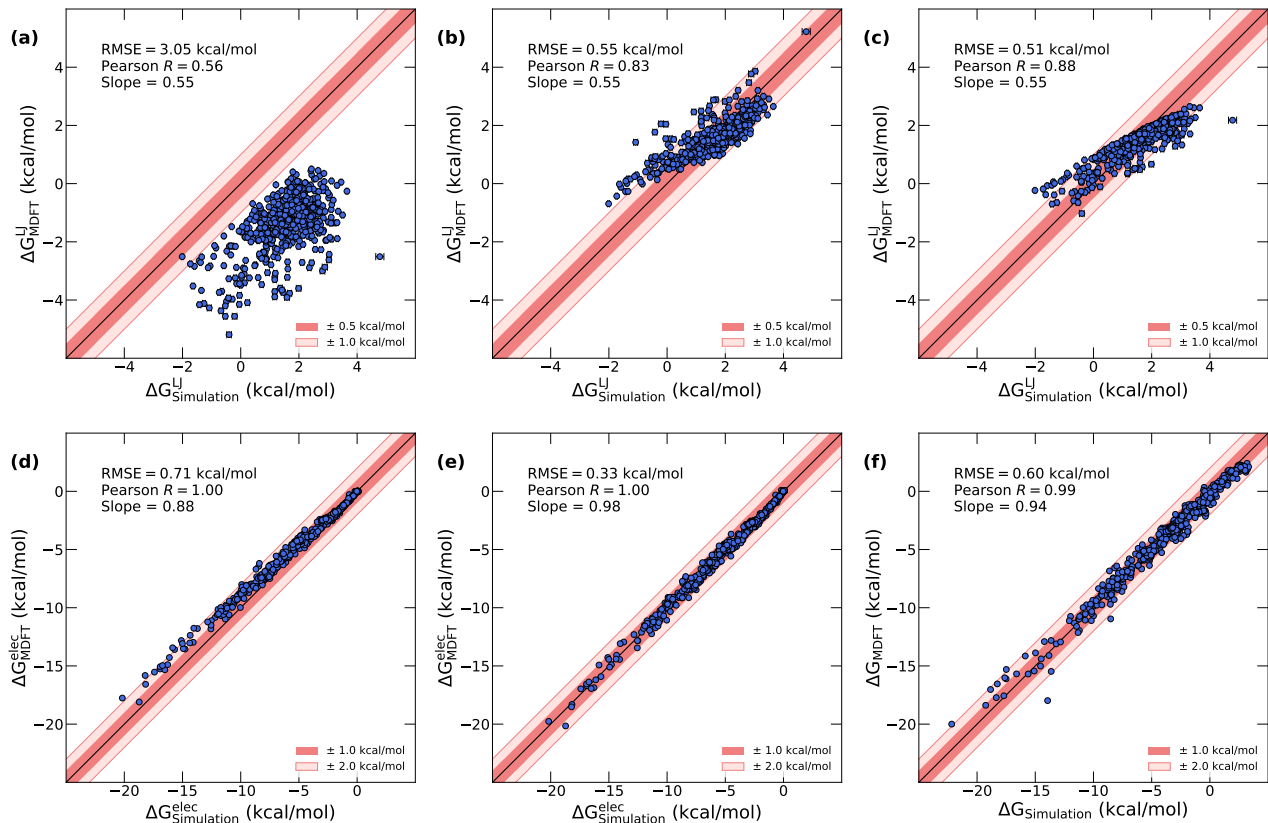


Figure 9. Comparison between MDFT-HNC and simulation for the FreeSolv database. The top line corresponds to the bare MDFT outputs for the non-polar part (LJ) plus **a.** a parameter-free pressure correction, $-P_{hnc}V_{pm}$, **b.** a renormalized pressure correction $-0.86P_{hnc}V_{pm}$ and **c.** The volume/surface energy correction of eq. 11. The panels in the bottom line correspond to **d.** the bare MDFT outputs for the electrostatic solvation free energy, **e.** the same plus an empirical pressure-like correction $+0.6P_{hnc}\Delta V_{pm}$ and **f.** the total HFE including the corrections of **c.** and **e.**

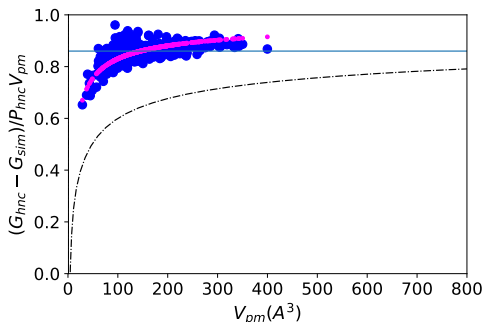


Figure 10. Exact correction to HNC normalized by the first order-pressure correction (the quantity $a(R^*)$ in eq. 10) for the Lennard-Jones contribution to the solvation free-energy. Each blue dot represents a molecule in Mobley's dataset. The dots in magenta correspond to the analytical fit of eq. 11, using an effective spherical radius $R^* = (3V_{pm}/4\pi)^{1/3}$. The parameter free-result for a purely repulsive sphere in TIP3P water is given by the dashed line.

$P_{exc}/P_{hnc} = 0.86$, that is exactly what we now suggest by introducing surface contributions. With our current understanding, we consider this agreement as satisfying, but fortuitous.

In Fig 9-d, we have correlated MDFT-HNC and simulations for the electrostatic contribution of the HFE, i.e. $\Delta G^{elec} = \Delta G - \Delta G^{LJ}$, where ΔG is computed for the fully charged solutes. Without any correction, we observe already a very good agreement with a RMSE of roughly half a kcal/mol and a correlation of 1, but nevertheless a mean slope of 0.9 instead of 1. We find that the agreement can even be improved to a RMSE of 0.3 kcal/mol and a slope of nearly 1 by adding a pressure-like correction $+0.6P\Delta V_{pm}$, involving the difference of the partial molar volumes with and without charges; see Fig 9-e. ΔV_{pm} is always negative, and this new correction goes with an opposite sign with respect to the standard one. It means that the regular pressure/surface tension correction above, roughly $-0.86P\Delta V_{pm}$, is overcompensated by electrostatic effects that we do not yet understand; this correction remains empirical at this stage. As discussed in Sec. IV B, it also improves greatly the free-energy predictions for the charged entities, including the anions. The improved RMSE proves at least that a physically relevant descriptor has been identified.

Overall, Fig 9-f displays the final correlation results adding both the Lennard-Jones and electrostatic contributions. For each solute, this requires two independent minimisations, with and without the solute partial charges.

One important contribution of this work was to show that HNC should be corrected, at least, by both volume and surface effects when solvating micro- to nano-solutes. Thus reducing the parametrisation to its minimum to capture the correct physics, i.e. a single parameter $\Delta\gamma$ correcting the pressure correction by surface effects, and no empirical correction of the electrostatics, yields a RMSE of 0.8 kcal/mol. Accounting for the full story reported above, i.e. incorporating in addition to the well justified one-parameter correction for the LJ contribution, another one-parameter correction for the electrostatic contribution yields a RMSE of 0.6 kcal/mol, a correlation of 0.99 and a correlation slope very close to 1. This is accomplished with a speed-up of, at least, 3 orders of magnitude compared to direct simulations. It takes less than an hour to compute HFEs for the whole Freesolv database by MDFT-HNC.

VI. CONCLUSION

In this paper we have explored the capability of MDFT at its lowest degree of approximation, i.e. with a functional developed at order $\Delta\rho^2$ with respect to the uniform fluid, to predict the hydration free-energies and microscopic structure of a variety of molecular solutes. To assess the method, a detailed comparison was done with respect to our own MC simulations performed under the very same conditions, identical force field and rigid solute geometries. We found many features that can be considered successful. In particular the shape of the cavities (vanishing densities and rising first peak) are perfectly reproduced, the structure around neutral, or moderately charged negative sites is perfectible but correct. The fine structure around positive charges is excellent, as exemplified by the monovalent cation series; so are the associated solvation free-energies. Overall, the electrostatic contribution to solvation free-energies of neutral molecules is correctly reproduced. The first obvious caveat concerns what should be the simplest case in view of the existing theories, in particular scaled-particle theory: the cost of creating cavities. Here the shortcuts of the second order approximation carried by HNC are well identified^{56,57,98}. In this paper we have proposed a new, one-parameter correction inspired by scaled-particle theory and accounting for both volume and surface effects, that makes it possible to improve *a-posteriori* the HFE prediction for the uncharged (Lennard-Jones only) solutes. This correction could certainly be improved by a better definition of the solute Van-der-Waals surface and, beyond, by making use of the morphometric approach of Roth and collaborators^{17–19}. The other important deficiency concerns the microscopic structure around cations and more generally H-bond acceptors which evidently lacks some tetrahedral order, as was noted before⁵⁰. This affects their solvation free energies. One lesson of this work, already noted by others, is that for future improvement it is certainly wise to proceed in two steps, as is done usually in simulations: first introducing the nude LJ interaction, then adding the charges in a second step. For the first part, one can reasonably hope that spherical bridge corrections, be it a hard-sphere bridge⁵⁵ or even simpler weighted density variants⁵⁸ will do the job.

New closures requiring pressure and free energy consistency have been proposed recently and could be tried also⁹⁹. Re-assembling those ideas in a consistent way and validating them for the systems studied in this paper will be the focus of a forthcoming work. Finding an appropriate bridge functional to correct the problem for anions will be more difficult, since it will require to tackle the angular dependence; we think that this is not out of reach. A proposal for reinforcing tetrahedral order was already made in Ref.⁵⁰; it is not a solute-independent bridge functional as we would like to aim for.

For the time being, including the appropriate cavity corrections described above, the method does yield already a precision of 0.8 kcal/mol for the hydration free energies of the whole Mobley's dataset compared to simulations. Inclusion of a second cavity term slightly correcting the electrostatic part brings it to 0.6 kcal/mol. Such precisions are already smaller than that obtained by straight MD simulations comparison to experiments (roughly 1 kcal/mol) so that a direct comparison of MDFT to experiment, and direct improvement of the force fields, begins to make sense. This is accomplished with a three-orders of magnitude numerical speed-up with respect to molecular simulations.

ACKNOWLEDGEMENT

This work has been supported by the Agence Nationale de la Recherche, projet ANR BRIDGE AAP CE29. DB thanks the "Programma Visiting Professor/Scientist" funded by Regione Autonoma della Sardegna, L.R. 7/2007, for sponsoring the visiting professor grant he spent at the University of Cagliari, Department of Physics, where part of his contribution was achieved.

REFERENCES

- ¹B. ROUX and T. SIMONSON, *Biophys. Chem.* **78**, 1 (1999).
- ²J. SRINIVASAN, T. CHEATHAM, P. CIEPLAK, P. KOLLMAN, and D. CASE, *J. Am. Chem. Soc.* **120**, 9401 (1998).
- ³D. BASHFORD and D. A. CASE, *Ann. Rev. Phys. Chem.* **51**, 7 (2000).
- ⁴J. MONGAN, C. SIMMERLING, J. A. MCCAMMON, D. A. CASE, and A. ONUFRIEV, *J. Chem. Theory Comput.* **3**, 156 (2007).
- ⁵J. DZUBIELLA, J. M. J. SWANSON, and J. A. MCCAMMON, *Phys. Rev. Lett.* **96**, 1 (2006).
- ⁶R. W. ZWANZIG, *J. Chem. Phys.* **22**, 1420 (1954).
- ⁷J. G. KIRKWOOD, *J. Chem. Phys.* **3**, 300 (1935).
- ⁸C. H. BENNETT, *J. Comput. Phys.* **22**, 245 (1976).
- ⁹S. KUMAR, J. M. ROSENBERG, D. BOUZIDA, R. H. SWENDSEN, and P. A. KOLLMAN, *J. Comp. Chem.* **13**, 1011 (1992).
- ¹⁰M. R. SHIRTS and J. D. CHODERA, *J. Chem. Phys.* **129**, 124105 (2008).
- ¹¹R. ABEL, T. YOUNG, R. FARID, B. J. BERNE, and R. A. FRIESNER, *J. Am. Chem. Soc.* **130**, 2817 (2008).
- ¹²T. YOUNG, R. ABEL, B. KIM, B. J. BERNE, and R. A. FRIESNER, *Proc. Natl. Acad. Sci. USA* **104**, 808 (2007).
- ¹³N. MATUBAYASI and M. N. NAKAHARA, *J. Chem. Phys.* **117**, 3605 (2002).
- ¹⁴N. MATUBAYASI and M. N. NAKAHARA, *J. Chem. Phys.* **119**, 9686 (2003).
- ¹⁵H. S. ASHBAUGH and L. R. PRATT, *Rev. Mod. Phys.* **78**, 159 (2006).
- ¹⁶R. ROTH, *J. Phys. Condens. Matter* **22** (2010).

- ¹⁷R. ROTH, Y. HARANO, and M. KINOSHITA, *Phys. Rev. Lett.* **97**, 1 (2006).
- ¹⁸S. CHIBA, Y. HARANO, R. ROTH, M. KINOSHITA, and M. SAKURAI, *J. Comput. Chem.* **33**, 550 (2012).
- ¹⁹Y. HARANO, R. ROTH, and S. CHIBA, *J. Comput. Chem.* **34**, 1969 (2013).
- ²⁰P. KOEHL and M. DELARUE, *J. Chem. Phys.* **132**, 064101 (2010).
- ²¹R. RAMIREZ, R. GEBAUER, M. MARESCHAL, and D. BORGIS, *Phys. Rev. E* **66**, 031206 (2002).
- ²²R. RAMIREZ and D. BORGIS, *J. Phys. Chem. B* **109**, 6754 (2005).
- ²³D. CHANDLER and H. C. ANDERSEN, *J. Chem. Phys.* **57**, 1930 (1972).
- ²⁴F. HIRATA and P. J. ROSSKY, *Chem. Phys. Lett.* **83**, 329 (1981).
- ²⁵D. BEGLOV and B. ROUX, *J. Phys. Chem. B* **101**, 7821 (1997).
- ²⁶A. KOVALENKO and F. HIRATA, *Chem. Phys. Lett.* **290**, 237 (1998).
- ²⁷A. KOVALENKO and F. HIRATA, *J. Chem. Phys.* **113**, 2793 (2000).
- ²⁸A. KOVALENKO and F. HIRATA, *J. Phys. Chem. B* **103**, 7942 (1999).
- ²⁹Y. LIU, S. ZHAO, and J. WU, *J. Chem. Theory Comput.* **9**, 1896 (2013).
- ³⁰Y. LIU, J. FU, and J. WU, *J. Phys. Chem. Lett.* **4**, 3687 (2013).
- ³¹G. N. CHUEV, M. V. FEDOROV, and J. CRAIN, *Chem. Phys. Lett.* **448**, 198 (2007).
- ³²D. S. PALMER, A. I. FROLOV, E. L. RATKOVA, and M. V. FEDOROV, *Mol. Pharm.* **8**, 1423 (2011).
- ³³J. F. TRUCHON, B. M. PETTITT, and P. LABUTE, *J. Chem. Theory Comput.* **10**, 934 (2014).
- ³⁴M. MISIN, D. S. PALMER, and M. V. FEDOROV, *J. Phys. Chem. B* **120**, 5724 (2016).
- ³⁵T. LUCHKO, N. BLINOV, G. C. LIMON, K. P. JOYCE, and A. KOVALENKO, *J. Comput. Aided. Mol. Des.* **30**, 1115 (2016).
- ³⁶J. JOHNSON, D. A. CASE, T. YAMAZAKI, and T. LUCHKO, *J. Phys. Cond. Matt.* **28**, 344002 (2016).
- ³⁷D. ROY and A. KOVALENKO, *J. Phys. Chem. A* **123**, 4087 (2019).
- ³⁸N. YOSHIDA, T. IMAI, S. PHONGPHANPHANEE, A. KOVALENKO, and F. HIRATA, *J. Phys. Chem. B* **113**, 873 (2009).
- ³⁹T. LUCHKO, S. GUSAROV, D. R. ROE, C. SIMMERLING, A. DAVID, J. TUSZYNSKI, and A. KOVALENKO, *J. Chem. Theory Comput.* **6**, 607 (2010).
- ⁴⁰M. C. STUMPE, N. BLINOV, D. WISHART, A. KOVALENKO, and V. S. PANDE, *J. Phys. Chem. B* **115**, 319 (2011).
- ⁴¹J. JOHNSON, D. A. CASE, T. YAMAZAKI, S. GUSAROV, A. KOVALENKO, and T. LUCHKO, *J. Phys. Cond. Matt.* **28** (2016).
- ⁴²C. NGUYEN, T. YAMAZAKI, A. KOVALENKO, D. A. CASE, M. K. GILSON, T. KURTZMAN, and T. LUCHKO, *PLoS One* **14**, 1 (2019).
- ⁴³G. M. GIAMBASU, D. A. CASE, and D. M. YORK, *J. Am. Chem. Soc.* (2019).
- ⁴⁴J. RICHARDI, C. MILLOT, and P. H. FRIES, *J. Chem. Phys.* **110**, 1138 (1999).
- ⁴⁵M. LOMBARDERO, C. MARTÍN, S. JORGE, F. LADO, and E. LOMBA, *J. Chem. Phys.* **110**, 1148 (1999).
- ⁴⁶L. BELLONI and I. CHIKINA, *Mol. Phys.* **112**, 1246 (2014).
- ⁴⁷L. GENDRE, R. RAMIREZ, and D. BORGIS, *Chem. Phys. Lett.* **474**, 366 (2009).
- ⁴⁸D. BORGIS, L. GENDRE, and R. RAMIREZ, *J. Phys. Chem. B* **116**, 2504 (2012).
- ⁴⁹S. ZHAO, R. RAMIREZ, R. VUILLEUMIER, and D. BORGIS, *J. Chem. Phys.* **134** (2011).
- ⁵⁰G. JEANMAIRET, M. LEVESQUE, R. VUILLEUMIER, and D. BORGIS, *J. Chem. Phys. Lett.* **4**, 619 (2013).
- ⁵¹L. BLUM and A. J. TORRUELLA, *J. Chem. Phys.* **56**, 303 (1972).
- ⁵²L. BLUM, *J. Chem. Phys.* **57**, 1862 (1972).
- ⁵³L. DING, M. LEVESQUE, D. BORGIS, and L. BELLONI, *J. Chem. Phys.* **147**, 094107 (2017).
- ⁵⁴S. ZHAO, R. RAMIREZ, R. VUILLEUMIER, and D. BORGIS, *J. Chem. Phys.* **134**, 194102 (2011).
- ⁵⁵M. LEVESQUE, R. VUILLEUMIER, and D. BORGIS, *J. Chem. Phys.* **137**, 034115 (2012).
- ⁵⁶G. JEANMAIRET, M. LEVESQUE, and D. BORGIS, *J. Chem. Phys.* **139**, 154101 (2013).
- ⁵⁷G. JEANMAIRET, M. LEVESQUE, V. SERGHEVSKIY, and D. BORGIS, *J. Chem. Phys.* **142**, 154112 (2015).
- ⁵⁸C. GAGEAT, D. BORGIS, and M. LEVESQUE, *arXiv:1709.10139* (2017).
- ⁵⁹S. ZHAO, Z. JIN, and J. WU, *J. Chem. Phys.* **115**, 6971 (2011).
- ⁶⁰S. ZHAO, Z. JIN, and J. WU, *J. Phys. Chem. B* **115**, 15445 (2011).
- ⁶¹Y. LIU, S. ZHAO, and J. WU, *J. Chem. Theory Comput.* **9**, 1896 (2013).
- ⁶²L. BELLONI, *J. Chem. Phys.* **151**, 021101 (2019).
- ⁶³P. HOHENBERG and W. KOHN, *Phys. Rev.* **136**, B864 (1964).
- ⁶⁴W. KOHN and L. J. SHAM, *Phys. Rev.* **140**, A1133 (1965).
- ⁶⁵N. D. MERMIN, *Phys. Rev.* **137**, A1441 (1965).
- ⁶⁶R. EVANS, *Advances in Physics* **28**, 143 (1979).
- ⁶⁷R. EVANS, *Fundamentals of Inhomogeneous Fluids*, Marcel Dekker, Incorporated, 1992.
- ⁶⁸W. L. JORGENSEN, J. CHANDRASEKHAR, J. D. MADURA, and M. L. KLEIN, *J. Chem. Phys.* **79**, 926 (1983).
- ⁶⁹J. VAN LEEUWEN, J. GROENEVELD, and J. DE BOER, *Physica* **25**, 792 (1959).
- ⁷⁰J. PUIBASSET and L. BELLONI, *J. Chem. Phys.* **136**, 154503 (2012).
- ⁷¹L. BELLONI, *J. Chem. Phys.* **147**, 164121 (2017).
- ⁷²L. BELLONI and J. PUIBASSET, *J. Chem. Phys.* **147**, 224110 (2017).
- ⁷³D. BORGIS, R. ASSARAF, B. ROTENBERG, and R. VUILLEUMIER, *Mol. Phys.* **111**, 3486 (2013).
- ⁷⁴D. DE LAS HERAS and M. SCHMIDT, *Phys. Rev. Lett.* **120**, 218001 (2018).
- ⁷⁵S. W. COLES, D. BORGIS, R. VUILLEUMIER, and B. ROTENBERG, *J. Chem. Phys.* **151**, 064124 (2019).
- ⁷⁶C. JARZYNSKI, *Phys. Rev. Lett.* **78**, 2690 (1997).
- ⁷⁷C. JARZYNSKI, *Phys. Rev. E* **56**, 5018 (1997).
- ⁷⁸D. L. MOBLEY, Experimental and Calculated Small Molecule Hydration Free Energies. UC Irvine: Department of Pharmaceutical Sciences, UCI, 2013.
- ⁷⁹G. D. R. MATOS, D. Y. KYU, H. H. LOEFFLER, J. D. C., M. R. SHIRTS, and D. L. MOBLEY, *J. Chem. Eng. Data* **62**, 1559 (2017).
- ⁸⁰J. WANG, R. M. WOLF, J. W. CALDWELL, P. A. KOLLMAN, and D. A. CASE, *J. Comp. Chem.* **25**, 1157 (2004).
- ⁸¹A. JAKALIAN, B. L. BUSH, D. B. JACK, and C. I. BAYLY, *J. Comp. Chem.* **21**, 132 (2000).
- ⁸²A. JAKALIAN, D. B. JACK, and C. I. BAYLY, *J. Comp. Chem.* **23**, 1623 (2002).
- ⁸³J.-F. TRUCHON, B. M. PETTITT, and P. LABUTE, *J. Chem. Theory Comput.* **10**, 934 (2014).
- ⁸⁴D. S. PALMER, A. I. FROLOV, E. L. RATKOVA, and M. V. FEDOROV, *J. Phys. Cond. Matt.* **22** (2010).
- ⁸⁵D. S. PALMER, A. I. FROLOV, E. L. RATKOVA, and M. V. FEDOROV, *J. Phys. Cond. Matt.* **22** (2010).
- ⁸⁶V. SERGHEVSKIY, G. JEANMAIRET, M. LEVESQUE, and D. BORGIS, *J. Phys. Chem. Lett.* **5**, 1935 (2014).
- ⁸⁷V. SERGHEVSKIY, G. JEANMAIRET, M. LEVESQUE, and D. BORGIS, *J. Chem. Phys.* **143**, 184116 (2015).
- ⁸⁸F. M. FLORIS, M. SELMI, A. TANI, and J. TOMASI, *J. Chem. Phys.* **107**, 6353 (2009).
- ⁸⁹D. M. HUANG, P. L. GEISSLER, and D. CHANDLER, *J. Phys. Chem. B* **105**, 6704 (2001).
- ⁹⁰D. M. HUANG and D. CHANDLER, *J. Phys. Chem. B* **106**, 2047 (2002).
- ⁹¹J. DZUBIELLA and J.-P. HANSEN, *J. Chem. Phys.* **121**, 5514 (2004).
- ⁹²G. HUMMER, S. GARDE, A. E. GARCIA, A. POHORILLET, and L. R. PRATR, *Proc. Natl. Acad. Sci. USA* **93**, 8951 (1996).
- ⁹³C. VEGA and E. DE MIGUEL, *J. Chem. Phys.* **126**, 1 (2007).
- ⁹⁴D. HORINEK, S. MAMATKULOV, and R. NETZ, *J. Chem. Phys.* **130**, 124507 (2009).
- ⁹⁵M. A. KASTENHOLZ and P. H. HUNENBERGER, *J. Chem. Phys.* **124**, 124106 (2006).
- ⁹⁶M. A. KASTENHOLZ and P. H. HUNENBERGER, *J. Chem. Phys.* **124**, 224501 (2006).
- ⁹⁷H. J. C. BERENDSEN, D. VAN DER SPOEL, and R. VAN DRUNEN, *Computer Physics Communications* **91**, 43 (1995).
- ⁹⁸G. RICKAYZEN and A. AUGUSTI, *Molecular Physics* **52**, 1355 (1984).
- ⁹⁹T. TSEDNEE and T. LUCHKO, *Phys. Rev. E* **99** (2019).