



**HAL**  
open science

## Efficient molecular density functional theory using generalized spherical harmonics expansions

Lu Ding, Maximilien Levesque, Daniel Borgis, Luc Belloni

► **To cite this version:**

Lu Ding, Maximilien Levesque, Daniel Borgis, Luc Belloni. Efficient molecular density functional theory using generalized spherical harmonics expansions. *Journal of Chemical Physics*, 2017, 147, pp.094107. 10.1063/1.4994281 . cea-01564512

**HAL Id: cea-01564512**

**<https://cea.hal.science/cea-01564512>**

Submitted on 18 Jul 2017

**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.

Copyright

# Efficient molecular density functional theory using generalized spherical harmonics expansions

Lu Ding,<sup>1</sup> Maximilien Levesque,<sup>2,3</sup> Daniel Borgis,<sup>1,2,3</sup> and Luc Belloni<sup>4</sup>

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

<sup>2</sup>*PASTEUR, Département de chimie, école normale supérieure, UPMC Univ. Paris 06, CNRS, PSL Research University, 75005 Paris, France*

<sup>3</sup>*Sorbonne Universités, UPMC Univ. Paris 06, école normale supérieure, CNRS, Processus d'activation sélective par transfert d'énergie uni-électronique ou radiatif (PASTEUR), 75005 Paris, France*

<sup>4</sup>*LIONS, NIMBE, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France<sup>a)</sup>*

(Dated: July 6, 2017)

We show that generalized spherical harmonics are well suited for representing the space and orientation molecular density in the resolution of the molecular density functional theory. We consider the common system made of a rigid solute of arbitrary complexity immersed in a molecular solvent, both represented by molecules with interacting atomic sites and classical force fields. The molecular solvent density  $\rho(\mathbf{r}, \mathbf{\Omega})$  around the solute is a function of the position  $\mathbf{r} \equiv (x, y, z)$  and of the three Euler angles  $\mathbf{\Omega} \equiv (\theta, \phi, \psi)$  describing the solvent orientation. The standard density functional, equivalent to the HNC closure for the solute-solvent correlations in the liquid theory, is minimized with respect to  $\rho(\mathbf{r}, \mathbf{\Omega})$ . The up-to-now very expensive angular convolution products are advantageously replaced by simple products between projections onto generalized spherical harmonics. The dramatic gain in speed of resolution enables to explore in a systematic way molecular solutes of up to nanometric sizes in arbitrary solvents and to calculate their solvation free energy and associated microscopic solvent structure in at most a few minutes. We finally illustrate the formalism by tackling the solvation of molecules of various complexity in water.

## I. INTRODUCTION

The knowledge of the free energy of solvation or chemical potential of a molecular or macromolecular solute immersed in a molecular solvent like water is the starting point of many applications in different fields. Without surprise, beside experimental work, various numerical theories/simulations have been developed following different directions in order to predict such solvation free energy while minimizing the restitution time. The atomic/molecular level of description where the particles are described by sites interacting via classical force fields (essentially Lennard-Jones and coulombic contributions) offers a good compromise between expensive *ab-initio* treatments (with electronic, quantum mechanics description) and crude continuous solvent models. The numerical difficulty originates from the large number of solvent molecules to take into account. How to solve this statistical mechanical problem? The molecular dynamics or Monte Carlo simulations which explicitly consider up to millions of solvent molecules in a simulation cell seem to be methods of choice for an exact resolution but, in practice, are limited by prohibitive times to solution and associated large statistical uncertainties. Consequently, there is a clear demand for alternative theoretical routes. As usual, since the beginning of the liquid state theory field in the 1950-1960's, the approach based on the Ornstein-Zernike (OZ) equation, the integral equations (IE) or the classical density functional theory (DFT) formalism offers a good candidate for such calculations.<sup>1</sup> The goal is to derive the density of the solvent as a function of its position and its orientation in the vicinity of the solute. For polar solvent like water, the electrostatic couplings and resulting hydrogen-bonding correlations are highly anisotropic and the angular description requires a high level of sophistication. We briefly mention here the Reference Interaction Site Model (RISM) approach which ignores this full molecular analysis and replaces it by site-site correlations, only;<sup>2-6</sup> the gain in simplicity and speed is obvious since the interacting particles are spherical; the price to pay is to deal with phenomenological site-site OZ equation, correlation functions without proper statistical mechanical foundation, and ad-hoc closures. RISM is well-developed in its three-dimensional version<sup>7-12</sup> and has provided valuable insight to a number of physical-chemistry problems<sup>12-17</sup>, including the prediction of solvation free energies<sup>18-24</sup>. A RISM-based density functional theory has also been developed for similar applications<sup>25-27</sup>. To bypass the limitations of the RISM approximation, and some of its pitfalls, we propose here to stay at the more ambitious and demanding, but otherwise more fundamental, full molecular level of description. When the solvent and solute particles keep a simple shape, say 3-sites H<sub>2</sub>O molecules around a spherical ion, it is natural to express the solvent density  $\rho(\mathbf{r}, \mathbf{\Omega}_{12})$  in terms of solute-solvent

<sup>a)</sup>Electronic mail: luc.belloni@cea.fr

separation  $r$  and five Euler angles orientation. This radial description has been studied in great details both in bulk solvent and in solutions. Powerful formalisms which make use of expansions onto rotational invariants or generalized spherical harmonics enabled to solve the molecular Ornstein-Zernike equation (MOZ) and integral equations for various densities, temperatures, and compositions<sup>28-34</sup>. This approach breaks down when the molecular/macromolecular solute particle takes a complicated shape with many interacting sites. In such case, it is desirable to consider the solvent density  $\rho(\mathbf{r}, \mathbf{\Omega})$  as a function of its 3D absolute position  $\mathbf{r} \equiv (x, y, z)$  around the fixed solute and of its absolute orientation with respect to a laboratory frame, characterized by three Euler angles  $\mathbf{\Omega} \equiv (\theta, \phi, \psi)$ . Such approach was developed recently in a DFT framework (named MDFT, for molecular density functional theory, in reference to MOZ)<sup>35-39</sup>. In the current implementation, the formalism requires as input the full angular-dependent direct correlation function of the homogeneous solvent - a difficult problem in itself<sup>40,41</sup> especially to get it precisely at all wave-lengths<sup>33,42</sup>. The computation of the excess free energy requires a double integration over orientations for each spatial grid point, which made it prohibitive to tackle large molecular systems. This computational limitation has been overcome in special cases, such as the point-charge models of water, for which the free-energy functional can be further approximated and expressed in terms of two simpler fields than the full orientational density, namely the density and polarization density fields.<sup>43-47</sup>

The objective of the present work is to develop a formalism and numerical algorithms so efficient they unlock the resolution of 3D-DFT or OZ+IE theories in the general case. Section II recalls the 3D-DFT approach while Section III develops the formalism based on angular projections onto carefully chosen basis of spherical harmonics. A few examples applications are shown in Section IV.

## II. 3D MOLECULAR DFT AND ORNSTEIN-ZERNIKE APPROACH

The goal is to derive the local and orientational molecular solvent density  $\rho(\mathbf{r}, \mathbf{\Omega})$  where the vector  $\mathbf{r} \equiv (x, y, z)$  defines the position of the rigid solvent molecule and  $\mathbf{\Omega} \equiv (\theta, \phi, \psi)$  represents its orientation with respect to a fixed laboratory frame. The direction of the main axis of the molecule is characterized by the colatitude  $\theta$  and longitude  $\phi$  while  $\psi$  is the angle of rotation around this axis. The choice of solvent's origin and main axis should take advantage of the molecular symmetry group. For instance, for the water molecule of point group  $C_{2v}$ , the origin is chosen at the oxygen site while the  $z$  axis is its  $C_2$  main symmetry axis and points from the oxygen to the mid-point between the hydrogens. It will be shown below that choosing high symmetry axes implies notable simplifications.

The starting point of the liquid-state density functional theory (DFT) consists in writing a functional  $F[\rho(\mathbf{r}, \mathbf{\Omega})]$  of the molecular density to be minimized. It is defined as the difference between the grand potential of the solvated solute and the grand potential of the homogeneous solvent at density  $\rho_{\text{bulk}}$ . It is thus by definition the solvation free energy of the solute. Without approximation for the moment, it may be splitted into ideal, external and excess contributions<sup>48,49</sup>:

$$F = F_{\text{ideal}} + F_{\text{ext}} + F_{\text{excess}} \quad (1)$$

The ideal term, coming from the entropy of mixing of the solvent molecules, reads

$$F_{\text{ideal}} = k_{\text{B}}T \iiint d\mathbf{r} \iiint d\mathbf{\Omega} \left[ \rho(\mathbf{r}, \mathbf{\Omega}) \ln \frac{\rho(\mathbf{r}, \mathbf{\Omega})}{\rho_{\text{bulk}}} - \Delta\rho(\mathbf{r}, \mathbf{\Omega}) \right] \quad (2)$$

where  $T$  is the temperature,  $k_{\text{B}}$  is the Boltzmann constant,  $k_{\text{B}}T$  is the thermal energy, and  $\Delta\rho(\mathbf{r}, \mathbf{\Omega}) \equiv \rho(\mathbf{r}, \mathbf{\Omega}) - \rho_{\text{bulk}}$  with  $\rho_{\text{bulk}} \equiv n_{\text{bulk}}/8\pi^2$ .  $n_{\text{bulk}}$  is the bulk density. The external contribution comes from the interaction potential  $V_{\text{ext}}$  between the solute molecule and one solvent molecule:

$$F_{\text{ext}} = \iiint d\mathbf{r} \iiint d\mathbf{\Omega} \rho(\mathbf{r}, \mathbf{\Omega}) V_{\text{ext}}(\mathbf{r}, \mathbf{\Omega}). \quad (3)$$

In the usual case of spherically symmetric site-site interaction potentials,  $V_{\text{ext}}$  reads

$$V_{\text{ext}}(\mathbf{r}, \mathbf{\Omega}) = \sum_{i=\text{solvent site}} \sum_{j=\text{solute site}} v_{ij} (|\mathbf{r} + \mathbf{s}_i(\mathbf{\Omega}) - \mathbf{r}_j|) \quad (4)$$

where  $\mathbf{s}_i$  is the intra vector joining the solvent origin to the site  $i$ . When the DFT is solved inside a cubic cell of edge  $L$  with periodic boundary conditions, the contributions from the neighboring solute images must be added to Eq. 4 in the obvious and usual way. The coulombic  $1/r$  contribution to the external potential is derived by solving the Poisson equation inside the cell. Again, this imposed potential is constant in what follows.

The final, excess term involves the correlations between the solvent molecules perturbed by the neighboring solute. As usual in such liquid-state theory, an approximation must be assumed for this contribution. The bare, well developed and documented functional, first term in an infinite Taylor expansion around the liquid bulk density, reads:

$$\beta F_{\text{excess}} = -\frac{1}{2} \iiint d\mathbf{r}_1 \iiint d\mathbf{\Omega}_1 \iiint d\mathbf{r}_2 \iiint d\mathbf{\Omega}_2 \Delta\rho(\mathbf{r}_1, \mathbf{\Omega}_1) c(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) \Delta\rho(\mathbf{r}_2, \mathbf{\Omega}_2) \quad (5)$$

where  $c(\mathbf{r}_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$  is the bulk solvent-solvent molecular direct correlation function (DCF), which depends on the distance  $r_{12}$  between the two solvent molecules and the five Euler angles characterizing their relative orientation (invariant by translation and rotation of the ensemble  $(\mathbf{r}_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$  with respect to the fixed frame). We remind the reader that even if three Euler angles are necessary to define the orientation of a single molecule, five only are necessary for defining *relative* orientations. The function  $c$  of the bulk solvent for a given temperature and pressure is an input in the present approach and is provided by previous extensive Monte Carlo + IE bulk calculations<sup>33,34</sup>.

The formal functional differentiation of 2 leads to:

$$\rho(\mathbf{r}, \boldsymbol{\Omega}) = \rho_{\text{bulk}} \exp[-\beta V_{\text{ext}}(\mathbf{r}, \boldsymbol{\Omega}) + \gamma(\mathbf{r}, \boldsymbol{\Omega})] \quad (6)$$

where  $\beta = 1/k_{\text{B}}T$  and  $\gamma(\mathbf{r}, \boldsymbol{\Omega})$  represents the indirect (total minus direct) solute-solvent correlation function which is related to the previous functions via the solute-solvent Ornstein-Zernike equation:

$$\gamma(\mathbf{r}_1, \boldsymbol{\Omega}_1) = \iiint d\mathbf{r}_2 \iiint d\boldsymbol{\Omega}_2 c(\mathbf{r}_{12}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \Delta\rho(\mathbf{r}_2, \boldsymbol{\Omega}_2) \quad (7)$$

The integral equation 6 is nothing but the HNC approximation for the solute-solvent correlations, which ignores the so-called bridge function. Inclusion of more sophisticated excess functionals or bridge functions will be investigated in future works.

The excess free energy functional 5 may be written as:

$$F_{\text{excess}} = -\frac{1}{2} \iiint d\mathbf{r}_1 \iiint d\boldsymbol{\Omega}_1 \Delta\rho(\mathbf{r}_1, \boldsymbol{\Omega}_1) \gamma(\mathbf{r}_1, \boldsymbol{\Omega}_1). \quad (8)$$

In practice, the numerical resolution consists in general to describe the cubic cell with a 3D grid of  $N \times N \times N$  spatial positions (grid nodes) and mesh size  $L/N$ .  $N$  is typically below 256 for computer memory reasons. Generalization to parallelepiped cells or different directional mesh resolutions is straightforward. For each of the  $N^3$  grid points, the orientation is characterized by different  $\Omega$  triplets. For simplicity, we use  $N_\theta, N_\phi, N_\psi$  decoupled values  $\theta_i, \phi_j, \psi_k$  chosen from the Gauss quadrature. In general,  $0 \leq \theta_i < \pi$ ,  $0 \leq \phi_j < 2\pi$  and  $0 \leq \psi_k < 2\pi$ . In the case of the  $\text{H}_2\text{O}$  molecule (of symmetry group  $C_{2v}$ ),  $0 \leq \psi_k < \pi$  is sufficient. Typical numbers are 5 – 10 for each of the three angles.

The resolution consists either to numerically minimize the total DFT functional 2 with respect to the solvent density  $\Delta\rho(\mathbf{r}, \boldsymbol{\Omega})$  or, equivalently, to solve the integral equation 6. We choose the former route in the present study. The process is iterative. At convergence,  $\Delta\rho(\mathbf{r}, \boldsymbol{\Omega})$  gives the equilibrium solvent profiles around the solute and the value taken by  $F$  provides the free energy of solvation.

The most demanding and challenging part of the calculation is obviously the excess part 5 or 7 which requires a 6D spatial+angular convolution. The spatial one is naturally performed in the Fourier space where 7 becomes:

$$\hat{\gamma}(\mathbf{q}, \boldsymbol{\Omega}_1) = \iiint d\boldsymbol{\Omega}_2 \hat{c}(\mathbf{q}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \Delta\hat{\rho}(\mathbf{q}, \boldsymbol{\Omega}_2). \quad (9)$$

$\mathbf{q} \equiv (q_x, q_y, q_z)$  is the vector in the Fourier space, each component  $q_i$  is discretized in  $N$  values multiples of  $2\pi/L$ . The hat functions indicate the 3D Fourier transformed functions, defined as  $\hat{f}(\mathbf{q}, \boldsymbol{\Omega}) = \iiint f(\mathbf{r}, \boldsymbol{\Omega}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}$ . They are complex quantities. Of course, we use state-of-the-art FFT libraries to compute the space convolution with a complexity in  $\mathcal{O}(N \log N)$  instead of the  $\mathcal{O}(N^2)$  in the naive implementation. The angular convolution which remains in the MOZ equation 9, when implemented straightforwardly in refs<sup>36,37</sup>, represents the main barrier for an efficient resolution: for each of the  $N^3$  values of  $\mathbf{q}$  and for each of the orientation triplet  $\boldsymbol{\Omega}_1$ , one must perform a 3D integral over the whole orientation triplet  $\boldsymbol{\Omega}_2$  using angular quadratures! Indeed, even the naive implementation is not so straightforward in practice since, expressed in the laboratory frame, the 8-variables angular DCF  $\hat{c}(\mathbf{q}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$  is too large to be stored. This problem can be solved by storing the DCF in the so-called intermolecular frame, for which the  $z$ -axis is taken in the direction of  $\mathbf{q}$ , so that  $\hat{c}(q, \boldsymbol{\Omega}'_1, \boldsymbol{\Omega}'_2)$  can be expressed as a function of only 6 variables when accounting from rotational invariance around  $\mathbf{q}$ . For each value of  $\mathbf{q}$ , one thus needs also to infer the correspondence between orientations  $\boldsymbol{\Omega}_i$  and  $\boldsymbol{\Omega}'_i(\mathbf{q}, \boldsymbol{\Omega}_i)$  in the fixed and molecular frame, respectively. This process, whatever the algorithm (storing or recomputing), further impairs the numerical efficiency.

In the next Section, we show that the use of expansions onto basis of generalized spherical harmonics will (i) advantageously replace the angular convolution by simple products between projections, and (ii) reduce the memory footprint of the storage of the DCF.

### III. EXPANSION ONTO GENERALIZED SPHERICAL HARMONICS

The angular dependency of the solvent density  $\Delta\rho(\mathbf{r}, \boldsymbol{\Omega})$  is expanded for each point  $\mathbf{r}$  of the 3D network onto a basis of carefully chosen functions, the generalized spherical harmonics  $R_{\mu'\mu}^m(\boldsymbol{\Omega})$  following Messiah and Blum's notations<sup>28,50</sup>:

$$\Delta\rho(\mathbf{r}, \boldsymbol{\Omega}) = \sum_{m=0}^{n_{\max}} \sum_{\mu'=-m}^m \sum_{\mu=-m}^m f_m \Delta\rho_{\mu'\mu}^m(\mathbf{r}) R_{\mu'\mu}^m(\boldsymbol{\Omega}), \quad (10)$$

with

$$R_{\mu'\mu}^m(\boldsymbol{\Omega}) = r_{\mu'\mu}^m(\theta) e^{-i\mu'\phi - i\mu\psi}, \quad (11)$$

where  $r_{\mu'\mu}^m(\theta)$  is the generalized Legendre polynomial and  $f_m = \sqrt{2m+1}$  is a normalization factor. Each labelled coefficient, the so-called projections, in the sum 10 is obtained by angular integral of the original function (projection onto the corresponding basis vector):

$$\Delta\rho_{\mu'\mu}^m(\mathbf{r}) = f_m \iiint \Delta\rho(\mathbf{r}, \boldsymbol{\Omega}) R_{\mu'\mu}^{m*}(\boldsymbol{\Omega}) d\boldsymbol{\Omega}. \quad (12)$$

The expansion in 10 is in principle infinite. In practice, it is truncated at  $m \leq n_{\max}$ , which defines the basis  $\{n_{\max}\}$  of angular functions. In order to be consistent with the prescription of the Gauss quadrature, the number of angles for  $\theta$ ,  $\phi$ , and  $\psi$  will verify  $N_\theta = n_{\max} + 1$ ,  $N_\phi = 2n_{\max} + 1$  and  $N_\psi = 2(n_{\max}/s) + 1$  where  $s$  is the order of the symmetry axis used as main molecular axis for the solvent molecule ( $s = 2$  for  $C_{2V}$  molecules like water) and the division is an integer division. Since the input function  $\Delta\rho(\mathbf{r}, \boldsymbol{\Omega})$  is *real-valued*, a symmetry relation follows between the complex-valued projections  $\Delta\rho_{\mu'\mu}^m(\mathbf{r})$ :

$$\Delta\rho_{\underline{\mu}\mu}^m(\mathbf{r}) = (-1)^{\mu'+\mu} \Delta\rho_{\mu'\mu}^{m*}(\mathbf{r}), \quad (13)$$

where  $\underline{\mu} \equiv -\mu$ . As a consequence, it is sufficient to deal here with  $\mu' \geq 0$  (or  $\mu \geq 0$ ). For  $\text{H}_2\text{O}$  solvent,  $\mu$  is even and the total number of independent projections per spacial grid node is 4, 19, 40, 85, 140 for  $n_{\max} = 1, 2, 3, 4$  and 5, as shown in table I.

The transformation from  $\Delta\rho(\mathbf{r}, \boldsymbol{\Omega})$  to  $\Delta\rho_{\mu'\mu}^m(\mathbf{r})$  through Eq. 10 and 12 is numerically performed using a fast 3-step algorithm<sup>51</sup> described in the appendix. Each  $r$ -projection is then Fourier transformed by FFT

$$\Delta\hat{\rho}_{\mu'\mu}^m(\mathbf{q}) = \iiint \Delta\rho_{\mu'\mu}^m(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r}. \quad (14)$$

Of course, since the angle  $\boldsymbol{\Omega}$  is defined with respect to a fixed frame, independent of  $\mathbf{r}$ , this means that

$$\Delta\hat{\rho}(\mathbf{q}, \boldsymbol{\Omega}) = \sum_{m=0}^{n_{\max}} \sum_{\mu'=-m}^m \sum_{\mu=-m}^m f_m \Delta\hat{\rho}_{\mu'\mu}^m(\mathbf{q}) R_{\mu'\mu}^m(\boldsymbol{\Omega}), \quad (15)$$

and the symmetry relation 13 becomes:

$$\Delta\hat{\rho}_{\underline{\mu}\mu}^m(\mathbf{q}) = (-1)^{\mu'+\mu} \Delta\hat{\rho}_{\mu'\mu}^{m*}(-\mathbf{q}), \quad (16)$$

which halves the number of  $\mathbf{q}$  values to consider.

In the same way, the bulk function can be decomposed into:

$$\hat{c}(\mathbf{q}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \sum_{mnl\mu\nu} \hat{c}_{\mu\nu}^{mnl}(q) \Phi_{\mu\nu}^{mnl}(\hat{\mathbf{q}}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2), \quad (17)$$

where the coefficients  $\hat{c}_{\mu\nu}^{mnl}(q)$  depend here on the norm  $q$  only and the rotational invariants are defined such as to verify the invariance by rotation of the ensemble:

$$\Phi_{\mu\nu}^{mnl}(\hat{\mathbf{q}}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = f_m f_n \sum_{\mu'\nu'\lambda'} \begin{pmatrix} m & n & l \\ \mu' & \nu' & \lambda' \end{pmatrix} R_{\mu'\mu}^m(\boldsymbol{\Omega}_1) R_{\nu'\nu}^n(\boldsymbol{\Omega}_2) R_{\lambda'0}^l(\hat{\mathbf{q}}). \quad (18)$$

The coefficients  $\begin{pmatrix} m & n & l \\ \mu' & \nu' & \lambda' \end{pmatrix}$  are the usual 3-j-symbols. The complex projections  $\hat{c}_{\mu\nu}^{mnl}$  verify symmetry relations because  $c$  is a real-valued function and the solvent molecules 1, 2 are identical:

$$\hat{c}_{\underline{\mu}\underline{\nu}}^{mnl} = (-1)^{m+n+\mu+\nu} \hat{c}_{\mu\nu}^{mnl*} \quad (19)$$

$$\hat{c}_{\nu\mu}^{nml} = (-1)^{m+n} \hat{c}_{\mu\nu}^{mnl} \quad (20)$$

In the case of H<sub>2</sub>O symmetry,  $\mu$  and  $\nu$  are even and  $\hat{c}_{\mu\nu}^{mnl}$  is real-valued if  $l$  is even and pure imaginary if  $l$  is odd. Consequently:

$$\hat{c}_{\underline{\mu\nu}}^{mnl} = (-1)^{m+n} \hat{c}_{\mu\nu}^{mnl*} = (-1)^{m+n+l} \hat{c}_{\mu\nu}^{mnl}. \quad (21)$$

In that case, the number of independent real coefficients is 4, 27, 79, 250, 549 for  $n_{\max} = 1, 2, 3, 4, 5$ , respectively.

What is the interest of all these projections? The angular integral over  $\Omega_2$  in equation 9 now concerns only two spherical harmonics  $R(\Omega_2)$  in equations 15 and 18: it can now be performed analytically!

The calculation is again simplified and accelerated by switching to the local, molecular frame linked to  $\hat{q}$ , taken as principal axis. The orientation of the solvent molecule in this frame is noted  $\Omega'$ . Composition relations between spherical harmonics during the transformation (rotation) from fixed to local frames are simple matrix products, one for each  $m$  indices:

$$\mathbf{R}^m(\Omega) = \mathbf{R}^m(\hat{\mathbf{q}})\mathbf{R}^m(\Omega') \quad (22)$$

$$R_{\mu'\mu}^m(\Omega) = \sum_{\chi} R_{\mu'\chi}^m(\hat{\mathbf{q}})R_{\chi\mu}^m(\Omega') \quad (23)$$

In the local frame, the expansion analogous to equation 15 becomes:

$$\Delta\hat{\rho}(\mathbf{q}, \Omega') = \sum_{m\mu\chi} f_m \Delta\hat{\rho}_{\mu;\chi}^m(\mathbf{q})R_{\chi\mu}^m(\Omega'). \quad (24)$$

The new coefficients (be careful of the new lower indices notation consistent with Blum's) are deduced from the previous ones by a transformation analogous of the so-called  $\chi$ -transform of Blum<sup>28,29</sup>:

$$\Delta\hat{\rho}_{\mu;\chi}^m(\mathbf{q}) = \sum_{\mu'} \Delta\hat{\rho}_{\mu'\mu}^m(\mathbf{q})R_{\mu'\chi}^m(\hat{\mathbf{q}}). \quad (25)$$

For each discrete value of  $\mathbf{q}$ , the ensemble of  $R_{\mu'\chi}^m(\hat{\mathbf{q}})$  projections is calculated using fast recurrence relations depending only on the Cartesian coordinates of  $\mathbf{q}$ <sup>52</sup>. Moreover,  $\mathbf{q}$  and  $-\mathbf{q}$  require a single treatment since

$$R_{\mu'\chi}^m(-\hat{\mathbf{q}}) = (-1)^m R_{\mu'\underline{\chi}}^m(\hat{\mathbf{q}}) = (-1)^{m+\mu'+\chi} R_{\underline{\mu}'\chi}^m(\hat{\mathbf{q}}) \quad (26)$$

In notation  $\chi$ , the general symmetry relation of equation 16 becomes:

$$\Delta\hat{\rho}_{\mu';\chi}^m(\mathbf{q}) = (-1)^{m+\mu'+\chi} \Delta\hat{\rho}_{\mu';\chi}^{m*}(-\mathbf{q}) \quad (27)$$

In the same way, the bulk  $\hat{c}$  function reads in this new frame<sup>28</sup>:

$$\hat{c}(q, \Omega'_1, \Omega'_2) = \sum_{mn\mu\nu\chi} f_m f_n \hat{c}_{\mu\nu;\chi}^{mn}(q) R_{\chi\mu}^m(\Omega'_1) R_{\chi\nu}^n(\Omega'_2) \quad (28)$$

where the new coefficients are deduced from the old ones through the Blum's " $\chi$ -transform":

$$\hat{c}_{\mu\nu;\chi}^{mn}(q) = \sum_{\chi} \begin{pmatrix} m & n & l \\ \chi & \underline{\chi} & 0 \end{pmatrix} \hat{c}_{\mu\nu}^{mnl}(q) \quad (29)$$

Some symmetry relations apply, even for molecules without symmetry:

$$\hat{c}_{\underline{\mu\nu};\chi}^{mn} = (-1)^{m+n+\mu+\nu} \hat{c}_{\mu\nu;\chi}^{mn*} \quad (30)$$

and

$$\hat{c}_{\nu\mu;\chi}^{nm} = (-1)^{m+n} \hat{c}_{\mu\nu;\chi}^{mn}. \quad (31)$$

In the specific case of water,

$$\hat{c}_{\mu\nu;\underline{\chi}}^{mn} = \hat{c}_{\underline{\mu\nu};\chi}^{mn} = (-1)^{m+n} \hat{c}_{\mu\nu;\chi}^{mn*}. \quad (32)$$

Finally, the insertion of expansions 24 and 28 into the OZ convolution product 9 (formally valid for any reference frame, so in particular for the local one) followed by an analytical integration over  $\Omega'_2$  (thanks to the orthogonality of the spherical harmonics) leads to a very simple OZ relation between  $\hat{c}$ ,  $\Delta\hat{\rho}$  and  $\hat{\gamma}$   $\chi$ -projections:

$$\hat{\gamma}_{\mu;\chi}^m(\mathbf{q}) = \sum_{n\nu} (-1)^{x+\nu} \hat{c}_{\mu\nu;\chi}^{mn}(q) \Delta\hat{\rho}_{\nu;\chi}^n(\mathbf{q}) \quad (33)$$

This OZ relation constitutes the main result of the present formalism and manuscript. It replaces the expensive angular convolution product 9 by simple algebraic products between projections in the local frame! This can be seen as the angular analogous of the replacement of spatial convolution in equation 7 by direct product in Fourier space 9. It is important to note that different  $\chi$  values do not mix in 33; there is one simple matrix multiplication for each  $\chi$  value.

Once the  $\hat{\gamma}_\chi$  projections have been derived from the OZ equation, the return to the laboratory frame follows a relation inverse of equation 24:

$$\hat{\gamma}_{\mu'\mu}^m(\mathbf{q}) = \sum_{\chi} \hat{\gamma}_{\mu;\chi}^m(\mathbf{q}) R_{\mu'\chi}^{m*}(\hat{\mathbf{q}}) \quad (34)$$

Note that the transformation between fixed and local frames in 25 and 34 invoke the spherical harmonics  $R(\hat{\mathbf{q}})$ , where  $(\hat{\mathbf{q}})$  is understood as the rotation which goes from the fixed to the local frames. This last one is not defined univocally because there is freedom in the choice of the rotation angle around  $\hat{\mathbf{q}}$ . Fortunately, it is satisfying to verify in the previous analysis that this angle is completely irrelevant in the final result: indeed, the ensemble 25, 33, 34 involves products of the form  $R_{\gamma\chi}^m(\hat{\mathbf{q}}) R_{\gamma\chi}^{m*}(\hat{\mathbf{q}})$  which are really independent of it.

Finally, we apply an inverse FFT to  $\hat{\gamma}_{\mu'\mu}^m(\mathbf{q})$ , then gather all projections of  $\gamma_{\mu'\mu}^m(\mathbf{r})$  like in equation 10. We end up with the desired indirect correlation function  $\gamma(\mathbf{r}, \Omega)$ :

$$\gamma(\mathbf{r}, \Omega) = \sum_{m\mu'\mu} f_m \gamma_{\mu'\mu}^m(\mathbf{r}) R_{\mu'\mu}^m(\Omega) \quad (35)$$

The very expensive original OZ equation 7 has thus been replaced by the series of cheap steps 12, 15, 25, 33, 34, 35:

$$\Delta\rho(\mathbf{r}, \Omega) \rightarrow \Delta\rho_{\mu'\mu}^m(\mathbf{r}) \rightarrow \Delta\hat{\rho}_{\mu'\mu}^m(\mathbf{q}) \rightarrow \Delta\hat{\rho}_{\mu;\chi}^m(\mathbf{q}) \rightarrow \hat{\gamma}_{\mu;\chi}^m(\mathbf{q}) \rightarrow \hat{\gamma}_{\mu'\mu}^m(\mathbf{q}) \rightarrow \gamma_{\mu'\mu}^m(\mathbf{r}) \rightarrow \gamma(\mathbf{r}, \Omega). \quad (36)$$

#### IV. IMPLEMENTATION AND EXAMPLES

We apply the present DFT approach in the case of the SPC/E model of water. The H<sub>2</sub>O solvent molecule is characterized by one LJ site localized at the O site and three partial charges at the O, H, H sites. For this model, the DCF projections  $\hat{c}_{\mu\nu;\chi}^{mn}(q)$  at different orders of accuracy  $n_{max}$  have been previously obtained by combining Monte Carlo simulation data at short distances and HNC closure at long distances and solving the resulting 1D MOZ+mixed IE. The temperature is 298.15 K and the bulk density is 997 g/L<sup>33,42</sup>.

As mentioned above, the functional of equation 1 is minimized with respect to  $\rho(\mathbf{r}, \Omega)$  using the quasi-Newton minimizer L-BFGS<sup>53</sup>. The density is usually initiated at  $\rho_{bulk} \exp(-\beta V_{ext}(\mathbf{r}, \Omega))$ . LBFGS requires at each minimization step the value of the functional and its gradient

$$\frac{\beta\delta F}{\delta\rho(\mathbf{r}, \Omega)} = \log\left(\frac{\rho(\mathbf{r}, \Omega)}{\rho_{bulk}}\right) + \beta V_{ext}(\mathbf{r}, \Omega) - \gamma(\mathbf{r}, \Omega). \quad (37)$$

A typical minimization process using the new method described above for computing  $\gamma(\mathbf{r}, \Omega)$  is illustrated in Fig. 1 for the CO<sub>2</sub> molecule in water; it is seen that in this case a relative error of  $10^{-4}$  is reached after only  $\approx 20$  cycles. To our experience, convergence is reached before  $\approx 35$  steps or never.

We begin by comparing the numerical efficiency of the new method to that of the original, direct method, described by equation 7.

The old, direct method requires to pre-compute and store the DCF in local frame,  $\hat{c}(q, \Omega'_1, \Omega'_2)$  using eq. 28 and then involves three successive steps, namely (i) to fast Fourier transform  $\Delta\rho(\mathbf{r}, \Omega)$  to  $\Delta\rho(\mathbf{q}, \Omega)$ , (ii) to compute  $\gamma(\mathbf{q}, \Omega)$  using the MOZ equation 9 in  $q$  space with the stored DCF, and finally (iii) to inverse Fourier Transform to  $\gamma(\mathbf{r}, \Omega)$ .

On the other side, the new method involves the succession of seven steps described in the previous section. Note that steps 1-2 and 6-7, i.e. angular transforms and spatial transforms, could well be inverted. However it is more efficient to go with the angular transforms first, since for a given  $n_{max}$  there are less projections than angles (see table I), and thus less functions to Fourier transform.

We show in Fig. 2 that the CPU time requested to compute  $F_{exc}$  and  $\gamma(\mathbf{q}, \Omega)$  is indeed much lower with the new algorithm and that it behaves linearly with respect to the chosen number of orientations per grid point,  $N_\Omega$ , whereas it

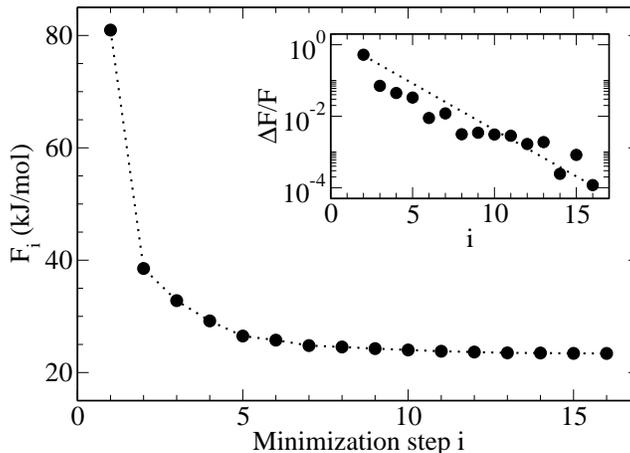


Figure 1. Convergence of the free energy estimation during the minimization process for a  $\text{CO}_2$  molecule in water with a box size  $L = 24 \text{ \AA}$ ,  $N = 72$  and  $n_{max} = 3$ . The inset shows the evolution of the relative difference in the free energy functional between successive steps.

$n_{max}$	for generic solvent molecules			for $C_{2V}$ molecules like $\text{H}_2\text{O}$		
	$N_{\Omega}$	$N_{\text{complex-valued projections}}$	$N_{\text{independent real-valued projections}}$	$N_{\Omega}$	$N_{\text{complex-valued projections}}$	$N_{\text{independent real-valued projections}}$
1	18	10	7	6	4	4
2	75	35	22	45	19	14
3	196	84	50	84	40	28
4	405	165	95	225	85	55
5	726	286	161	330	140	88

Table I. Correspondance between the number of orientations,  $N_{\Omega} = N_{\theta} \times N_{\phi} \times N_{\psi}$ , and the number of projections for generic solvent molecules and for water. The number of independent real projections uses the symmetry rule from equation 13. For  $C_{2V}$  molecules, we add the constraint that  $\psi$  lies in  $[0, \pi[$  and not  $[0, 2\pi[$  and that  $\mu$  is even.

is quadratic in the direct algorithm. The numerical gain is of a factor 200 for  $n_{max} = 3$  (84 orientations) and 750 for  $n_{max} = 5$  (330 orientations). No surprise, the dependence with respect to the number of spatial grid points  $N$  is clearly cubic, as shown in the bottom panel of Fig.2. The quoted CPU times refer to calculations on a single thread on an INTEL Sandy Bridge © processor at 2 GHz. No parallelism of whatever kind is included here. The important information to take from that last figure is that even in that single-thread case, and for  $n_{max} = 3$ , the calculation of  $\gamma(\mathbf{q}, \mathbf{\Omega})$  takes a few seconds for a grid of size  $72^3$ , and above a minute for  $200^3$ .

In Fig. 3 we show the decomposition of the CPU time along the different steps of the algorithm for different  $n_{max}$ : Although the Fast Generalized Spherical Harmonics Transform (FGSHT) is the most time-consuming, the different steps are rather equilibrated. None of them is a bottleneck.

We show furthermore in Fig. 4 that despite the complexity of the computation of  $F_{exc}$  with respect to the straightforward calculation of the local quantities  $F_{ext}$  and  $F_{id}$ , the computational overhead for  $F_{exc}$  appears only a factor 2 with respect to  $F_{id}$  and a factor 8 with respect to  $F_{ext}$ . All in all, with a sufficient grid resolution of 3 points per angstrom and angular resolution  $n_{max} = 3$  (see below), this makes it possible to handle, even on a single core, the solvation of small molecules (typically  $L = 25 \text{ \AA}$ ,  $N \sim 75$ ) within a minute, and that of much larger molecules (e.g.,  $L = 60 \text{ \AA}$ ,  $N \sim 180$ ) in tens of minutes. Those latter calculations were absolutely out of reach with the direct algorithm.

In Fig. 5, we examine the precision of the method for the solvation free energies, taking as example a small organic molecule, pyrimidine, dissolved in water. The three-dimensional solvent structure resulting from the functional minimisation is shown on top of the figure. For this neutral molecule as for many others, we observe that in order to converge the solvation free energy, MDFIT requires a grid resolution of 3 points per Angstrom, a box length of 28 Å (say a dozen of Angstrom of “solvent buffer” from the molecule to the box edge in every direction), and an angular resolution corresponding to  $n_{max} = 3$  (84 orientations).

Such results are corroborated for charged entities too, as shown in Fig. 6 for the toy model corresponding to an hypothetical  $\text{CH}_4^q$  entity, that is a single Lennard-Jones center with parameters corresponding to a unified-atoms representation of methane ( $\sigma = 3.73 \text{ \AA}$ ,  $\epsilon = 1.23 \text{ kJ/mol}$ ) from Asthagiri et al.<sup>54</sup>, with a charge  $q$  at its center. For this very specific spherically symmetric test cases, the 1D integral equation theory is able to solve exactly the same HNC problem, which we use as a test bed. More precisely, MDFIT results are compared to a direct integral equation resolution of the two component system with the solute at infinite dissolution<sup>34</sup>. This last approach implies spherical boundaries that tend

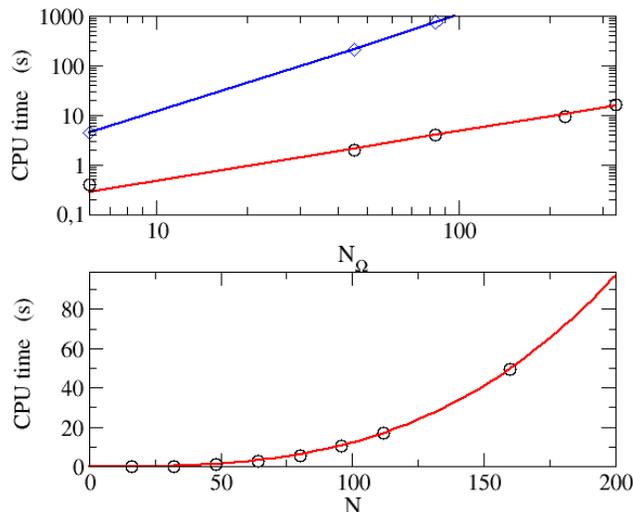


Figure 2. (Top) CPU time for the computation of  $F_{exc}$  within a single minimization step using the direct algorithm (blue diamonds) and the new one (black circles) with  $N = 72$  as a function of number of discrete orientations per spatial grid point,  $N_\Omega$ . The displayed numbers  $N_\Omega = 6, 45, 84, 225, 330$  correspond to  $n_{max} = 1, 2, 3, 4, 5$ , respectively. The red and blue lines represent the best fit to linear behavior,  $T = aN_\Omega$ , and quadratic behavior,  $T = bN_\Omega^2$ , respectively. (Bottom) same quantity as function of spatial grid point number  $N$  using the new algorithm with  $n_{max} = 3$ . The red line represents the best fit to cubic behavior  $T = cN^3$ .

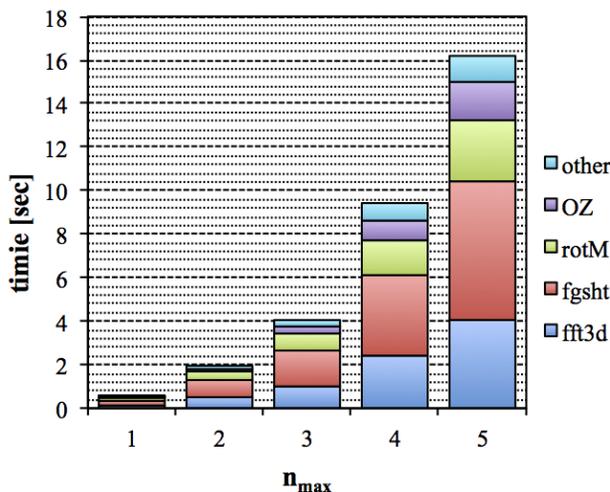


Figure 3. Decomposition of the CPU time of the different steps involved in the calculation of  $\gamma(\mathbf{r}, \Omega)$  from  $\Delta\rho(\mathbf{r}, \Omega)$  at different angular resolutions for different  $n_{max}$ : Fast generalized spherical harmonics transforms (red, steps 1+7); 3D-FFT (blue, steps 2+6); rotations between laboratory and local frames (green, steps 3+5); and resolution of the molecular Ornstein-Zernike equation in the local frame (purple, step 4).

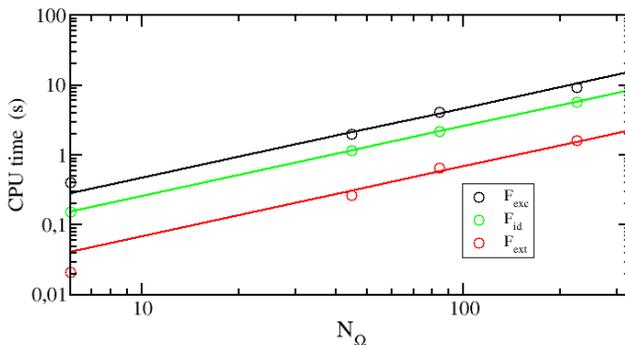


Figure 4. CPU time for the computation of the different components  $F_{id}$ ,  $F_{ext}$ ,  $F_{exc}$  of the solvation free energy for a cubic grid of size  $72^3$  and  $n_{max} = 3$ .

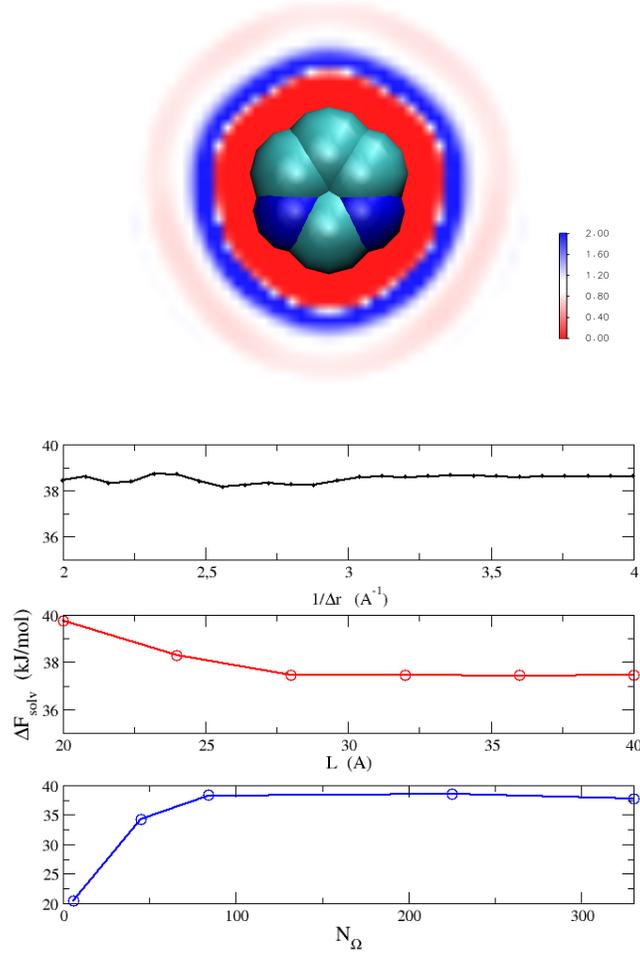


Figure 5. On top, the pyrimidine molecule with CH groups in green and N atoms in blue. The water density map in the plane of the molecule is also shown. Computed solvation free energy of a pyrimidine molecule in water as a function of spatial resolution for  $L = 25 \text{ \AA}$ ,  $n_{\max} = 4$  (top panel), or of box size  $L$  at a given resolution and number of orientations,  $\Delta r^{-1} = 4 \text{ \AA}^{-1}$ ,  $n_{\max} = 4$  (middle panel), or of number of orientation per grid point at fixed box size and spatial resolution ( $L = 25 \text{ \AA}$ ,  $\Delta r^{-1} = 4 \text{ \AA}^{-1}$ ). In this last plot, the 5 points are for  $n_{\max} = 1$  to 5.

toward infinity. In the molecular density functional theory, no restriction apply to the symmetry of the solute molecule and we use a finite box with periodic boundary conditions. Consequently, the results of the minimisation have to be corrected twice for charged systems<sup>55–57</sup>. The first correction is of the Madelung type and accounts for the contribution of the periodic images of the solute and solvent (so-called correction of type B)<sup>56</sup>

$$\Delta F_B = -\xi \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{2L} + \mathcal{O}(L^{-2}), \quad (38)$$

with  $\xi \approx 2.873$  and  $\epsilon = 71$  for SPC/E water<sup>58,59</sup>. The other one originates from the periodic treatment of the electrostatic potential, yielding a vanishing charge density at the box boundary and a finite electrostatic potential in the uniform solvent (type C)

$$\Delta F_C = -(6\epsilon_0)^{-1} q n_{\text{bulk}} \gamma_0, \quad (39)$$

where  $\gamma_0$  is the quadrupole moment of the SPC/E water molecule.

In Fig. 6, we show that the MDFT free energies, including the above corrections, do match the rigorous (but spherically symmetric only) converged HNC-IET results when the angular resolution is increased; within the resolution of the figure, convergence of the solvation free energy is reached for  $n_{\max} = 3$ . We note that both IET and MDFT diverge for  $q = -1$  and  $n_{\max} > 3$ , a failure of the HNC approximation for this artificial solute. Fortunately, this is not the case with Lennard-Jones parameters fitted to model halides, e.g. from<sup>60</sup>. In Fig. 7 and 8, we show the effect of the angular resolution on the solvent structure for  $q = +1, 0$  and  $-0.6$ . We plot there the corresponding radial distribution function (or reduced solvent density) around the solute,  $g(r) = \int d\Omega \rho(\mathbf{r}, \Omega) / n_{\text{bulk}}$  and radial solvent polarisation,  $P(r) = \int d\Omega (\Omega \cdot \hat{\mathbf{r}}) \rho(\mathbf{r}, \Omega) / n_{\text{bulk}}$ .

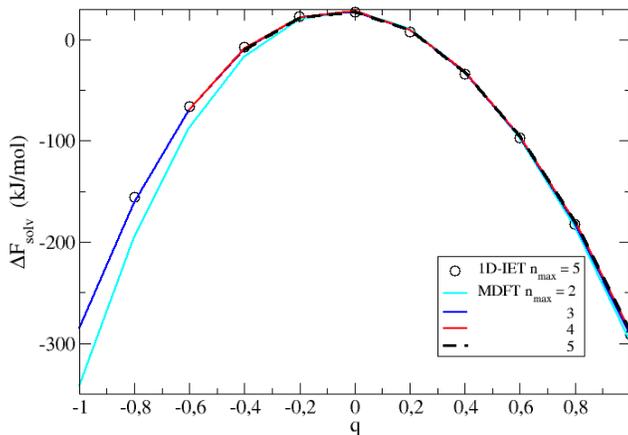


Figure 6. Solvation free energy of a hypothetical  $\text{CH}_4^q$  molecule calculated by MDFT-HNC as a function of its charge  $q$  for different angular resolution  $n_{max} = 2$  to 5. The finite size corrections of equations 38 and 39 are included. For comparison, we also show the exact 1D-IET results that can be calculated for this spherically symmetric case.

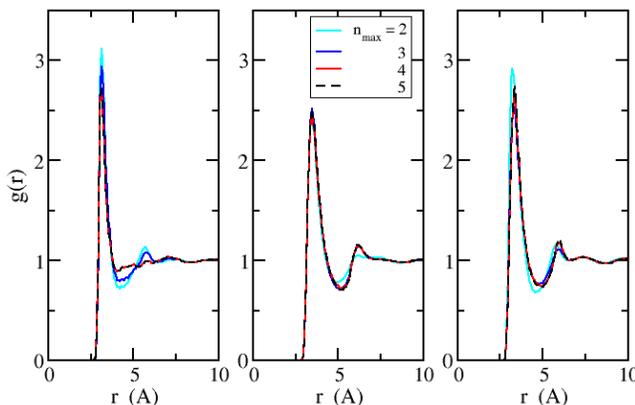


Figure 7. Reduced water density around  $\text{CH}_4^q$  for different angular resolution  $n_{max} = 2$  to 5 and with  $q = +1, 0$  and  $-0.6$  in the left, middle, and right panels, respectively.

Although it appears that for the cationic case  $q = +1$ , only  $n_{max} = 4$  gives a full convergence of the fine structure beyond the first peak,  $n_{max} = 3$  does provides overall an acceptable compromise. It is remarkable that for the neutral case, despite a vanishing electric field, the solute induces an expected but small finite polarisation due to density-orientation couplings. This fine effect is slower to converge with  $n_{max}$ .

In order to illustrate the intrinsic molecular nature of the molecular density functional theory and thus its major advantage over other site-based liquid state theories like 3D-RISM, we show in figure 9 the distribution function  $g$  between  $\text{CH}_4^+$  and the oxygen atom of water as a function of the distance between the two sites,  $r$ , and of the cosine of the angle  $\theta'$  between the water dipole and the axis joining those sites, averaged over all intrinsic rotations  $\psi'$ . We note that in the local framework there is invariance over angle  $\phi'$ . We see that the maximum probability is found for a distance of 3.1 Å and a cosine between 0.7 and 0.9. A cosine of 1 accounts for the oxygen atom pointing exactly toward the cation. Without solvent-solvent correlations, all water molecules would have their dipole pointing exactly toward the cation and would thus have such cosine of 1. It is not the case here : it is favorable to point slightly off the cation ( $\cos \theta_{max} \approx 0.8$ , not 1) but to keep a more favorable short range order, i.e., to keep more of the hydrogen bond network.

For a given distance  $r = 3.1$  Å, that is in the maximum of the radial distribution function, we show the effect of  $\cos \theta'$  and  $\psi'$  in figure 10. First, we see that the distribution is symmetric around  $\psi' = \pi/2$ , as expected from a  $C_{2v}$  molecule in the reference framework. The maximum of the distribution is again found for  $\cos \theta'$  between 0.65 and 0.75, that is for a dipole pointing roughly toward the cation. For a dipole perpendicular to the solute-oxygen vector, that is for  $\cos \theta' = 0$ , we see that the internal rotation of  $\psi' = \pi/2$  that produces the two hydrogen the farthest from the cation is much more probable than other internal rotations.

We conclude by a proof of concept to show that this formalism is efficient enough to unlock the description of the solvation around large molecular solutes. In Fig. 11, we show the water structure around a protein made of 4000 atoms corresponding to 230 residues. We use a grid of  $128^3$  nodes, an angular resolution corresponding to  $n_{max} = 3$  and a discretization of

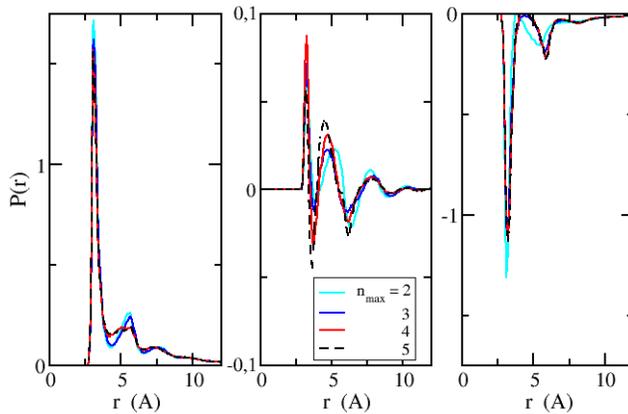


Figure 8. Polarisation density around  $\text{CH}_4^q$  for different angular resolution  $n_{max} = 2 - 5$  and with  $q = +1, 0, -0.6$  in the left, middle, and right panels, respectively.

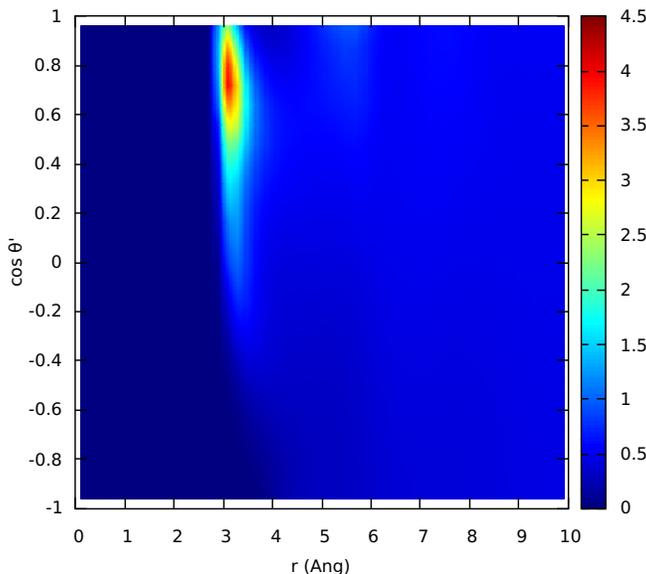


Figure 9. Distribution function between a  $\text{CH}_4^+$  and a water molecule as a function of the  $\text{CH}_4^+ - \text{O}$  distance  $r$  and the cosine of the angle  $\theta'$  between the water dipole and the axis joining the two sites. For each  $r$  and  $\cos \theta'$ , we average over all values of intrinsic rotation angle  $\psi'$ . The distribution is not dependent of  $\phi'$  in this local frame. This is thus a plot of  $\langle g(r, \cos \theta') \rangle_{\psi'}$ .

the grid of  $0.5 \text{ \AA}$ . The overall minimisation took 2 minutes on 24 distributed cores. Using MD simulations, an equivalent statistics for the water density requires at least 100 ns and hundreds of cpu-hours with the same computer resources. It would be even more challenging to get the whole angular-dependent density  $\rho(\mathbf{r}, \boldsymbol{\Omega})$ , a direct output of the functional approach.

## V. CONCLUSION

The three-dimensional density functional theory and integral equation formalism at the molecular level of description of the solvent has been greatly improved by using the concept of expansions/projections onto generalized spherical harmonics. The present analysis of the Ornstein-Zernike convolution product follows that previously developed in bulk systems. The resulting algorithm decreases the time-to-solution by many orders of magnitude. This makes it possible to study in a systematic and routine way many solute/solvent mixtures and to provide free energies of solvation with restitution times of at most a few minutes. Applications to simple molecular solutes in water have been presented. A detailed assessment of the method with respect to reference MD calculations or experimental data, as well as examination of large molecular systems of biological interest, like the prediction of protein hydration, will be reported soon in a companion paper.

The general algorithm presented in this paper could be accelerated following different directions –not speaking of making it highly parallel. First, it is important to note that the  $\gamma$  function is a convolution product. It is thus smoother (both

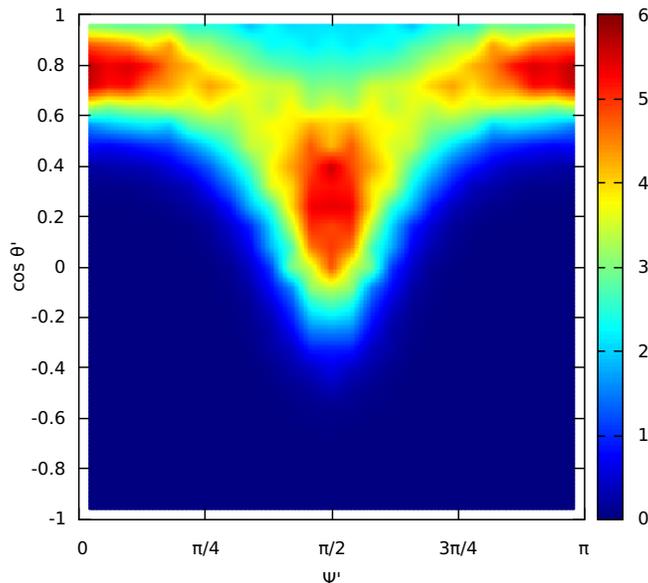


Figure 10. Distribution function between a  $\text{CH}_4^+$  and the water molecule separated by the distance  $r = 3.1 \text{ \AA}$  as a function of the cosine of the angle  $\theta'$  between the site-site axis and the water dipole, and of the intrinsic rotation angle  $\psi'$ .

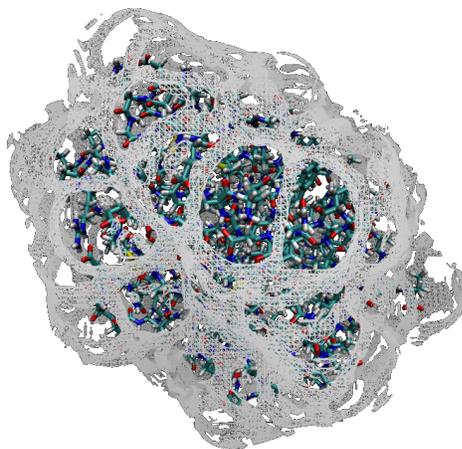


Figure 11. Water density around a protein made of 230 residues and 4000 atomic sites (4M7G: Streptomyces Erythraeus Trypsin). The displayed isosurface correspond to 3 times the bulk density.

in spatial and angular dependence) than its two building blocks  $\Delta\rho$  and  $c$ . As a consequence, it is legitimate to use a degraded basis  $\{n'_{\max}\}$  with  $n'_{\max} < n_{\max}$  during the entire process. Also, inhomogeneous grids in space and orientations seem logical extensions to the important milestone reported therein. These techniques may lead to further substantial decrease in time-to-solution without altering the precision.

Now that the numerical barrier has been unlocked remains an important question. As usual in such liquid-state theories, the validity of the HNC-like DFT functional has to be challenged, and one will have to go beyond this approximation by building solute-solvent bridge function(al)s. We already have several suggestions in that directions, either based on global thermodynamic corrections<sup>44,45,61,62</sup> or on a detailed understanding of the bridge functions for simple molecular systems<sup>63</sup>.

## APPENDIX: ANGULAR REPRESENTATION VERSUS PROJECTIONS

The expansion (10) and the projection (12) which transform triplets of angles  $\mathbf{\Omega} \equiv (\theta, \phi, \psi)$  to indices  ${}^m_{\mu'\mu}$  or vice versa follow a three-step algorithm originally developed for bulk systems<sup>51</sup>:

First and second steps: transform  $\phi$  and  $\psi$  into  $\mu'$  and  $\mu$ :

$$\Delta\rho_{\mu'\mu}(\theta) = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \Delta\rho(\theta, \phi, \psi) e^{+i\mu'\phi + i\mu\psi} d\phi d\psi \quad (40)$$

$$\Delta\rho(\theta, \phi, \psi) = \sum_{\mu'=-n_{\max}}^{n_{\max}} \sum_{\mu=-n_{\max}}^{n_{\max}} \Delta\rho_{\mu'\mu}(\theta) e^{-i\mu'\phi - i\mu\psi} \quad (41)$$

The 2D angular integral 40 is performed by trapezoidal rule (or Gauss Chebychev quadrature):

$$\Delta\rho_{\mu'\mu}(\theta) = \frac{1}{N_\phi N_\psi} \sum_{j=0}^{N_\phi-1} \sum_{k=0}^{N_\psi-1} \Delta\rho(\theta, \phi_j \equiv j \frac{2\pi}{N_\phi}, \psi \equiv k \frac{2\pi}{N_\psi}) e^{+2i\pi(\frac{\mu'j}{N_\phi} + \frac{\mu k}{N_\psi})} \quad (42)$$

One recognizes a discrete 2D Fourier transform which can be efficiently performed by 2D FFT, provided  $N_\phi = N_\psi = 2n_{\max} + 1$ . Same remark for the inverse transformation 41. The case of H<sub>2</sub>O symmetry can be adapted by choosing  $N_\psi = 2(n_{\max}/2) + 1$  angles between 0 and  $\pi$ . This operation must be performed for each  $\theta$  value.

Third step: transformation between  $\theta$  and  $m$ :

$$\Delta\rho_{\mu'\mu}^m = f_m \int_{-1}^1 \frac{d \cos \theta}{2} \Delta\rho_{\mu'\mu}(\theta) r_{\mu'\mu}^m(\theta) = f_m \sum_{i=1}^{N_\theta} w_i \Delta\rho_{\mu'\mu}(\theta_i) r_{\mu'\mu}^m(\theta_i) \quad (43)$$

$$\Delta\rho_{\mu'\mu}^m = \sum_{m=\max(|\mu'|, |\mu|)}^{n_{\max}} \Delta\rho_{\mu'\mu}^m r_{\mu'\mu}^m(\theta) \quad (44)$$

The integral over  $\theta$  is performed using Gauss-Legendre quadrature with  $N_\theta = n_{\max} + 1$  and associated weights  $w_i$ . It is performed for each pair  $\{\mu', \mu\}$ .

Despite the lack of "Fast" transform in this last step, the whole procedure is fast enough not to be the limiting process in the OZ convolution calculation. Overall, we can qualify the whole angles-to-projections process, analogous to a FFT for the angular variable, as a Fast Generalized Spherical Harmonics Transform (FGSHT).

## ACKNOWLEDGMENTS

This work was supported by the Energy oriented Centre of Excellence (EoCoE), grant agreement number 676629, funded within the Horizon2020 framework of the European Union.

## REFERENCES

- <sup>1</sup>J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids: With Applications to Soft Matter*. Academic Press, Amstersdam, 4th edition, 2013.
- <sup>2</sup>David Chandler and Hans C. Andersen. Optimized cluster expansions for classical fluids. II. theory of molecular liquids. *The Journal of Chemical Physics*, 57:1930–1937, 1972.
- <sup>3</sup>Fumio Hirata and Peter J Roscky. An extended rism equation for molecular polar fluids. *Chemical Physics Letters*, 83(2):329–334, 1981.
- <sup>4</sup>B. Montgomery Pettitt and Peter J. Roscky. Integral equation predictions of liquid state structure for waterlike intermolecular potentials. *The Journal of Chemical Physics*, 77:1451–1457, 1982.
- <sup>5</sup>Kippi M. Dyer, John S. Perkyns, and B. Montgomery Pettitt. A site-renormalized molecular fluid theory. *The Journal of Chemical Physics*, 127:194506, 2007.
- <sup>6</sup>Kippi M. Dyer, John S. Perkyns, George Stell, and B. Montgomery Pettitt. A molecular site-site integral equation that yields the dielectric constant. *The Journal of Chemical Physics*, 129:104512, 2008.
- <sup>7</sup>F. Hirata. *Molecular Theory of Solvation*. Springer, 2003.
- <sup>8</sup>Thomas Kloss and Stefan M. Kast. Treatment of charged solutes in three-dimensional integral equation theory. *The Journal of Chemical Physics*, 128:134505–7, 2008.
- <sup>9</sup>Yutaka Maruyama, Norio Yoshida, Hiroto Tadano, Daisuke Takahashi, Mitsuhsa Sato, and Fumio Hirata. Massively parallel implementation of 3D-RISM calculation with volumetric 3D-FFT. *Journal of Computational Chemistry*, 35(18):1347–1355, 2014.
- <sup>10</sup>Volodymyr P Sergiievskiy, Wolfgang Hackbusch, and Maxim V Fedorov. Multigrid solver for the reference interaction site model of molecular liquids theory. *Journal of computational chemistry*, 32:1982–1992, 2011.
- <sup>11</sup>Volodymyr Sergiievskiy. *Modelling of solvation thermodynamics using a combination of reference interaction site model theory and multi-grid numerical methods*. PhD thesis, University of Strathclyde, 2012.

- <sup>12</sup>F. Hoffgaard S. M. Kast, J. Heil. Integral equation theory as a solvation model for classical and quantum solute systems. In G. Sutmann, J. Grotendorst, G. Gompper, and D. Marx, editors, *IAS Series Vol. 28: Computational Trends in Solvation and Transport in Liquids*, pages 419–434, 2015.
- <sup>13</sup>Takashi Imai, Ryusuke Hiraoka, Andriy Kovalenko, and Fumio Hirata. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins: Structure, Function, and Bioinformatics*, 66:804–813, 2006.
- <sup>14</sup>Norio Yoshida, Takashi Imai, Saree Phongphanphanee, Andriy Kovalenko, and Fumio Hirata. Molecular recognition in biomolecules studied by statistical-mechanical integral-equation theory of liquids. *The Journal of Physical Chemistry B*, 113:873–886, 2009.
- <sup>15</sup>David Casanova, Sergey Gusarov, Andriy Kovalenko, and Tom Ziegler. Evaluation of the SCF combination of KS-DFT and 3d-RISM-KH; solvation effect on conformational equilibria, tautomerization energies, and activation barriers. *Journal of Chemical Theory and Computation*, 3:458–476, 2007.
- <sup>16</sup>Jakub W. Kaminski, Sergey Gusarov, Tomasz A. Wesolowski, and Andriy Kovalenko. Modeling solvatochromic shifts using the orbital-free embedding potential at statistically mechanically averaged solvent density. *The Journal of Physical Chemistry A*, 114:6082–6096, 2010.
- <sup>17</sup>Thomas Kloss, Jochen Heil, and Stefan M. Kast. Quantum chemistry in solution by combining 3d integral equation theory with a cluster embedding approach. *The Journal of Physical Chemistry B*, 112:4337–4343, 2008.
- <sup>18</sup>David S. Palmer, Volodymyr P. Sergiievskiy, Frank Jensen, and Maxim V. Fedorov. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *The Journal of Chemical Physics*, 133:044104, 2010.
- <sup>19</sup>Volodymyr P. Sergiievskiy and Maxim V. Fedorov. 3drism multigrid algorithm for fast solvation free energy calculations. *Journal of Chemical Theory and Computation*, 8:2062–2070, 2012.
- <sup>20</sup>Jean-François Truchon, B. Montgomery Pettitt, and Paul Labute. A cavity corrected 3d-RISM functional for accurate solvation free energies. *Journal of Chemical Theory and Computation*, 10:934–941, 2014.
- <sup>21</sup>Volodymyr Sergiievskiy, Guillaume Jeanmairet, Maximilien Levesque, and Daniel Borgis. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. *The Journal of Chemical Physics*, 143:184116, 2015.
- <sup>22</sup>M. Misin, M.V Fedorov, and D. Palmer. Hydration free energies of molecular ions from theory and simulation. *The Journal of Physical Chemistry B*, 120:975–983, 2016.
- <sup>23</sup>J. Johnson, D A Case, T. Yamazaki, and T. Luchko. Small molecule hydration energy and entropy from 3d-rism. *Journal of Physics Condensed Matter*, 28:344002, 2016.
- <sup>24</sup>Nicolas Tielker, Daniel Tomazic, Jochen Heil, Thomas Kloss, Sebastian Ehrhart, Stefan Güssregen, K. Friedemann Schmidt, and Stefan M. Kast. The SAMPL5 challenge for embedded-cluster integral equation theory: solvation free energies, aqueous  $pK_a$ , and cyclohexane-water  $\log D$ . *Journal of Computer-Aided Molecular Design*, 30:1035–1044, 2016.
- <sup>25</sup>Yu Liu, Shuangliang Zhao, and Jianzhong Wu. A site density functional theory for water: Application to solvation of amino acid side chains. *Journal of Chemical Theory and Computation*, 9:1896–1908, 2013.
- <sup>26</sup>J. Fu, L. Liu, and J. Wu. Fast prediction of hydration free energies for sampl4 blind test from a classical density functional theory. *J. computer-aided molecular design*, 28:299, 2014.
- <sup>27</sup>J. Fu, L. Liu, and J. Wu. Molecular density functional theory for multiscale modeling of hydration free energy. *Chem. Engineering Science*, 126:370, 2015.
- <sup>28</sup>L. Blum and A. J. Torruella. Invariant expansion for two-body correlations: Thermodynamic functions, scattering, and the ornstein-zernike equation. *The Journal of Chemical Physics*, 56:303–310, 1972.
- <sup>29</sup>L. Blum. Invariant expansion. II. the ornstein-zernike equation for nonspherical molecules and an extended solution to the mean spherical model. *The Journal of Chemical Physics*, 57:1862–1869, 1972.
- <sup>30</sup>P. H. Fries and G. N. Patey. The solution of the hypernetted-chain approximation for fluids of nonspherical particles. a general method with application to dipolar hard spheres. *The Journal of Chemical Physics*, 82:429–440, 1985.
- <sup>31</sup>J. Richardi, C. Millot, and P. H. Fries. A molecular ornstein-zernike study of popular models for water and methanol. *The Journal of Chemical Physics*, 110:1138–1147, 1999.
- <sup>32</sup>M. Lombardero, C. Martín, S. Jorge, F. Lado, and E. Lomba. An integral equation study of a simple point charge model of water. *The Journal of Chemical Physics*, 110:1148, 1999.
- <sup>33</sup>Joel Puibasset and Luc Belloni. Bridge function for the dipolar fluid from simulation. *The Journal of Chemical Physics*, 136:154503, 2012.
- <sup>34</sup>L. Belloni and I. Chikina. Efficient full newton-raphson technique for the solution of molecular integral equations example of the SPC/E water-like system. *Mol. Phys.*, 112:1246, 2014.
- <sup>35</sup>Rosa Ramirez, Ralph Gebauer, Michel Mareschal, and Daniel Borgis. Density functional theory of solvation in a polar solvent: Extracting the functional from homogeneous solvent simulations. *Physical Review E*, 66:031206–031206–8, 2002.
- <sup>36</sup>Lionel Gendre, Rosa Ramirez, and Daniel Borgis. Classical density functional theory of solvation in molecular solvents: Angular grid implementation. *Chemical Physics Letters*, 474:366–370, 2009.
- <sup>37</sup>Shuangliang Zhao, Rosa Ramirez, Rodolphe Vuilleumier, and Daniel Borgis. Molecular density functional theory of solvation: From polar solvents to water. *The Journal of Chemical Physics*, 134:194102, 2011.
- <sup>38</sup>Daniel Borgis, Lionel Gendre, and Rosa Ramirez. Molecular density functional theory: Application to solvation and electron-transfer thermodynamics in polar solvents. *The Journal of Physical Chemistry B*, 116:2504–2512, 2012.
- <sup>39</sup>Volodymyr P. Sergiievskiy, Guillaume Jeanmairet, Maximilien Levesque, and Daniel Borgis. Fast computation of solvation free energies with molecular density functional theory: Thermodynamic-ensemble partial molar volume corrections. *The Journal of Physical Chemistry Letters*, 5:1935–1942, 2014.
- <sup>40</sup>Rosa Ramirez, Michel Mareschal, and Daniel Borgis. Direct correlation functions and the density functional theory of polar solvents. *Chemical Physics*, 319:261–272, 2005.
- <sup>41</sup>Shuangliang Zhao, Honglai Liu, Rosa Ramirez, and Daniel Borgis. Accurate evaluation of the angular-dependent direct correlation function of water. *The Journal of Chemical Physics*, 139:034503–1–034503–10, 2013.
- <sup>42</sup>Luc Belloni. Angular-dependent bridge functions for SPC/E water. *to be published*.
- <sup>43</sup>Maximilien Levesque, Rodolphe Vuilleumier, and Daniel Borgis. Scalar fundamental measure theory for hard spheres in three dimensions: Application to hydrophobic solvation. *The Journal of Chemical Physics*, 137:034115, 2012.
- <sup>44</sup>Guillaume Jeanmairet, Maximilien Levesque, Rodolphe Vuilleumier, and Daniel Borgis. Molecular density functional theory of water. *The Journal of Physical Chemistry Letters*, 4:619–624, 2013.
- <sup>45</sup>Guillaume Jeanmairet, Maximilien Levesque, and Daniel Borgis. Molecular density functional theory of water describing hydrophobicity at short and long length scales. *The Journal of Chemical Physics*, 139:154101–1–154101–9, 2013.
- <sup>46</sup>Guillaume Jeanmairet, Maximilien Levesque, Volodymyr Sergiievskiy, and Daniel Borgis. Molecular density functional theory for water with liquid-gas coexistence and correct pressure. *The Journal of Chemical Physics*, 142:154112, 2015.

- <sup>47</sup>Guillaume Jeanmairet, Nicolas Levy, Maximilien Levesque, and Daniel Borgis. Molecular density functional theory of water including density-polarization coupling. *Journal of Physics: Condensed Matter*, 28:244005, 2016.
- <sup>48</sup>R. Evans. In D. Henderson, editor, *Fundamental of Inhomogeneous Fluids*, New York, 1992. Marcel Dekker.
- <sup>49</sup>R. Evans. Density functional theory for inhomogeneous fluids i: Simple fluids in equilibrium. In *Lecture notes at 3rd Warsaw School of Statistical Physics*. 2009.
- <sup>50</sup>Albert Messiah. *Quantum Mechanics: v. 2*. North-Holland Publishing Co, 1961.
- <sup>51</sup>F. Lado, E. Lomba, and M. Lombardero. Integral equation algorithm for fluids of fully anisotropic molecules. *The Journal of Chemical Physics*, 103:481, 1995.
- <sup>52</sup>M. S. Gordon C. H. Choi, J. Ivanic and K. Ruedenberg. Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. *The Journal of Chemical Physics*, 111:8825, 1999.
- <sup>53</sup>R. H. Byrd, P. Lu, and J. Nocedal. A limited memory algorithm for bound constrained optimization. *SIAM J. Scient. Stat. Comp.*, 16:1190, 1995.
- <sup>54</sup>D. Asthagiri, Safir Merchant, and Lawrence R. Pratt. Role of attractive methane-water interactions in the potential of mean force between methane molecules in water. *The Journal of Chemical Physics*, 128:244512, 2008.
- <sup>55</sup>M. A. Kastenholtz and Philippe H. Hunenberger. Computation of methodology-independent ionic solvation free energies from molecular simulations. i. the electrostatic potential in molecular liquids. *The Journal of Chemical Physics*, 124:124106, 2006.
- <sup>56</sup>Mika A. Kastenholtz and Philippe H. Hunenberger. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. the hydration free energy of the sodium cation. *The Journal of Chemical Physics*, 124:224501, 2006.
- <sup>57</sup>Philippe Hünenberger and Maria Reif. *Single-ion solvation: experimental and theoretical approaches to elusive thermodynamic quantities*. RSC Pub., Cambridge, 2011.
- <sup>58</sup>H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91:6269–6271, 1987.
- <sup>59</sup>I. M. Svishchev P. G. Kusalik. The spatial structure in liquid water. *Science*, 265:1219–1221, 1994.
- <sup>60</sup>Dominik Horinek, Shavkat I. Mamatkulov, and Roland R. Netz. Rational design of ion force fields based on thermodynamic solvation properties. *The Journal of Chemical Physics*, 130:124507, 2009.
- <sup>61</sup>M. Levesque, R. Vuilleumier, and D. Borgis. Scalar fundamental measure theory for hard spheres in three dimensions. application to hydrophobic solvation. *The Journal of Chemical Physics*, 137:034115, 2012.
- <sup>62</sup>Cédric Gageat, Daniel Borgis, and Maximilien Levesque. Bridge functional for the molecular density functional theory with consistent pressure and surface tension. *to be published*.
- <sup>63</sup>J. Puibasset and L. Belloni. Bridge function for the dipolar fluid from simulation. *The Journal of Chemical Physics*, 136:154503, 2012.