Finite-size Effects on the Calculation of X-ray Scattering from Molecular Dynamics Simulations

A. O. Dohn[1,2], V. I. Markmann[1], A. Nimmrich[1,‡], K. Haldrup[1],
K. B. Møller[3,†], and M. M. Nielsen[1]

[1]: Department of Physics, Technical University of Denmark, 2800 Lyngby, Denmark
[2]: Science Institute and Faculty of Physical Sciences, VR-III, University of Iceland, Reykjavík 107, Iceland
[3]: Department of Chemistry, Technical University of Denmark, 2800 Lyngby, Denmark
[‡]: Curr. affiliation: Department of Chemistry, University of Washington, Seattle, WA 98195-1700, United States
[†]Correspondence: kbmo@kemi.dtu.dk

Abstract

Structural studies using x-ray scattering methods to investigate molecules in solution are shifting focus towards including the role and effects of the surrounding solvent. However, forward models based on molecular dynamics (MD) simulations to simulate structure factors and x-ray scattering from interatomic distributions such as radial distribution functions (RDFs) face limitations imposed by simulations, particularly at low values of the scattering vector q. In this work, we show how the value of the structure factor at $q = 0$ calculated from RDFs sampled from finite MD simulations is entirely dependent on the size of the simulation cell. We derive a new, simple method to correct the RDFs to eliminate this error, and compare it to two other RDF-correction methods developed for Kirkwood-Buff theory applications. We present a quantitative test to assess the reliability of the simulated low-q scattering signal, and show that our RDF-correction successfully recovers the correct $q = 0$ limit for neat water. We investigate the effect of MD-sampling time on the RDF-corrections, before advancing to a more realistic example system comprised of a solvated transition metal complex, sampled in a series of water cells with increasing density. We show that our correction recovers the correct $q = 0$-behaviour for increasing density. Furthermore, we employ a simple continuum scattering model to dissect the total scattering signal from the solvent-solvent structural correlations in a solute-solvent model system to find two distinct contributions: a non-local 'density'-contribution and a local contribution from the solvent shell. We show how the second contribution can be approximated without also including the density-contribution. Finally, we provide a 'best-practices'-checklist for experimentalists planning to incorporate explicit solvation MD simulations in future work, offering guidance for improving the accuracy and reliability of structural studies using x-ray scattering methods in solution.

Introduction

The term *solvation* describes the processes of how solvent molecules interact with solute molecules in a solution. Solvation and solvation dynamics can significantly influence the properties and function of a solute. Since so much chemistry takes place in solution, studying the role of solvation dynamics on a molecular level can be the key to obtaining a complete understanding of reaction mechanisms that are central within (photo)chemical processes with sought-after purposes such as light-harvesting [1–5], charge-transfer [6–8], catalysis [9, 10], or photoswitching [11–14]. Lastly, studies of the neat solvent alone [15] are especially pertinent when that solvent is water [16, 17], simply because of its unquestionable importance both in biology and elsewhere. Direct structural probes such as (time-resolved) x-ray scattering have proved a crucial tool for obtaining mechanistic understanding on an atomic scale [10,18–21]. However, the random orientation of solutes in a solution results in the scattered signal being comprised of an average of all possible orientations of the solute and its solvation shell, which thus reduces the extractable information content in the recorded signal, necessitating support from theoretical and computational modeling. Often, classical molecular dynamics (MD) are used, driven by either quantum [22–24], classical [8,18,23,25,26], or mixed potentials [23,27,28], to provide molecular structures or trajectories that evolve in time, providing models of the studied process. These models need to be linked to experiment through a forward model, calculating the theoretical experimental signal from the molecular simulations [29,30]. As the focus moves further towards the role of the solvent [8,19,26,31–35], the structural changes of solvation shells are more diffuse than for, say, covalent bond-lengths in a molecule. This necessitates more accurate forward models that fully capture these subtle effects across the entire range of the scattering vector q. Furthermore, the models need to be able to discern between real changes in distributions of interatomic distances and artificial contributions to the predicted scattering signals. Such artificial contributions arise from approximations in the simulations, such as the finite size of the simulation cell, and density fluctuations in the solvent that cannot fully be averaged out within reasonable simulation times.

In a previous study, we presented the fundamentals of deriving a forward model for calculating coherent x-ray scattering signals from solute-solvent systems, based on pairwise radial distributions (RDFs) sampled from molecular dynamics simulations [29]. In this work, we investigate how finite-size simulation cells introduce errors in the long-range behaviour of the sampled RDFs (finite-size artefacts), and what effects this has on the predicted scattering signal. We present new methods of eliminating these artefacts by correcting the RDFs to approximate being sampled from infinite systems, and critically asses it alongside two other methods methods to correct the RDFs developed with different purposes, but can also be used to suppress artificial oscillations in the scattering signal. We present methods to discern features in the simulated signal corresponding to real interatomic structure from finite-size artefacts, and analyse a method to obtain contributions to the total x-ray scattering term from solvent-solvent correlations within solvation shells [26].

The paper is structured as follows: First we present the theoretical background for calculating x-ray scattering from RDFs sampled from finite-size systems and introduce our new correction along two previously developed corrections. We then test the performance of the corrections on systems of increasing complexity, going from a simple Lennard-Jones liquid, to neat water, to a single-atom solute in water, before ending up with a many-atom complex solvated in water. Finally, we go into detail in analysing further contributions to the total x-ray scattering signal from solvent shells, before presenting a best-practices recipe for future forward modeling.

The methods described and used in this work have been implemented in Python and is available from the The Python Package Index [36]. The data, plots, and code to reproduce them can be found in an online data-repository [37].

Background

It is well known that the isotropic x-ray scattering can be simulated from radial distribution functions (RDFs) which describe the ratio of local probability density to the bulk density, $g(r) = \rho(r)/\rho$ [38–40]. For a pair of 'atom types' l, m - elements, or atoms which scatter in the same way - the pairwise RDF, $g_{lm}(r)$ can be sampled from molecular dynamics (MD) trajectories of $i \in l$ and

$j \in m$ atom positions, sampled in spherical shells along r from the i'th particle to every j'th particle, $\delta(|\mathbf{r}_i - \mathbf{r}_j| - r)$:

$$g_{lm}(r) = \frac{1}{\rho_m} \left\langle \sum_{j \neq i}^{N_m} \delta(|\mathbf{r}_i - \mathbf{r}_j| - r) \right\rangle_i = \lim_{dr \to 0} \frac{N_{lm}(r)}{\rho_m 4\pi r^2 dr} = \frac{\rho(r)}{\rho_m}, \tag{1}$$

where $\rho_m = N_m/V$ is the average density of m-type particles in a simulation cell of volume V, and $N_{lm}(r)$ is the number of m-type atoms found at distance $r + dr$ from atoms of type l, most often averaged over a set of MD frames. As we want to separate the various contributions to the total scattering signal, we introduce the indices w_l, w_m, since for a solute in a solvent, the atom types can be further classified as belonging to either the sol*u*te, $w_l = u$, or the sol*v*ent: $w_l = v$. Then, the total scattered x-ray intensity *I(q)* as a function of the scattering vector q from a solute in a solvent can be calculated as the double sum of the individual terms (the form of which we shall return to in eq. 3):

$$I(q) = \sum_{w_l}^{\{u,v\}} \sum_{w_m}^{\{u,v\}} I_{w_l w_m}(q) = I_{uu}(q) + I_{uv}(q) + I_{vu}(q) + I_{vv}(q). \tag{2}$$

As discussed previously [29], the diagonal terms constitute solute-solute scattering $I_{uu}(q)$, which will not be addressed further in this work, and solvent-solvent scattering, $I_{vv}(q)$. The off-diagonal, or "cross" terms constitute the scattering terms from the solute-solvent correlations, $I_{uv}(q) + I_{vu}(q) = I_c(q)$. These terms are sometimes also called "solvent cage" scattering, which is misleading, as the total scattering signal from the solvent cage *also* includes scattering from solvent-solvent correlations in e.g. a denser solvent shell around a solute. The $I_{w_l w_m}(q)$ terms can be calculated from the RDFs via [29, 40]:

$$I_{w_l w_m}(q) = \delta_{w_l w_m} \sum_{l}^{w_l} N_l f_l(q)^2 + \sum_{l}^{w_l} \sum_{m}^{w_m} f_l(q) f_m(q) \frac{N_l(N_m - \delta_{lm})}{V} 4\pi \int_0^\infty (g_{lm}(r) - g^0_{w_l w_m}) \frac{\sin(qr)}{qr} r^2 dr, \tag{3}$$

where $f_l(q), f_m(q)$ are the atomic form factors that can be parameterized within the Independent Atom Model [41] and N_l (N_m) the number of atoms of type l (m) in the simulation cell. $g^0_{w_l w_m} = 0$ when $w_l = u$ and $w_m = u$, and 1 otherwise [29]. The equation is split up in two terms: the first term is the self-scattering and the second term contains the interatomic structure. The Kronecker delta, $\delta_{w_l w_m}$, assures there is no double-counting of self-scattering terms.

For this work, the two relevant contributions are the solvent and cross terms, $I_{vv}(q)$ and $I_c(q)$, i.e. with $g^0_{w_l w_m} = 1$. The fundamental measurable quantity for these two terms, agnostic to the nature of the scatterer, is the structure factor $S_{lm}(q)$, which for uniform, isotropic systems comprised of particles of type m and l (with at least one type within the solvent) can be defined as [42]:

$$S_{lm}(q) = 1 + \rho_m 4\pi \int_0^\infty (g_{lm}(r) - 1) \frac{\sin(qr)}{qr} r^2 dr. \tag{4}$$

We can express the x-ray scattering signal as a function of structure factors for the different pairs of atom types in our simulation. Keeping in mind that $N_l(N_m - \delta_{lm}) = (N_l - \delta_{lm})N_m$, we can rewrite eq. 3 as:

$$I_{w_l w_m}(q) = \delta_{w_l w_m} \sum_{l}^{w_l} N_l f_l(q)^2 + \sum_{l}^{w_l} \sum_{m}^{w_m} f_l(q) f_m(q) (N_l - \delta_{lm})(S_{lm}(q) - 1) \tag{5}$$

Note that with the definition of the structure factor in eq. 4, this expression is only valid for the cross- and solvent-solvent terms. Lastly, while $g_{lm}(r) = g_{ml}(r)$, the same is not true for $S_{lm}(q)$ and $S_{ml}(q)$, as N_m and N_l should also switch place, and thus we have not reduced the double sum further. In the following, we will use the shorthand index c to label the total cross-term scattering and structure factors, and v for the solvent-solvent term. Note that since the first term in eq. 3 is zero for $I_c(q)$, this term can be negative.

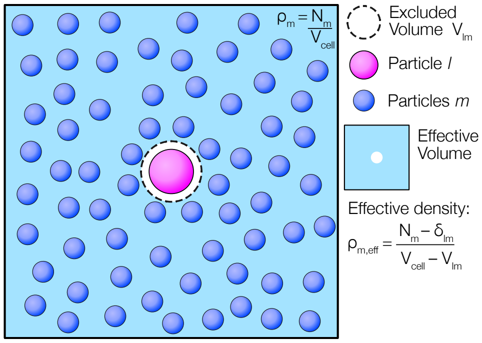

Figure 1: Illustration of the central concepts of volume and densities in a finite (periodic) system with volume V_{cell}. Particle l takes up some space V_{lm} in the simulation cell, where the other particles thus cannot exist. Note that this volume is *also* dependent on the extension of the particle of type m. When scaling the RDF to ρ_m, the excluded volume of particle l, V_{lm}, is included in V_{cell}, even though it is not available to the m-type particles. Instead, the "effective volume", i.e. the total volume of the simulation cell minus the excluded volume of particle l should be used to obtain an "effective" density $\rho_{m,\text{eff}}$ with which $\rho(r)$ in eq. 1 should be normalized.

If we let $q \to 0$, the integral in eq. 4, except for the density ρ_m, is recognized as the Kirkwood-Buff Integral (KBI) [43]:

$$G_{lm} = 4\pi \int_0^\infty \left(g_{lm}(r) - 1\right) r^2 dr \tag{6}$$

which provides a link between atomic-scale interactions and thermodynamic properties of fluids. The experimentally measurable isothermal compressibility for a system of a single atom-type $l = m$, χ_T can be obtained from the structure factor at $q = 0$ (i.e. the KBI from eq. 6) [44, 45]:

$$S(0) = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} = \rho k_b T \chi_T, \tag{7}$$

for $\rho = \langle N \rangle / V$, temperature T, and where k_B is the Boltzmann constant. However, KBIs are only defined for infinite systems [43]. In infinite systems, the theoretical long-r limit of the RDF is 1. This means, that for the integrals in eqs. 3, 4, and 6, when the RDF has converged to this limit at some distance $r = R$, we can safely truncate the integral at R, since the integrand above this R will be 0. However, the limit of the RDF when sampled in a finite-size system of N particles, even if periodic, becomes different from 1. As an example of this, statistical mechanics show that for an ideal gas system, $g(r) = 1 - (1/N)$ [45, 46]. The deviation from 1 arises because the analysis leading to deviation from 1 does not take into account the following: A *real*, finite system contains a fixed number of particles with finite volumes. Thus, the probability of finding a particle in the volume element $d\mathbf{r}_j$, given that another particle is in the element $d\mathbf{r}_i$ *must* take into account the (excluded) volume taken up by particle i, since the two particles cannot occupy the same volume (see fig. 1). As the system size increases, the excluded:total volume ratio decreases, and the limit will approach the infinite-size limit of 1.

This means that when we sample RDFs from finite-system molecular dynamics simulations, we must correct the limit. From $\rho_{m,\mathrm{eff}}$ in fig. 1, we see that there can only be two overarching strategies for doing so:

1. by correcting the effective *volume*, or
2. by correcting the effective *number of particles*.

Since this is a fundamental problem of RDFs sampled in finite systems, the usage of KBIs is also hindered by it. In the last two decades, much work has gone into developing corrections for the RDF that recovers the correct KBI [47–62]. In the following, we will study and test the most prominent from an x-ray scattering perspective. MD simulations and subsequent RDF sampling are often carried out in a "black box"-fashion as just one of many steps in a complex structural modeling strategy. Thus, we will limit ourselves to corrections that can be applied to RDFs sampled from MD simulations carried out in readily available and highly efficient MD codes, and RDFs sampled from already available codes, where no (cost-increasing or otherwise) modifications have to be implemented in either methodology, in order to apply the correction.

Volume Corrections

For RDFs sampled in finite systems, $g_{lm}^{N}(r)$, the RDF in eq. 1, the RDF is normalized to the average density of particles m, $\rho_m = N_m/V_{\mathrm{cell}}$. We must therefore rescale the RDF with the difference between the cell volume and the effectively accessible volume to recover the RDF as if sampled from an infinite system, $g_{lm}^{\infty}(r)$:

$$g_{lm}^{\infty}(r) = \frac{\rho_m}{\rho_{\mathrm{eff}}} g_{lm}^{N}(r) = g_{lm}^{N}(r)\rho_m \frac{V_{\mathrm{cell}} - V_{lm}}{N_m - \delta_{lm}}, \tag{8}$$

where V_{lm} is the excluded volume, i.e. the volume taken up by particle of type l in which particles m cannot be found. This volume is nontrivial to obtain exactly, but as a first-order approximation, we assume it to be spherical, with radius R_{lm}: $V_{lm} = (4/3)\pi R_{lm}^3$. Note that V_{lm} is the volume inaccessible to the remaining particles, and not just the total volume divided by the total number of particles (which would be larger than V_{lm}).

Particle Number Corrections: Ganguly & van der Vegt

An alternative route is presented by P. Ganguly and N. F. van der Vegt, developed for correcting KBIs. The authors state that "excess (depletion) of particles of type m around particles of type l at a local scale is compensated by depletion (excess) of particles of type m at long distances, as the total number of particles is fixed." [57] Based on this observation, the authors propose to correct the error in the ρ-normalization of the RDF by correcting the number N of particles throughout r:

$$\begin{aligned} g_{\mathrm{lm}}^{\infty}(r) &= g_{\mathrm{lm}}^{N}(r) \frac{N_m}{N_m - \left[(\Delta N_{\mathrm{lm}}(r) + \delta_{lm}) \left(1 - \frac{(4/3)\pi r^3}{V_{\mathrm{cell}}}\right)^{-1} \right]} \\ &= g_{\mathrm{lm}}^{N}(r) \frac{N_m \left(1 - \frac{(4/3)\pi r^3}{V_{\mathrm{cell}}}\right)}{N_m \left(1 - \frac{(4/3)\pi r^3}{V_{\mathrm{cell}}}\right) - (\Delta N_{lm}(r) + \delta_{lm})}, \end{aligned} \tag{9}$$

where N_l is the number of particles of type l in the system, V is its volume, and δ_{lm} the Kronecker delta. The parameter $\Delta N_{lm}(r)$ is defined as "the excess number of particles of type m within a sphere of radius r around particle of type l". Usually [55, 56, 60] one estimates this quantity via:

$$\Delta N_{lm}(r') = 4\pi\rho_m \int_0^{r'} (g_{lm}^N(r) - 1) r^2 dr, \tag{10}$$

for r'-values less than half the simulation box length.[1]

Another N-focused correction uses a Taylor expansion of the difference between the finite- and infinite-system RDF long-r behaviour [52], using an unknown function to describe the deviation, which is estimated from two simulations with differing number of molecules [56, 59]. This method thus requires twice the amount of simulations, and relies on the difference between RDFs sampled from almost identical systems, making it very sensitive to sampling noise. Thus, we will not be employing this method here.

Heuristic Corrections: Perera *et al.*

Lastly, Perera *et al* propose a more heuristic correction to the RDF between particles of type l and m [64]:

$$g_{lm}^{\infty}(r) = g_{lm}^N(r)\left[1 + \frac{1 - g_{lm}^{N,0}}{2}\left(1 + \tanh\left(\frac{r - \kappa_{lm}}{\alpha_{lm}}\right)\right)\right]. \tag{11}$$

The difference between the correct asymptotic limit, 1, and the sampled, $g_{lm}^{N,0}$, $1 - g_{lm}^{N,0}$ is used in a smooth switching function, where α_{lm} controls the smoothness of the switch from $g_{lm}^{N,0}$ to 1. κ_{lm} represents an effective diameter with which particle of type l and m displace each other, as it is taken to be twice the distance at which $g_{lm}(r)$ starts to be different from zero. This RDF-correction grows in at larger r, and leaves the local structure untouched. However, from the definition of the RDF in eq. 1, the $1/\rho_m$ normalization should be applied over the entire r-range. This correction is more akin to the often applied "damping" or "windowing" functions used ubiquitously in forward modeling as well as in inversion methods [8, 9, 16, 26, 65–67], which we will address in the next section.

Damping functions

As we shall see in the following sections, none of the above physics-based corrections are guaranteed to make the RDF go *exactly* to 1 at long r, we thus cannot be certain that the integral in eq. 3 (or in eq. 4) can be truncated safely. Therefore, in practice, we must often employ a window function (or 'damping function') to the integral in eq. 3, $\omega(r)$, which is not grounded in physical considerations about the RDF. One such window function is used by Lorch [68]:

$$\omega_{\text{Lorch}}(r) = \frac{\sin(\pi r/R)}{\pi r/R}, \tag{12}$$

where R is the largest r-value of the sampled RDF (most often half of the simulation cell sides). However, to avoid over-correction when combined with any of the previously mentioned RDF-corrections, one should take care that the applied damping function does as little as possible to

---

[1] Some programs, such as VMD [63] uses an average density in the normalization of the RDF that does not count the particle from which the histogramming starts, in cases of $l = m$, i.e. $\rho_m = (N_m - \delta_{lm})/V$. If such a normalization is used, then the Kronecker delta should be omitted from eq. 9.

change the structure encoded in the RDF. Dhabal and coworkers compared the correlation functions of water simulated using the TIP4P/2005 potential with experimental results using the following windowing function:

$$\omega_{\mathrm{Dhabal}}(r) = \begin{cases} 1 - 3(r/R)^2, & \text{if } r < R/3 \\ 3/2\left(1 - 2(r/R) + (r/R)^2\right), & \text{if } R/3 < r < R \\ 0, & \text{if } r > R \end{cases} \tag{13}$$

which was later modified into an smooth step function from R_{cut} to $R_{\mathrm{cut}} + R$, which leaves the RDF completely untouched before R_{cut} [5]:

$$\omega_{\mathrm{ZK}}(r) = \begin{cases} 1, & \text{if } r < R_{\mathrm{cut}} \\ 1 - 3\left(\frac{r - R_{\mathrm{cut}}}{R - R_{\mathrm{cut}}}\right)^2, & \text{if } R_{\mathrm{cut}} < r \leq \frac{2R_{\mathrm{cut}} + R}{3} \\ \frac{3}{2}\left(1 - \frac{r - R_{\mathrm{cut}}}{R - R_{\mathrm{cut}}}\right)^2, & \text{if } R > r > \frac{2R_{\mathrm{cut}} + R}{3} \\ 0, & \text{if } r > R, \end{cases} \tag{14}$$

Which reduces to $\omega_{\mathrm{Dhabal}}(r)$ for $R_{\mathrm{cut}} = 0$. For large cells, the Lorch-like damping might be most appropriate, as a very slow change in the RDF is less aggressive than a more abrupt cut provided by a smooth step function, but in other cases, where e.g. smaller cells have been sampled, it might in some cases be more important to avoid damping the local structure encoded at shorter r values in the RDF, than to enforce a smoothly changing damping window. See fig. 2 for a comparison of the shapes of the functions.

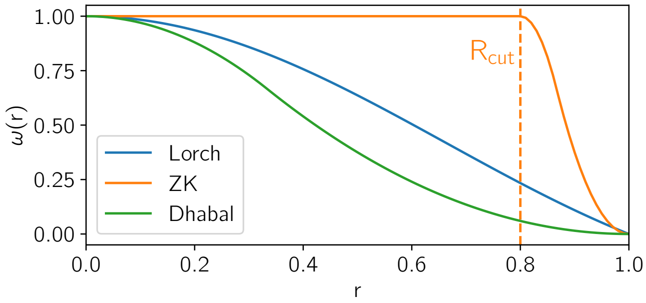

Figure 2: Illustration of the shapes of the three window functions described in the main text, with $R = 1$, and in the case of the ZK-damping, $R_{\mathrm{cut}} = 0.8$

Testing the Corrections

Based on the considerations in the previous sections, we can devise a useful test of the various corrections by evaluating

$$S_{lm}(q; R) = 1 + \rho_m 4\pi \int_0^R \omega(r)\left(g_{lm}^{\infty}(r) - 1\right)\frac{\sin(qr)}{qr} r^2 dr \tag{15}$$

at $q = 0$ Å^{-1}, for increasing values of R, since a successful recovery of $g_{lm}^{\infty}(r)$ from $g_{lm}^{N}(r)$ will make this integral converge to a fixed value, as previously discussed.

The test systems in the following are chosen to gradually increase in complexity, starting from a neat LJ liquid, a solute-solvent LJ liquid system, then switching to liquid water before finally ending up with solute-solvent water systems. For the systems with many l, m-combinations, we will use eq. 5 at $q = 0$ instead.

Computational Details

The test systems were chosen to gradually increase in complexity, starting from a neat LJ liquid, a solute-solvent LJ liquid system, then switching to liquid water before finally ending up with solute-solvent water systems.

All MD simulations were carried out in OpenMM [69]. The LJ simulations were set up using OpenMMTools [70], placing 20713 particles in a (100 x 100 x 100) Å cell with the LJ-parameters $\sigma_{\text{solvent}} = 3.4$ Å and $\epsilon = 0.238$ kcal/mol. The σ_{solute} values were manually changed before starting each separate simulation. All simulations were equilibrated for 1 ns each in the NVT ensemble using a Langevin thermostat with a 2 fs timestep at 94.4 K. Each production run was carried out with the same thermostat for 100 ns, as recommended elsewhere [57], saving all particle positions every 0.5 ps. The RDFs were sampled using VMD with a numerical dr of 0.05 Å.

The neat water system was created using Packmol [71] to place 4095 water molecules in a (50 x 50 x 50) Å cell, and and AmberTools22 [72] to generate the topology for the TIP4P$_{\text{EW}}$ potential [73]. The simulation was equilibrated for 1 ns, and sampled for 99 ns, both in the NVT ensemble at 300 K, using a Langevin thermostat with a 2 fs timestep. Then two simulations were spawned from the end of the first: one in which the cell side lengths were reduced each by 0.2 Å, and the other where they were increased by the same amount. The new simulations were again equilibrated for 1 ns, and sampled for 99 ns. The process was then repeated for further reductions and expansions of the cell. The RDFs were sampled using VMD with a numerical dr of 0.01 Å.

The first water-based solute-solvent system was chosen to be a single Ag^+ in water, but could have been any single-atom ion. The Ag^+-Water cells were created with AmberTools22 [72], placing a single ion in 17535 water molecules, using the non-bonded Ag^+ parameters optimized for TIP4P$_{\text{EW}}$ from Li & Merz [74]. Since the solvent-cell routine in AmberTools creates water with a very low density, the system was equilibrated for 2 ns in the NPT ensemble at 1 bar, maintained by a Monte Carlo barostat, and the system was propagated in 2 fs timesteps at 300 K by the Langevin integrator. The final cell dimensions for the production runs were (85.8 x 85.8 x 85.8) Å. The RDFs were sampled using VMD with a numerical dr of 0.05 Å.

For the final solute-solvent system, we chose the $[Fe(bpy)_3]^{2+}$ (bpy = 2,2'-bipyridine) complex in water, as it has been central to many previous x-ray solution scattering experiments [21,25,75] and theoretical studies of solute-solvent interactions [22,24,76]. The complex was parameterized within the General AMBER (GAFF) potential, using the 'MCPB.py' method, as described elsewhere [77]. The complex was then solvated in a box of 7329 water molecules modelled with the TIP4P$_{\text{EW}}$ potential [73], using the 'leap' program from the 'Ambertools' toolbox [72]. Two Cl^- counterions were also added using leap, to charge-neutralize the system. The complex was restrained in the center of the box, and the counterions were restrained roughly 22 Å away from the complex using positional restraints with 500 kcal/mol force constants. The NPT equilibration resulted in a final cell size of (58.7 x 61.1 x 61.4) Å. The density and temperature were equilibrated using OpenMM for 500 ps in the NPT ensemble, using a Monte Carlo Thermostat to keep the pressure at 1 bar, and a Langevin propagator to keep the temperature at 300 K [69]. As with the previous set of simulations the simulation cell size was changed to sample each global density in the NVT ensemble for 20 ns.

Results and Discussion

Test System 1: The Lennard-Jones Liquid

This section analyses the effects on calculating the cross-term structure factor from finite-system sampled RDFs of Lennard-Jones (LJ) liquids, and tests the ability of the chosen corrections to eliminate finite-size artefacts. The liquid structure of LJ liquid is governed by the well-known potential between particles i and j:

$$E(r_{ij}) = 4\epsilon_{\mathrm{ij}} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{16}$$

where σ_{ij} determines the distance at which the potential becomes repulsive, and ϵ the depth of the minimum. Each atom type (element) are assigned their own σ and ϵ parameters, and are combined via simple geometric combination rules, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, also called the Lorentz-Berthelot combination rules [78, 79]. By employing the LJ potential, we prioritize computational conveniences such as differentiability over a completely well-defined excluded volume, meaning that pressure and density will affect the resulting excluded volume.

First, we simulate a 'neat' LJ liquid where all particles are assigned $\sigma = 3.4$ Å, sample the RDFs, and calculate the structure factor using eq. 15 at $q = 0$ Å^{-1} for gradually increasing truncations of the integral, R, $S_{lm}(0; R)$, towards convergence if the correction is successful. This type of plot is thus a helpful tool in judging whether the low-q behaviour of the structure factor (and thus the scattering signal) will be influenced by finite-size sampling errors of the RDF. Since $l \in v, m \in v$, we label this structure factor $S_v(0; R)$. The top plot in figure 3 shows the LJ-potential (red line) used in sampling the RDF shown in black. The middle plot in the figure shows the structure factor $S_v(0; R)$, using the 'raw' RDF in the top plot, before applying any RDF-corrections (black dashed line). The plot also shows structure factors calculated from RDFs corrected with the three main methods described previously (solid lines). The uncorrected $S_v(0; R)$-value starts a smooth increase from around roughly 40 Å, while the structural correlations die out. This indicates that the RDF is converging towards a value slightly above 1 such that the value of the integral truncated at R is increasing proportionally to r^3, and the low-q structure factor intensity will be entirely dependent on which cell size was chosen for the MD cell, and/or when one chooses to stop sampling the RDF. For the Perera-correction, k_{lm} is simply set to the first r-value that gives $g_{lm}(r) > 0$, 2.925 Å, as suggested in the paper [64]. Due to the finite bin-size of 0.05 Å in sampling the RDF, as well as the fact that an LJ particle is not completely hard, this value is of course associated with some inaccuracy. For the volume correction, R_{lm} was set to 2.8240 Å using a simple fitting-method to find the excluded volume which we will describe in the next section on solute-solvent LJ systems. The volume- and van der Vegt correction both flatten out $S_v(0; R)$ at R-values roughly above 40 Å to slightly oscillate around their average value in this region. The Perera-correction seems to have slightly overcorrected the RDF such that the $S_v(0; R)$-value starts to systematically decrease at long R-values. This small overcorrection can most likely be eliminated by further tuning the involved parameters.

The volume- and van der Vegt correction both makes the $S_v(0; R)$-value converge to a fixed value, which is the expected physical behaviour from eq. 7. However, in practice, the small oscillations still present even at these large distances will cause an unwanted truncation-distance dependence

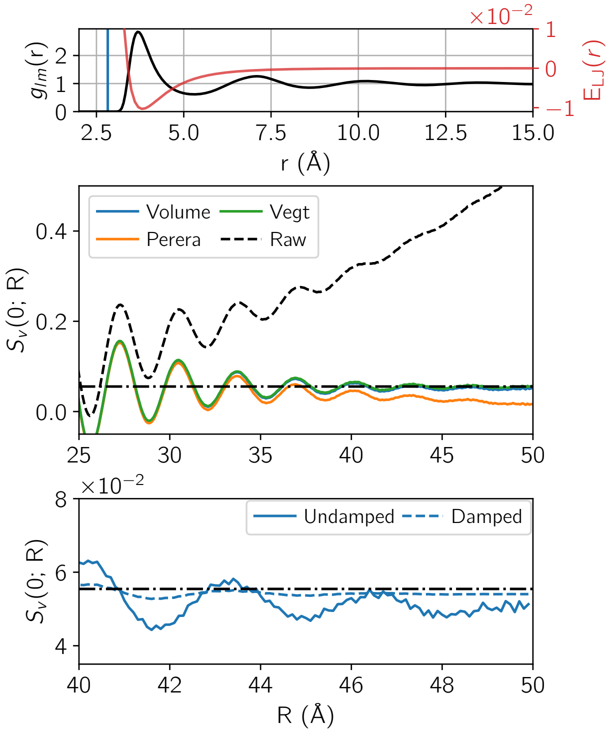

Figure 3: Top: The RDF of a neat Argon LJ liquid with $\sigma = 3.4$ Å. The RDF is sampled for 100 ns in a 100 Å box. The red curve shows the energy from the corresponding LJ potential. Middle: $S_v(0; R)$ for the corrections presented in the previous section, calculated without damping ($\omega(r) = 1$), as well as from the uncorrected ('raw') RDF. The R represents the truncation length of the integral in eq. 15. The dash-dotted line shows the average $S_v(0; R)$-value for $r > 40$ Å of the van der Vegt-corrected result. Bottom: Zooming in on the $R > 40$ Å region, and comparing two Volume-corrected results, the same, undamped calculation as found in the main bottom plot, and a calculation that also employed the Lorch-like damping with $L = 50$ Å.

of the resulting predicted scattering signal. The bottom plot in the fig. 3 shows how employing the Lorch-like damping window converges the integral to a constant value such that it can be safely truncated. This procedure will be used throughout the rest of this work.

Fig. 4 shows the low-q behaviour of the structure factor, with and without corrections, as well as the entire signal up to $q = 4$ Å^{-}1 in the inset. The uncorrected structure factor artificially increases as q goes toward zero, consistent with the $S_v(0; R)$ from fig. 3. The black dashed curves with gradually increasing opacity represents results from integral-truncation distances ranging from 30 Å to 50 Å, showing how the low-q amplitude is entirely dependent on the (arbitrarily) chosen truncation distance. As the damping functions also effectively decrease the area under the RDF-curves, only applying damping (black, dashed curve) has some degree the same effect on the resulting structure factor as applying any of the RDF-corrections, but they fail to fully correct the signal, and have no grounding in the root physical problem behind sampling RDFs from finite-size systems. By correcting the RDFs for their finite-sampling, these low-q artefacts can be removed. The van der Vegt- and the volume-corrections are in closest agreement on the $q < 0.2$ Å^{-1} signal shape and magnitude, whereas the Perera-correction produces a slightly smaller signal in this region. For these large 100 Å cells, the artificial increase in the uncorrected signal sets in around $2\pi/(50$ Å$)$

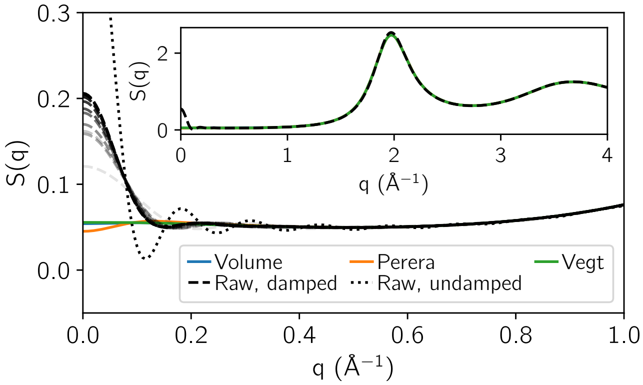

Figure 4: The performance of the three corrections on the low-q behaviour of the structure factor. The structure factor calculated from the uncorrected RDF artificially increases as q goes toward zero, as a result of the finite-size RDF-sampling. The dashed black curves increase in opacity according to R values going from 30 Å to 50 Å.

$= 0.125$ Å^{-1}, with some ringing (i.e. truncation oscillations) present up until $\sim$0.2 Å^{-1}. However, not all types of simulation are inexpensive enough to allow cells of these sizes to be employed, making corrections crucial, even at higher q. Especially *ab initio* molecular dynamics (AIMD) simulations, where the particles are classically propagated from gradients of *ab initio* potentials [21, 22, 24], are still too expensive to perform using such large cells. Multiscale methods that couple Quantum Mechanical / Molecular Mechanical (QM/MM) potentials can employ larger (MM) cells [3,10,80,81], but not without increasing the coupling cost in evaluating the interactions between the two subsystems, and introducing approximations in how the coupling is carried out [82].

Solute-Solvent LJ systems

We now extend the analysis by systematically increasing the σ-value of a single 'solute' particle in a series of separate MD simulations, to simulate how an increase of the effective particle-size affects the resulting estimated excluded volume, RDF convergence, and corrections. To find the excluded volume for an l-type particle in an m-type solvent, we fitted $V_{lm} = 4/3\pi R_{lm}^3$ of the volume-correction in eq. 8 by optimizing R_{lm}^3 to produce an RDF that minimizes the residual between $g_{lm}(r > r_{\max})$ and 1, where the the parameter $r_{\max}$ was chosen such that correlations within the liquid have all but died out. This will circumvent the limit imposed by the finite bin-size when numerically sampling the RDF, but the fit will be sensitive to the numerical sampling noise, which even for 100 ns trajectories is relatively large compared to the difference between the actual asymptotic limit of the RDF and 1. The first 5 rows of fig. 5 again uses the $S(0; R)$ method to analyse the results, this time labeled $S_c(0; R)$, since $l \in u$ and $m \in v$. Since the differences between the RDF value and 1 are so small, the sampling noise makes the fit sensitive to the choice of $r_{\max}$, but most often 7σ proved adequate in fitting volumes that converges $S_c(0; R)$. The bottom plot shows a linear increase in the excluded volume with increasing σ, with two distinct slopes depending on whether $\sigma_{\text{solute}} > \sigma_{\text{solvent}}$ or vice versa: When the solute is smaller than the solvent, the extension of the neighbouring solvent molecules play a larger role in how close they can pack around the solute, as they also sterically hinder each other, showing the m-dependence on V_{lm}.

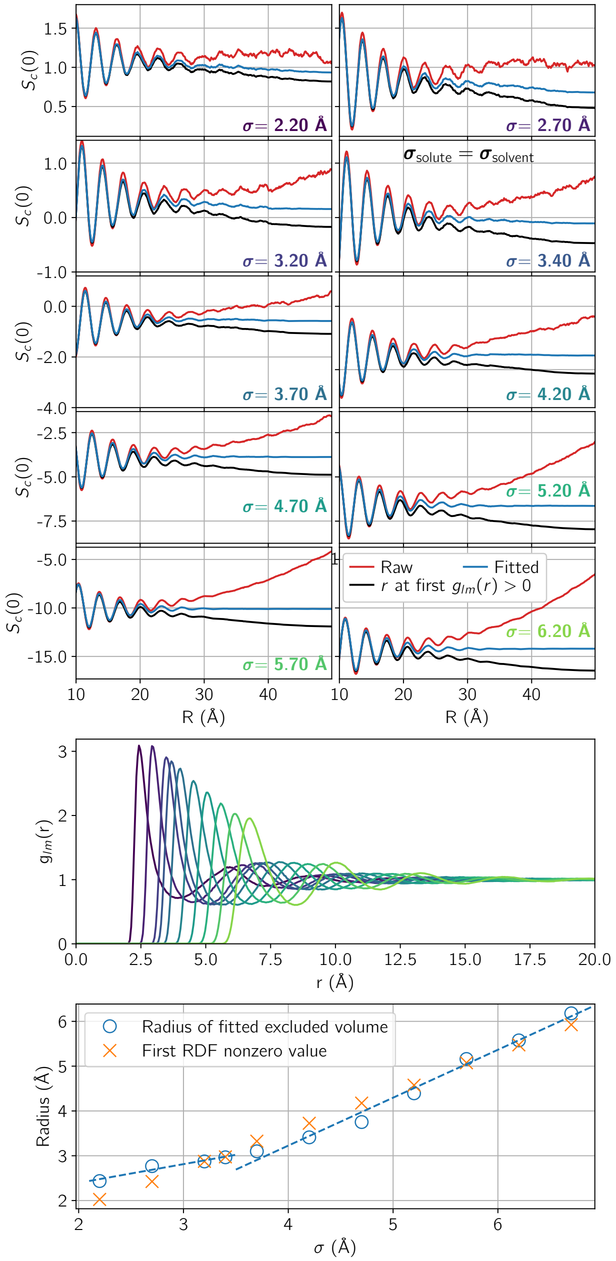

Figure 5: Top 5 rows: Comparing the volume-correction for $S_c(0; R)$ where the excluded volume is either determined via the first non-zero RDF-r-value (black), or via the fitting procedure described in the main text (blue), for a selection of the entire set of simulations. Note that we only calculate the cross-term. The actual measured signal would also contain both particle self-scattering and all the other terms apart from the term shown here. Middle: Solute-solvent RDFs used in the plots above. Bottom: Fitted radii of the excluded volume V_{lm} as a function of solute-solvent σ values (blue circles). The blue dashed lines show linear fits to the two regions $\sigma_{\text{solute}} < \sigma_{\text{solvent}}$ and $\sigma_{\text{solute}} > \sigma_{\text{solvent}}$ (with $\sigma_{\text{solvent}} = 3.4$ Å).

Test System 2: Water

We now focus on neat water as a test system. As shown in eq. 7, the isothermal compressibility can be obtained from $S(q = 0)$. Therefore, we can calculate the isothermal compressibility χ_T from

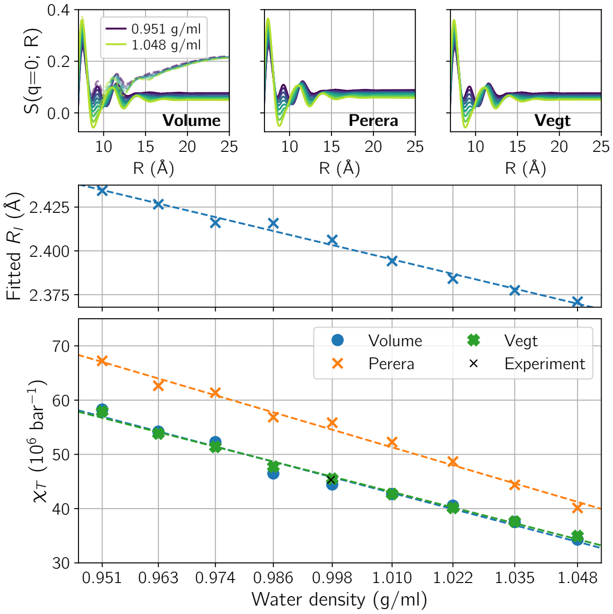

Figure 6: Top: The three corrections of the $S(0; R)$-value, for each density of 4095 molecules of TIP4P$_{\mathrm{EW}}$ water sampled for 100 ns each, in cells of decreasing size. The leftmost plot also shows the uncorrected curves as dashed lines. Middle: The fitted radius of the excluded volume as a function of density, where the fit minimizes the residual of $g(r > r_{\max}) - 1$, where an $r_{\max}$-value of 22 Å was used. Bottom: The isothermal compressibility χ_T as a function of the water density.

the simulations and compare to its experimental value. Fig. 6 collects the results from simulating neat water at a range of densities.

As the main contribution to the scattering will come from the O-O correlation, we calculate the $q = 0$ structure factor of this atom-type pair, and correct it with the three previous corrections, again fitting the excluded volume as explained in the previous section. The middle plot shows how the excluded volume decreases as expected with increasing the density. The lowest plot shows how this translates into a linearly decreasing isothermal compressibility, which can again be heuristically understood as the more we have already compacted the water, the less it is willing to be further compacted. As we are only using the O-O pair, and using the relatively simple TIP4P$_{\mathrm{EW}}$ water model that e.g. employs rigid water molecules, we cannot expect a complete agreement with the experimental value, yet both the van der Vegt- and volume correction produces compressibilities very close to the experimental value of ambient water density at room temperature [83, 84]. This shows that the corrections not only make the integral converge at some constant value, but a value as close to the experimental one as can be expected from the MM potential, whereas the Perera correction seems to slightly overestimate it. This does not change whether we set κ_{lm} from eq. 11 to the value found using our excluded-volume fitting method, or the first non-zero $g_{lm}(r)$ r-value, but we cannot rule out that further adjustments would increase the accuracy with respect to the experimental result. However, our result is consistent with the previous finding that the van der Vegt- and Perera- methods give different KBIs at long distances [57], most likely since the Perera-correction leaves the RDFs unchanged at shorter r-values, even though the sampled RDF should be corrected across all r.

Test System 3: A Single-atom Ion in Water - Sampling Convergence

In the previous section we took advantage of the fact that there are $N(N-1)$ water-water correlations in the simulation cell, which greatly increases the sampling statistics. However, many scenarios involve a single solute molecule in solution, where the cross term RDFs are then sampled on only N correlations per independent MD frame. Figure 7 tests the three corrections on a simple solute-solvent system comprised of a single Ag^+ ion in water, with RDFs sampled from between 1 ns to 133 ns. If the sampling is too short, the corrections cannot converge the integral. The $S_c(0; R > 39\text{Å})$ values converge with simulation time when either the Volume- and Perera correction is applied, whereas the van der Vegt correction is more unstable. For the 1 ns sampling, the bottom plot shows large artificial oscillations at low q with visible ringing above 1 Å^{-1}. The dashed curve represents scattering calculated from a $dr = 0.2$ Å-sampled RDF, showing that a 4 x increase in numerical bin size of dr in eq. 1 only affects the oscillatory artefacts significantly at $q > 0.2$ Å^{-1}.

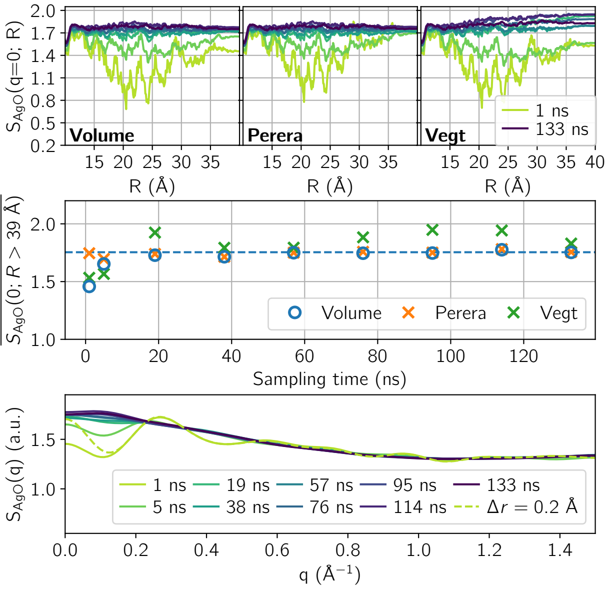

Figure 7: Top row: Performance of the volume, Perera, and van der Vegt corrections on the Ag^+-O RDF, with respect to sampling time. Middle: the average $S_c(0; R > 39\text{Å})$ as a function of sampling time. The Perera correction converges fastest, but all corrections reach a stable plateau if sampled more than roughly 20 ns. The dashed blue line shows the total average of all $S_c(0; R > 39\text{Å})$ averages for sampling times of 60 ns or more. Bottom: The structure factor at low q as a function of sampling time, calculated from RDFs corrected using the Volume correction. The dashed line shows a the result of reducing the sampling noise of the 1 ns sample by calculating the scattering using a numerical dr of 0.2 Å.

Real systems: Simulation cell sizes

Real systems, like a molecule in solution, will most often be comprised of many different elements. This section will use the transition metal complex $[\mathrm{Fe(bpy)_3}]^{2+}$ as an example, as it has been the center for many experimental studies using x-ray techniques [75, 85, 86], including scattering [18, 21, 25], as well as computational investigations [22, 24, 76]. The complex is shown in the lower

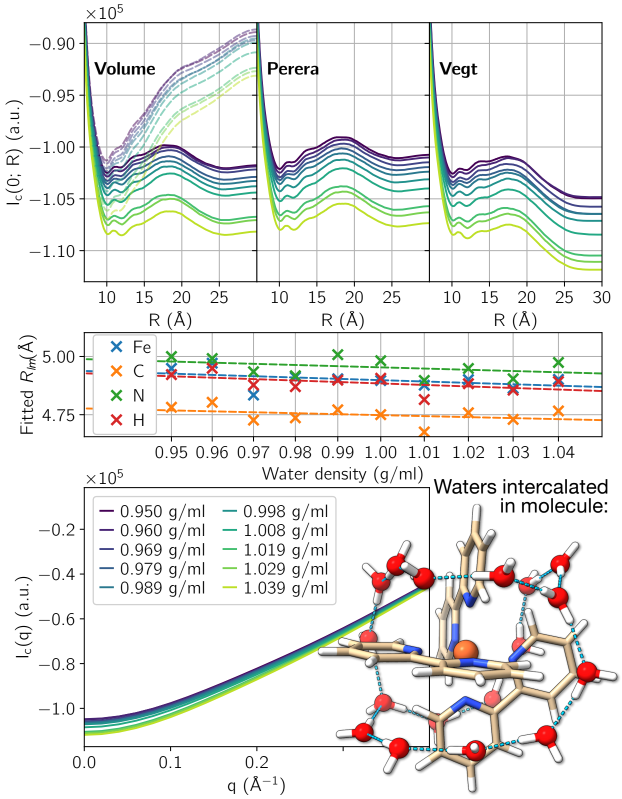

Figure 8: Top row: Performance of the three corrections tested on the total coherent cross-scattering term $I_c(q = 0)$ as a function of the integral truncation distance R, calculated from all solute-solvent pairwise RDFs of the $[\mathrm{Fe(bpy)_3}]^{2+}$ molecule in water. Middle: The fitted excluded volume of the different solute atom types as a function of water density. Bottom left: the cross-term at low q as a function of water density. Bottom right: Illustration of the (hydrogen bonded) network of intercalated water molecules in the first solvation shell of the molecule.

right corner of figure 8. From the geometry of the complex, it is evident that the simple spherical approximation of the excluded volume will be much less accurate than previously. The complex is comprised of a single Fe atom, plus atoms of the elements C, N, and H, and according to eq. 3 the coherent cross term scattering signal $I_c(q)$ is a stoichiometrically weighed sum of the integrals over the cross-term RDFs, each term multiplied with the atomic form factors of the atom types of the RDF. It is thus clear that each pairwise correlation does not contribute equally to the recorded signal, but that its intensity is dependent on the number of atoms of each type present in the system, as well as the number of electrons in each atom type. But $I_c(0; R)$ will show the same behaviour as $S_c(0; R)$, as the form of the integral is the same. Therefore, to make a single assessment of the convergences of *all* the integrals that are evaluated to obtain $I_c(q)$, taking into account how much each term contributes to the total cross-term signal, we use $I_c(0; R)$ instead of all the individual $S_c(0; R)$ terms. The top three plots in figure 8 shows the results of applying each of the three corrections to each of the cross-term RDFs. For the volume-correction, the excluded volume was fitted separately for each RDF. For this relatively large molecule, the first thing we observe is how the two first corrections struggle to make the running integral reach convergence within the limited box size. However, close inspection of the results from the van der Vegt correction shows that the curves seem to flatten out above ∼27 Å, indicating that convergence might be reached for this correction. We repeated the strategy of performing the simulations at increasing densities in order to study the robustness of the fitted excluded volume from the long-r RDF asymptote. The middle plot shows the result of these fits. We first note that the excluded-volume values are larger than expected for a single atom of the same type in a solution. This is because the rest of the particles in the solute will hinder access of the solvent molecules. Yet, in spite of the now somewhat crude spherical approximation of the excluded volume, we still observe the expected behaviour of the volume decreasing with increased density, no matter which of the solute-solvent RDFs we fit. While all pairs produce excluded volumes of similar sizes for each water density, the intercalation of water molecules between the ligands of the complex (see the inset in fig. 8) mean that the waters of the first solvation shell can get closer to especially the carbon atoms of the complex. This effect causes the smaller fitted excluded volume values from those correlations.

The solid curves in the bottom plot show the total cross-term scattering signal at low q-values, corrected using the van der Vegt correction. The signal is negative since there is no atomic self-scattering for this term in eq. 3, but while this term can easily be calculated, it evidently cannot be measured without *also* measuring the scattering from the rest of the terms, which make the total signal positive for all q.

We conclude this section by emphasizing that larger solute molecules create larger excluded volumes, which again require larger total simulation cell sizes if one is to robustly assess and correct the RDFs to calculate reliable signals in the low-q region. Overall, this section has given examples on how to use evaluation of $S(0;R)$ and $I(0;R)$ to assess how artefacts from finite-size sampling affects the low-q region, and how well one can correct for them. We thus put forward the strategy presented in the above analysis as an effective way to gauge the trustworthiness of simulated scattering signals and discern artifacts from physical information, in work where high accuracy in the low-q region is required. The strategy is summarized in the summary and outlook section of this work.

Solvent-Solvent Contributions to Scattering

Since our RDF-correction is based on an estimation of the excluded volumes in a finite simulation cell, it makes sense to further analyse the contributions of the excluded volume itself to the total scattering signal. Since the excluded volume will essentially constitute a scattering contrast between the solvent and the 'hole' left by a solute, we must analyse the solvent-solvent contributions to the total scattering signal. All tallied, the solute-solvent system leads to the follow scattering terms: (1) $I_{uv}(q)+I_{vu}(q) = I_{\mathrm{c}}(q)$ from the solute-solvent RDFs (see eqn 2), (2) The part of the solvent-solvent scattering term $I_{vv}(q) = I_v(q)$ that encodes the local structure in the shell around the solute, and (3) the scattering from the contrast between the excluded volume 'hole' and the shell- and bulk solvent. The second and third contribution are encoded in the solvent-solvent RDFs and thus $I_{\mathrm{v}}(q)$, but the subtle, diffuse structural differences are averaged out over the bulk solvent average pairwise particle distances, and are thus extremely sensitive to numerical sampling noise. Furthermore, for fixed-cell (NVT) simulations, studies of how changes in molecular solute structure leads to change in its solvation shell, a local expansion of the solute (and its excluded volume) will lead to an - albeit small - decrease in effective volume available to the solvent, resulting in a bulk-solvent density increase.

Small-Angle X-ray Scattering (SAXS) methods estimate the scattering from the excluded volume by constructing form factors from an average density of the surrounding solvent to model the scattering from it [87–90]. The same central principle has been employed elsewhere [26, 35] by using the cross-term RDFs to construct an 'excluded volume' form factor $f_{\mathrm{l}\in\mathrm{u,EV}}(q)$:

$$f_{\mathrm{l}\in\mathrm{u,EV}}(q) = 4\pi \sum_{m\in v} f_m(q) \frac{N_m}{V} \int (g_{lm}(r) - 1) \frac{\sin(qr)}{qr} r^2 dr, \tag{17}$$

which was then used in the discrete Debye-formulation on all solute distances [26], effectively using the solute-atomic positions to approximate the size and shape of the excluded volume. The solute in the original study of Panman et al. was comprised of particles of a single atom type, but as we observed in figure 8, when different atom types make up the solute, the excluded volume experienced by each atom type can be slightly different, thus requiring an $f_{\mathrm{l,EV}}(q)$ per atom type $l \in u$. Due to the equivalence of the discrete Debye equation and eq. 3 [29], we can generalize this procedure to obtain the excluded-volume scattering from solute-solute radial distribution functions:

$$I_{EV}(q) = \sum_{l\in u} N_l f_{l,\mathrm{EV}}(q)^2 + \sum_{l\in u} \sum_{m\in u} f_{l,\mathrm{EV}}(q) f_{m,\mathrm{EV}}(q) \frac{N_l(N_m - \delta_{lm})}{V} 4\pi \int_0^R g_{lm}(r) \frac{\sin(qr)}{qr} r^2 dr. \tag{18}$$

We can assess this method by comparing it to the direct solvent-solvent term $I_v(q)$ for our simple LJ liquid with varying solute sizes. We thus return to the LJ liquid system and calculate $I_v(q)$, as well as the scattered intensity using $f_{l,\mathrm{EV}}(q)$. As our system only contains a single particle, the excluded-volume intensity is simply the self-scattering term.

$$I_{\mathrm{EV}}(q) = \sum_{l\in u} N_l f_{\mathrm{l,EV}}(q)^2 = f_{\mathrm{l,EV}}(q)^2. \tag{19}$$

The contribution of the solvation shell to the total solvent-solvent term $I_v(q)$ will be completely overshadowed by the bulk solvent signal. Thus we first calculate the *difference*-scattering signal from a solute-expansion by subtracting the solvent-scattering of the 'neat' LJ liquid from that of the system with the biggest solute: $\Delta I_v(q) = I_v^{6.7}(q) - I_v^{3.4}(q)$, where both simulations were performed with identical box sizes.

The black curve in fig. 9 shows this calculated signal. The signal has two distinct features, a $q < 1$ Å^{-1} peak and a 1 Å$^{-1} < q < 3$ Å^{-1} oscillation, which we call the 'low-q' and 'high-q' feature, respectively. As mentioned in the introduction to this section, there must be two contributions to this signal: (1) The increase in the excluded volume inaccessible to the solvent due to the increase in solute size, and (2) the increase in average density of the bulk solvent as the effectively accessible volume to the solvent decreases as the excluded volume of the solute *increases*. The second term can be isolated by simply running a simulation of 'neat' $\sigma_{ij} = 3.4$ Å-particles in a *smaller* box, where we have reduced the volume with the same amount as the total accessible volume has decreased for $\sigma = 3.4$ Å-particles in the original box, due to the single $\sigma_{ij} = 6.7$ Å-solute. We estimate this volume by using the fitted radius of the excluded volume in figure 5. For 100 Å box sides, this change constitutes a reduction of the total volume by less than 0.1 %. Yet, this minute change gives the difference-scattering signal $\Delta I_\rho(q)$, displayed with blue in fig. 9. There is noticeable agreement between the high-q solvent-solvent feature and the global average density increase, indicating that this feature arises from the global density increase of the bulk solvent. The differences between the black and blue curve in the high-q feature must arise from the the inaccuracies in exactly determining the excluded volume from the simple model discussed previously (see fig. 5 and accompanying text), as well as the fact that the signal from the global density increase does not contain any local solvent shell structure.

The low-q feature is not present in the global density term $\Delta I_\rho(q)$, but is also present in $\Delta I_{\mathrm{EV}}(q) = I_{\mathrm{EV}}^{6.7}(q) - I_{\mathrm{EV}}^{3.4}(q)$. Thus, the excluded-volume scattering expression in eq. 18 describes the local

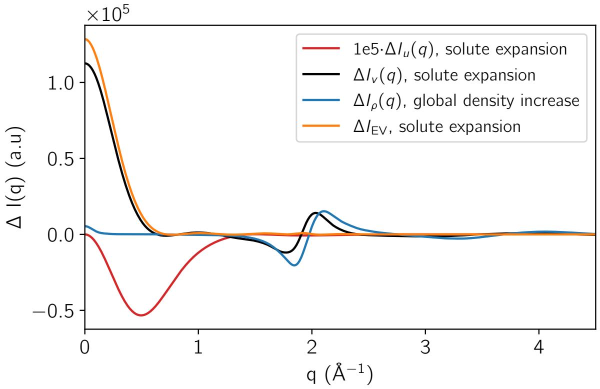

Figure 9: Difference-scattering curves of the individual contributions to the solvent-solvent term (black curve), calculated from the difference of the scattering of the LJ RDFs containing the $\sigma_{ij} = 6.7$ Å solute, and the 'neat' LJ liquid, where all particles have a σ_{ij} of 3.4 Å. The blue curve shows the global density difference from calculating $\Delta I(q)$ from reducing the NVT box volume with the same amount as we estimate the excluded volume increases in the $\sigma = 6.7$ Å case. The orange curve shows the excluded volume term calculated via eq. 17. The red curve shows a simple model of difference-scattering from a homogeneous sphere that is expanding by 3.21 Å, according to fitted values of the excluded volume radii.

structural changes in the shell and of the 'hole' left by the solute, for this simple test system. We can test this characterization by comparing the absolute I_{EV} term to an analytical 'hole'-shell model, inspired by SAXS-methodologies [89]. Such models are based on homogeneous volumes, and not scattering from distributions of discrete particles with preferred interatomic distances, so we cannot expect a 1:1 agreement between scattering from an analytic hole-shell model and the I_{EV} term. Nevertheless, we re-purpose a well-known core-shell model [89], where the analytical form factor F(q) of a core particle with scattering length density ρ_c and volume V_c is surrounded by a shell of scattering length density ρ_s and volume V_s is [89] (see fig. 10):

$$F(q) = \frac{3}{V_s}\big(V_c(\rho_c - \rho_s)\Psi(qr_c) + V_s(\rho_s - \rho_{\mathrm{bulk}})\Psi(qr_s)\big), \tag{20}$$

with

$$\Psi(qr) = \frac{\sin(qr) - qr\cos(qr)}{(qr)^3},$$

where ρ_{bulk} is the scattering length density of the bulk solvent. In our case, the scattering length density of the core particle is zero, since the solute particle scattering should not be included twice (both here, and in the solute-term). Thus, the form factor becomes:

$$F(q) = \frac{3}{V_s}\big(V_s(\rho_s - \rho_{\mathrm{bulk}})\Psi(qr_s) - V_c\rho_s\Psi(qr_c)\big), \tag{21}$$

and we see that the second term indeed gives the negative scattering length density of the 'hole' left in the shell. Since the local density of particles in the shell will differ slightly from the bulk solvent, the first term of the form factor needs to reflect the change in scattering contrast between these two (albeit not very well-defined) regions of solvent, $\rho_s - \rho_{\mathrm{bulk}}$, as is indeed the case in eq. 21. The model is illustrated in the top right part of fig. 10. We calculate the scattered intensity as

$$I(q) = \alpha F(q)^2, \tag{22}$$

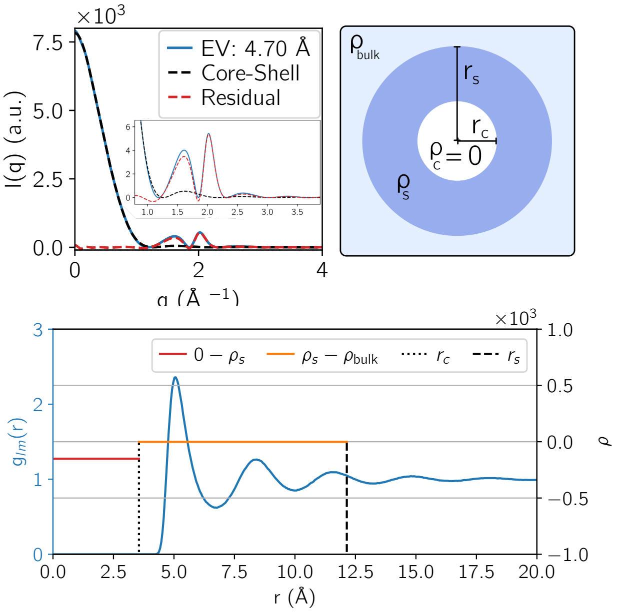

Figure 10: Top left: The blue curve shows the excluded volume term for the $\sigma_{ij} = 4.7$ Å LJ solute system, calculated with eq. 22. The dashed black line shows scattering from the analytical core-shell model, calculated from fitting the scattering length densities and shell radii in eq. 21, as well as the overall scaler α in eq. 19 to minimize the difference (red dashed curve) to the EV term. The inset shows a zoom-in on the high-q features. The analytic model can capture the low-q feature, but not the later oscillations. Top right: Core-hole continuum model scattering densities, made from subtracting a core sphere with the same scattering length density as in the shell, ρ_s from a bigger sphere of ρ_s, such that $\rho_c = 0$. Bottom plot: Comparison of the RDF used to obtain the EV form factor in eq. 17 to the fitted scattering length densities and radii of the core-hole r_c, as well as the shell r_s.

with α being a free parameter to scale the total intensity to account for the number of particles used in the MD simulations. We can thus now fit the core (hole) and shell radii, r_c and r_s, as well as the scattering length densities to the excluded volume scattering signal, to see if the fitted values give reasonable estimates while reproducing I_{EV}. Figure 10 shows this for the absolute scattering of the solute-$\sigma_{ij} = 4.70$ Å LJ liquid, which was chosen since the both its high- and low-q features are visible within the same scale. The top right plot in fig. 10 shows the results of this fit. The plot shows the $I^{4.7}_{\mathrm{EV}}(q)$ signal (blue curve) and the scattering from the analytical form factor as dashed black lines. The bottom plot shows the fitted radii and scattering length densities. The fitted radii corresponds well with the structure of the shell encoded in the RDF, even for this simple description. The difference in fitted scattering length densities between the bulk solvent and the shell is only around 1%, which is to be expected as the two regions are made up of the same type of particles. Returning to the top left plot, we observe how the model captures the low-q increase arising from the excluded volume, but not the oscillations at longer q (see also the inset). We thus assign them to the non-homogeneous structure of the solvent shell, which cannot be captured by a continuum model, but is encoded in the cross-term RDF from which the excluded volume signal is constructed.

In conclusion, this analysis has identified the two main contributions of the excluded volume term of eq. 18 as the solvent-solvent scattering from the excluded volume created by the solute, as well as the local structure of the solvent shell. The excluded volume formulation thus provides a method for calculating the local solvent-solvent scattering from the shell (and hole). For difference-scattering analyses, this method can prove advantageous since it is free from (possibly unwanted) contributions from global density changes due to changes in the excluded volume changing the average solvent-solvent distances in the bulk solvent for simulations of the same cell size.

Summary and Outlook

With this work, we have analysed the effect of the excluded volume on the global density normalization in the sampling of RDFs from finite systems on the low-q behaviour of the structure factor and the coherent x-ray scattering signal. We have done so in order to design a robust set of best practices for the forward modelling of coherent x-ray scattering across the entire q-range, all the way to 0 Å^{-1}. We showed how the density normalization of the RDFs sampled from particles l in finite size systems does not take into account the excluded volume of the l-type particles themselves, and how this makes the RDFs converge to a value different from 1. This causes the integral in the central scattering equations (3, 4) to diverge, if the RDFs are not corrected for this type of finite-size artifacts. We have developed a conceptually simple way of correcting finite-size sampled RDFs, by accounting for this excluded volume in the density normalization of the RDF. This correction was then evaluated alongside two previously published corrections which we repurposed for x-ray scattering. We have also shown the evaluation of $S(0;R)$ and $I(0;R)$ at increasing cutoff-values R of the central integral in the scattering equations as a robust way of assessing whether or not the low-q region of the resulting scattering signal has converged to its final value, as these quantities converge to a constant value with increasing R. We hope that this method will be used to include more of the low-q region in future work that focuses on the solvent-shell structure and changes therein.

In simple test-systems, we were able to use the volume-correction to identify the size of the excluded volume and correct the RDFs to provide converged scattering signals across the entire q-range. However, both the volume-correction and the correction provided by Perera and coworkers are reliant on parameters that have to be fitted or estimated, whereas the van der Vegt correction is parameter-free. We have used the isothermal compressibility of water to test the corrections against experimental results, and found the volume- and van der Vegt correction to recover the experimental isothermal compressibility with the best accuracy.

We also assessed the effects of MD-sampling time on the effectiveness of the RDF corrections, and have shown that for the cross-term RDFs, the significant reduction in particle-pairs that can be sampled for each frame compared to the solvent-solvent RDFs necessitates higher requirements on sampling-time for the corrections to make $S(0;R)$ converge. For the Ag^+ ion in water, roughly 20 ns sampling time was needed. Water correlations are on the 1-10 ps range, meaning that roughly 2,000-20,000 uncorrelated frames were needed for robust evaluations of the corrections. Since the computational cost of classical MD simulations has been greatly reduced with the advent of performing calculations on GPUs, it has become a routine operation to produce >20 ns trajectories on a single GPU node, and we therefore recommend always assessing the convergence of the corrections against sampling time, before continuing to employ the simulated signal in further modeling, if the computational cost of the employed potential allows for it.

We furthermore evaluated the corrections on a more intricate system, namely the $[Fe(bpy)3]^{2+}$ complex in water. We observe that larger excluded volumes require larger R-values to reach convergence, thus requiring larger simulation cells. For this system, only the van der Vegt correction managed to flatten out the $I(0;R)$-signal within the limits of the range of the sampled RDF. This system also illustrated the intricate nature and m-extension dependence of the excluded volume concept for real systems.

While there of course only exists a single *correct* RDF, the differences between the RDFs after applying the various corrections are so small that it can be advantageous to test multiple corrections from an operational perspective, before deciding on which to use going forward.

Finally, we analysed $I_{EV}(q)$ as an alternative to the solvent-solvent term $I_v(q)$ for including scattering from the local structure of the solvation shell and its excluded volume. We identified two different features of the $I_{EV}(q)$ term, arising from (1) the scattering from the contrast between the solvent and the hole left by the excluded volume, and (2) from the local structure in the solvent shell around the solute. We showed that, for analysis of difference scattering signals, e.g. from time-resolved experiments, the $I_{EV}(q)$ term is less affected by pertubations in the bulk solvent density that can manifest themselves when the overall extension of the solute is modified but the simulation cell is kept constant over both simulations.

Based on our findings in this work, we propose the following best-practices guidelines to minimize low-q oscillations from finite size systems:

1. Perform MD (equilibrium) production runs in the NVT ensemble if possible, to keep the cell volume fixed throughout the simulation. Non-equilibrium simulations should be carried out in the NVE ensemble [91].

2. If computationally tractable (i.e. as with classical MD), thoroughly sample the RDFs to avoid buildup of statistical noise - i.e. test convergence of $S(0;R)$ as a function of sampling time as in fig. 7 - and use simulation cells big enough to allow for the $S(0;R)$ analysis.

3. If using 'black-box' RDF-sampling codes and the van der Vegt correction, use a density-normalization for the RDF-sampling consistent with the correction (or vice versa). Similarly, if NPT production runs were performed, the $N_{lm}(r)$ histograms of the RDFs should be normalized with the current volume of the cell at each frame.

4. Perform the $S(0;R)$ or $I(0;R)$-analysis to ascertain converged integrals at long R-values and to benchmark the corrections.

5. Various window functions should be tested in conjunction with the RDF-corrections. There unfortunately isn't a 'one-size-fits-all' solution to all simulation cell sizes and systems, so their influence on the entire scattering signal should be minimized.

6. If reliability all the way to $q = 0$ is desired, either the calculated value should be tested against the isothermal compressibility if such an experimental quantity is available, or the $S(q = 0)$-dependence on density should be assessed to ensure that the $S(q = 0)$ limit exhibits the expected behaviour with respect to density changes.

7. If modeling dynamics or large structural changes of a solute in a solvent in the NVT ensemble, any significant changes to the size of the solute will be reflected in an overall change of the average solvent-solvent distances, creating a 'density'-contribution in the solvent-solvent scattering signal, as shown in fig. 9. Further treatment of this term is highly dependent on the specifics of the desired analysis. For ultra-fast studies of, say, an expanding solute, the surrounding solvent molecules might not have had time to re-arrange themselves to re-obtain the density it had before the expansion of the solute, and the term might also be present in

experimental data. One can imagine other cases where the solvent has had enough time to re-equilibrate, and thus including the density term can be avoided by calculating $I_{EV}(q)$ from eq. 22 instead. The density term can also be eliminated by expanding the simulation cell with the increase in excluded volume, or by re-equilibrating the cell size in an NPT simulation, stopping at the same density for the previous simulation of the non-expanded solute.

The software package 'grsq' has been created to apply the corrections and simulate the signals, and is available through the Python Package Index [36]. Further examples on how to use it apart from the ones already included can be found in the data-repository accompanying this work [37].

References

[1] A. Rondi, Y. Rodriguez, T. Feurer, and A. Cannizzo, "Solvation-driven charge transfer and localization in metal complexes," *Accounts of Chemical Research*, vol. 48, no. 5, pp. 1432–1440, 2015.

[2] P. Wernet, K. Kunnus, I. Josefsson, I. Rajkovic, W. Quevedo, M. Beye, S. Schreck, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. W. Hartsock, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, F. M. F. de Groot, K. J. Gaffney, S. Techert, M. Odelius, and A. Föhlisch, "Orbital-specific mapping of the ligand exchange dynamics of fe(CO)$_5$ in solution," *Nature*, vol. 520, no. 7545, pp. 78–81, 2015.

[3] G. Levi, E. Biasin, A. O. Dohn, and H. Jónsson, "On the interplay of solvent and conformational effects in simulated excited-state dynamics of a copper phenanthroline photosensitizer," *Physical Chemistry Chemical Physics*, vol. 22, no. 2, pp. 748–757, 2020.

[4] M. Pápai, M. Abedi, G. Levi, E. Biasin, M. M. Nielsen, and K. B. Møller, "Theoretical evidence of solvent-mediated excited-state dynamics in a functionalized iron sensitizer," *The Journal of Physical Chemistry C*, vol. 123, no. 4, pp. 2056–2065, 2019.

[5] D. B. Zederkof, K. B. Møller, M. M. Nielsen, K. Haldrup, L. González, and S. Mai, "Resolving femtosecond solvent reorganization dynamics in an iron complex by nonadiabatic dynamics simulations," *Journal of the American Chemical Society*, vol. 144, no. 28, pp. 12861–12873, 2022.

[6] S. E. Canton, K. S. Kjær, G. Vankó, T. B. van Driel, S. ichi Adachi, A. Bordage, C. Bressler, P. Chabera, M. Christensen, A. O. Dohn, A. Galler, W. Gawelda, D. Gosztola, K. Haldrup, T. Harlang, Y. Liu, K. B. Møller, Z. Németh, S. Nozawa, M. Pápai, T. Sato, T. Sato, K. Suarez-Alcantara, T. Togashi, K. Tono, J. Uhlig, D. A. Vithanage, K. Wärnmark, M. Yabashi, J. Zhang, V. Sundström, and M. M. Nielsen, "Visualizing the non-equilibrium dynamics of photoinduced intramolecular electron transfer with femtosecond x-ray pulses," *Nature Communications*, vol. 6, no. 1, 2015.

[7] A. O. Dohn, K. S. Kjær, T. B. Harlang, S. E. Canton, M. M. Nielsen, and K. B. Møller, "Electron transfer and solvent-mediated electronic localization in molecular photocatalysis," *Inorganic Chemistry*, vol. 55, no. 20, pp. 10637–10644, 2016.

[8] E. Biasin, Z. W. Fox, A. Andersen, K. Ledbetter, K. S. Kjær, R. Alonso-Mori, J. M. Carlstad, M. Chollet, J. D. Gaynor, J. M. Glownia, K. Hong, T. Kroll, J. H. Lee, C. Liekhus-Schmaltz, M. Reinhard, D. Sokaras, Y. Zhang, G. Doumy, A. M. March, S. H. Southworth, S. Mukamel,

K. J. Gaffney, R. W. Schoenlein, N. Govind, A. A. Cordones, and M. Khalil, "Direct observation of coherent femtosecond solvent reorganization coupled to intramolecular electron transfer," *Nature Chemistry*, vol. 13, no. 4, pp. 343–349, 2021.

[9] A. O. Dohn, E. Örn Jónsson, K. S. Kjær, T. B. van Driel, M. M. Nielsen, K. W. Jacobsen, N. E. Henriksen, and K. B. Møller, "Direct dynamics studies of a binuclear metal complex in solution: The interplay between vibrational relaxation, coherence, and solvent effects," *The Journal of Physical Chemistry Letters*, vol. 5, no. 14, pp. 2414–2418, 2014.

[10] K. Haldrup, G. Levi, E. Biasin, P. Vester, M. G. Laursen, F. Beyer, K. S. Kjær, T. B. van Driel, T. Harlang, A. O. Dohn, R. J. Hartsock, S. Nelson, J. M. Glownia, H. T. Lemke, M. Christensen, K. J. Gaffney, N. E. Henriksen, K. B. Møller, and M. M. Nielsen, "Ultrafast x-ray scattering measurements of coherent structural dynamics on the ground-state potential energy surface of a diplatinum molecule," *Physical Review Letters*, vol. 122, no. 6, 2019.

[11] C. Petermayer and H. Dube, "Indigoid photoswitches: Visible light responsive molecular tools," *Accounts of Chemical Research*, vol. 51, no. 5, pp. 1153–1163, 2018.

[12] H. Zulfikri, M. A. J. Koenis, M. M. Lerch, M. D. Donato, W. Szymański, C. Filippi, B. L. Feringa, and W. J. Buma, "Taming the complexity of donor–acceptor stenhouse adducts: Infrared motion pictures of the complete switching pathway," *Journal of the American Chemical Society*, vol. 141, no. 18, pp. 7376–7384, 2019.

[13] H. Xi, Z. Zhang, W. Zhang, M. Li, C. Lian, Q. Luo, H. Tian, and W.-H. Zhu, "All-visible-light-activated dithienylethenes induced by intramolecular proton transfer," *Journal of the American Chemical Society*, vol. 141, no. 46, pp. 18467–18474, 2019.

[14] M. Krell-Jørgensen, H. Zulfikri, M. G. Bonnevie, F. S. Bro, A. O. Dohn, and L. Laraia, "Red-shifted and thermally bistable one-way quantitative hemithioindigo-derived photoswitches enabled by isomer-specific excited state intramolecular proton transfer," *Chemical Communications*, 2023.

[15] H. Ki, S. Choi, J. Kim, E. H. Choi, S. Lee, Y. Lee, K. Yoon, C. W. Ahn, D.-S. Ahn, J. H. Lee, J. Park, I. Eom, M. Kim, S. H. Chun, J. Kim, H. Ihee, and J. Kim, "Optical kerr effect of liquid acetonitrile probed by femtosecond time-resolved x-ray liquidography," *Journal of the American Chemical Society*, vol. 143, no. 35, pp. 14261–14273, 2021.

[16] C. Huang, K. T. Wikfeldt, T. Tokushima, D. Nordlund, Y. Harada, U. Bergmann, M. Niebuhr, T. M. Weiss, Y. Horikawa, M. Leetmaa, M. P. Ljungberg, O. Takahashi, A. Lenz, L. Ojamae, A. P. Lyubartsev, S. Shin, L. G. M. Pettersson, and A. Nilsson, "The inhomogeneous structure of water at ambient conditions," *PNAS*, vol. 106, no. 36, pp. 15214–15218, 2009.

[17] G. N. I. Clark, G. L. Hura, J. Teixeira, A. K. Soper, and T. Head-Gordon, "Small-angle scattering and the structure of ambient liquid water," *Proceedings of the National Academy of Sciences*, vol. 107, no. 32, pp. 14003–14007, 2010.

[18] K. Haldrup, G. Vankó, W. Gawelda, A. Galler, G. Doumy, A. M. March, E. P. Kanter, A. Bordage, A. Dohn, T. B. van Driel, K. S. Kjær, H. T. Lemke, S. E. Canton, J. Uhlig, V. Sundström, L. Young, S. H. Southworth, M. M. Nielsen, and C. Bressler, "Guest–host interactions investigated by time-resolved x-ray spectroscopies and scattering at MHz rates: Solvation dynamics and photoinduced spin transition in aqueous $\mathrm{Fe(bipy)}_3^{2+}$," *The Journal of Physical Chemistry A*, vol. 116, no. 40, pp. 9878–9887, 2012.

[19] J. H. Lee, M. Wulff, S. Bratos, J. Petersen, L. Guerin, J.-C. Leicknam, M. Cammarata, Q. Kong, J. Kim, K. B. Møller, and H. Ihee, "Filming the birth of molecules and accompanying solvent rearrangement," *Journal of the American Chemical Society*, vol. 135, no. 8, pp. 3255–3261, 2013.

[20] K. H. Kim, J. G. Kim, S. Nozawa, T. Sato, K. Y. Oang, T. W. Kim, H. Ki, J. Jo, S. Park, C. Song, T. Sato, K. Ogawa, T. Togashi, K. Tono, M. Yabashi, T. Ishikawa, J. Kim, R. Ryoo, J. Kim, H. Ihee, and S. ichi Adachi, "Direct observation of bond formation in solution with femtosecond x-ray scattering," *Nature*, vol. 518, no. 7539, pp. 385–389, 2015.

[21] D. Khakhulin, L. M. Lawson Daku, D. Leshchev, G. Newby, M. Jarenmark, C. Bressler, M. Wulff, and S. Canton, "Visualizing the coordination-spheres of photoexcited transition metal complexes with ultrafast hard x-rays," *Phys. Chem. Chem. Phys.*, vol. 21, no. 18, pp. 9277–9284, 2019.

[22] L. M. Lawson Daku and A. Hauser, "Ab initio molecular dynamics study of an aqueous solution of $[\mathrm{Fe(bpy)_3}](\mathrm{Cl})_2$ in the low-spin and in the high-spin states," *J. Phys. Chem. Lett.*, vol. 1, no. 12, pp. 1830–1835, 2010.

[23] V. Pham, I. Tavernelli, C. Milne, R. van der Veen, P. D'Angelo, C. Bressler, and M. Chergui, "The solvent shell structure of aqueous iodide: X-ray absorption spectroscopy and classical, hybrid QM/MM and full quantum molecular dynamics simulations," *Chemical Physics*, vol. 371, no. 1-3, pp. 24–29, 2010.

[24] L. M. Lawson Daku, "Spin-state dependence of the structural and vibrational properties of solvated iron (ii) polypyridyl complexes from aimd simulations: aqueous $[\mathrm{Fe(bpy)_3}]\mathrm{Cl_2}$, a case study," *Phys. Chem. Chem. Phys.*, vol. 20, no. 9, pp. 6236–6253, 2018.

[25] K. Haldrup, W. Gawelda, R. Abela, R. Alonso-Mori, U. Bergmann, A. Bordage, M. Cammarata, S. E. Canton, A. O. Dohn, T. B. van Driel, D. M. Fritz, A. Galler, P. Glatzel, T. Harlang, K. S. Kjær, H. T. Lemke, K. B. Møller, Z. Németh, M. Pápai, N. Sas, J. Uhlig, D. Zhu, G. Vankó, V. Sundström, M. M. Nielsen, and C. Bressler, "Observing solvation dynamics with simultaneous femtosecond x-ray emission spectroscopy and x-ray scattering," *The Journal of Physical Chemistry B*, vol. 120, no. 6, pp. 1158–1168, 2016.

[26] M. R. Panman, E. Biasin, O. Berntsson, M. Hermann, S. Niebling, A. J. Hughes, J. Kübel, K. Atkovska, E. Gustavsson, A. Nimmrich, A. O. Dohn, M. Laursen, D. B. Zederkof, A. Honarfar, K. Tono, T. Katayama, S. Owada, T. B. van Driel, K. Kjaer, M. M. Nielsen, J. Davidsson, J. Uhlig, K. Haldrup, J. S. Hub, and S. Westenhoff, "Observing the structural evolution in the photodissociation of diiodomethane with femtosecond solution x-ray scattering," *Phys. Rev. Lett.*, vol. 125, p. 226001, 2020.

[27] T. B. van Driel, K. S. Kjær, R. W. Hartsock, A. O. Dohn, T. Harlang, M. Chollet, M. Christensen, W. Gawelda, N. E. Henriksen, J. G. Kim, K. Haldrup, K. H. Kim, H. Ihee, J. Kim, H. Lemke, Z. Sun, V. Sundström, W. Zhang, D. Zhu, K. B. Møller, M. M. Nielsen, and K. J. Gaffney, "Atomistic characterization of the active-site solvation dynamics of a model photocatalyst," *Nature Communications*, vol. 7, no. 1, 2016.

[28] T. Katayama, T.-K. Choi, D. Khakhulin, A. O. Dohn, C. J. Milne, G. Vankó, Z. Németh, F. A. Lima, J. Szlachetko, T. Sato, S. Nozawa, S. ichi Adachi, M. Yabashi, T. J. Penfold, W. Gawelda, and G. Levi, "Atomic-scale observation of solvent reorganization influencing

photoinduced structural dynamics in a copper complex photosensitizer," *Chemical Science*, 2023.

[29] A. O. Dohn, E. Biasin, K. Haldrup, M. M. Nielsen, N. E. Henriksen, and K. B. Møller, "On the calculation of x-ray scattering signals from pairwise radial distribution functions," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 48, no. 24, p. 244010, 2015.

[30] J. Kim, J. G. Kim, H. Ki, C. W. Ahn, and H. Ihee, "Estimating signal and noise of time-resolved x-ray solution scattering data at synchrotrons and XFELs," *J. Synchrotron Radiat.*, vol. 27, no. Pt 3, pp. 633–645, 2020.

[31] K. H. Kim, J. H. Lee, J. Kim, S. Nozawa, T. Sato, A. Tomita, K. Ichiyanagi, H. Ki, J. Kim, S. ichi Adachi, and H. Ihee, "Solvent-dependent molecular structure of ionic species directly measured by ultrafast x-ray solution scattering," *Physical Review Letters*, vol. 110, no. 16, 2013.

[32] D. J. Hsu, D. Leshchev, I. Kosheleva, K. L. Kohlstedt, and L. X. Chen, "Integrating solvation shell structure in experimentally driven molecular dynamics using x-ray solution scattering data," *J. Chem. Phys.*, vol. 152, no. 20, p. 204115, 2020.

[33] E. H. Choi, Y. Lee, J. Heo, and H. Ihee, "Reaction dynamics studied via femtosecond x-ray liquidography at x-ray free-electron lasers," *Chemical Science*, vol. 13, no. 29, pp. 8457–8490, 2022.

[34] P. Vester, K. Kubicek, R. Alonso-Mori, T. Assefa, E. Biasin, M. Christensen, A. O. Dohn, T. B. van Driel, A. Galler, W. Gawelda, T. C. B. Harlang, N. E. Henriksen, K. S. Kjær, T. S. Kuhlman, Z. Németh, Z. Nurekeyev, M. Pápai, J. Rittman, G. Vankó, H. Yavas, D. B. Zederkof, U. Bergmann, M. M. Nielsen, K. B. Møller, K. Haldrup, and C. Bressler, "Tracking structural solvent reorganization and recombination dynamics following e^- photoabstraction from aqueous i^- with femtosecond x-ray spectroscopy and scattering," *The Journal of Chemical Physics*, vol. 157, no. 22, p. 224201, 2022.

[35] A. Nimmrich, M. R. Panman, O. Berntsson, E. Biasin, S. Niebling, J. Petersson, M. Hoernke, A. Björling, E. Gustavsson, T. B. van Driel, A. O. Dohn, M. Laursen, D. B. Zederkof, K. Tono, T. Katayama, S. Owada, M. M. Nielsen, J. Davidsson, J. Uhlig, J. S. Hub, K. Haldrup, and S. Westenhoff, "Solvent-dependent structural dynamics in the ultrafast photodissociation reaction of triiodide observed with time-resolved x-ray solution scattering," *Journal of the American Chemical Society*, 2023.

[36] "grsq." `https://pypi.org/project/grsq`. Accessed: 2023-01-01.

[37] A. O. Dohn, "Finite-size effects on the calculation of x-ray scattering from molecular dynamics simulations," 2023.

[38] F. Zernike and J. A. Prins, "Die beugung von röntgenstrahlen in flüssigkeiten als effekt der molekülanordnung," *Zeitschrift für Physik A Hadrons and nuclei*, vol. 41, no. 2-3, pp. 184–194, 1927.

[39] B. E. Warren, *X-Ray Diffraction*. Dover Books on Physics, Mineola, NY: Dover Publications, 1990.

[40] J. Als-Nielsen and D. McMorrow, *Elements of modern X-ray physics*. Hoboken, NJ: Wiley-Blackwell, 2 ed., 2011.

[41] A. J. C. Wilson and V. Geist, *International Tables for Crystallography. Volume C: Mathematical, Physical and Chemical Tables.* Kluwer Academic Publishers, Dordrecht/Boston/London, (published for the International Union of Crystallography), 1993.

[42] T. E. Faber and J. M. Ziman, "A theory of the electrical properties of liquid metals," *Philosophical Magazine*, vol. 11, no. 109, pp. 153–173, 1965.

[43] J. G. Kirkwood and F. P. Buff, "The statistical mechanical theory of solutions. i," *The Journal of Chemical Physics*, vol. 19, no. 6, pp. 774–777, 1951.

[44] A. Guinier, *X-ray diffraction.* Dover Books on Physics, Mineola, NY: Dover Publications, 1994.

[45] J.-P. Hansen and I. MacDonald, *Theory of Simple Liquids.* Elsevier, 4th ed., 2013.

[46] P. A. Egelstaff, *An Introduction to the Liquid State.* Academic Press, 1967.

[47] I. Shulgin and E. Ruckenstein, "Kirkwood-buff integrals in aqueous alcohol systems: comparison between thermodynamic calculations and x-ray scattering experiments," *The Journal of Physical Chemistry B*, vol. 103, no. 13, pp. 2496–2503, 1999.

[48] S. Christensen, G. H. Peters, F. Y. Hansen, J. P. O'connell, and J. Abildskov, "Generation of thermodynamic data for organic liquid mixtures from molecular simulations," *Molecular Simulation*, vol. 33, no. 4-5, pp. 449–457, 2007.

[49] J. W. Nichols, S. G. Moore, and D. R. Wheeler, "Improved implementation of kirkwood-buff solution theory in periodic molecular simulations," *Physical Review E*, vol. 80, no. 5, 2009.

[50] M. A. Blanco, E. Sahin, Y. Li, and C. J. Roberts, "Reexamining protein–protein and protein–solvent interactions from kirkwood-buff analysis of light scattering in multi-component solutions," *The Journal of Chemical Physics*, vol. 134, no. 22, p. 225103, 2011.

[51] R. Wedberg, J. P. O'Connell, G. H. Peters, and J. Abildskov, "Pair correlation function integrals: Computation and use," *The Journal of Chemical Physics*, vol. 135, no. 8, p. 084113, 2011.

[52] P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, and J.-M. Simon, "Kirkwood–buff integrals for finite volumes," *The Journal of Physical Chemistry Letters*, vol. 4, no. 2, pp. 235–238, 2012.

[53] S. K. Schnell, P. Englebienne, J.-M. Simon, P. Krüger, S. P. Balaji, S. Kjelstrup, D. Bedeaux, A. Bardow, and T. J. Vlugt, "How to apply the kirkwood–buff theory to individual species in salt solutions," *Chemical Physics Letters*, vol. 582, pp. 154–157, 2013.

[54] A. Perera and B. Kežić, "Fluctuations and micro-heterogeneity in mixtures of complex liquids," *Faraday Discussions*, vol. 167, p. 145, 2013.

[55] J. Milzetti, D. Nayar, and N. F. A. van der Vegt, "Convergence of kirkwood–buff integrals of ideal and nonideal aqueous solutions using molecular dynamics simulations," *The Journal of Physical Chemistry B*, vol. 122, no. 21, pp. 5515–5526, 2018.

[56] N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon, and T. J. H. Vlugt, "Finite-size effects of kirkwood–buff integrals from molecular simulations," *Molecular Simulation*, vol. 44, no. 7, pp. 599–612, 2017.

[57] P. Ganguly and N. F. A. van der Vegt, "Convergence of sampling kirkwood–buff integrals of aqueous solutions with molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 9, no. 3, pp. 1347–1355, 2013.

[58] R. Cortes-Huerto, K. Kremer, and R. Potestio, "Communication: Kirkwood-buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 145, no. 14, p. 141103, 2016.

[59] N. Dawass, P. Krüger, S. K. Schnell, J.-M. Simon, and T. Vlugt, "Kirkwood-buff integrals from molecular simulation," *Fluid Phase Equilibria*, vol. 486, pp. 21–36, 2019.

[60] N. Dawass, P. Krüger, S. K. Schnell, O. A. Moultos, I. G. Economou, T. J. H. Vlugt, and J.-M. Simon, "Kirkwood-buff integrals using molecular simulation: Estimation of surface effects," *Nanomaterials*, vol. 10, no. 4, p. 771, 2020.

[61] M. Sevilla and R. Cortes-Huerto, "Connecting density fluctuations and kirkwood–buff integrals for finite-size systems," *The Journal of Chemical Physics*, vol. 156, no. 4, p. 044502, 2022.

[62] J.-M. Simon, P. Krüger, S. K. Schnell, T. J. H. Vlugt, S. Kjelstrup, and D. Bedeaux, "Kirkwood–buff integrals: From fluctuations in finite volumes to the thermodynamic limit," *The Journal of Chemical Physics*, vol. 157, no. 13, p. 130901, 2022.

[63] B. G. Levine, J. E. Stone, and A. Kohlmeyer, "Fast analysis of molecular dynamics trajectories with graphics processing units—radial distribution function histogramming," *Journal of Computational Physics*, vol. 230, no. 9, pp. 3556–3569, 2011.

[64] A. Perera, L. Zoranić, F. Sokolić, and R. Mazighi, "A comparative molecular dynamics study of water–methanol and acetone–methanol mixtures," *Journal of Molecular Liquids*, vol. 159, no. 1, pp. 52–59, 2011.

[65] A. K. Soper and E. R. Barney, "Extracting the pair distribution function from white-beam x-ray total scattering data," *Journal of Applied Crystallography*, vol. 44, no. 4, pp. 714–726, 2011.

[66] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore, "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide q-range," *The Journal of Chemical Physics*, vol. 138, no. 7, p. 074506, 2013.

[67] D. Dhabal, K. T. Wikfeldt, L. B. Skinner, C. Chakravarty, and H. K. Kashyap, "Probing the triplet correlation function in liquid water by experiments and molecular simulations," *Physical Chemistry Chemical Physics*, vol. 19, no. 4, pp. 3265–3278, 2017.

[68] E. Lorch, "Neutron diffraction by germania, silica and radiation-damaged silica glasses," *Journal of Physics C: Solid State Physics*, vol. 2, no. 2, pp. 229–237, 1969.

[69] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLOS Computational Biology*, vol. 13, no. 7, p. e1005659, 2017.

[70] J. Chodera, A. Rizzi, L. Naden, K. Beauchamp, P. Grinaway, J. Fass, A. Wade, B. Rustenburg, I. Pulido, G. A. Ross, M. Henry, A. Krämer, H. B. Macdonald, J. Rodríguez-Guerra, I. Zhang, A. Simmonett, D. W. Swenson, M. J. Williamson, Hb0402, J. Fennick, S. Roet, SimonBoothroyd, A. Silveira, and D. Rufa, "choderalab/openmmtools: 0.21.5," 2022.

[71] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "PACKMOL: A package for building initial configurations for molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 30, no. 13, pp. 2157–2164, 2009.

[72] D. Case, H. Aktulga, K. Belfon, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, G. Cisneros, V. Cruzeiro, T. Darden, R. Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, C. Jin, K. Kasavajhala, M. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N. Skrynnikov, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, Y. Xue, D. York, S. Zhao, , and P. Kollman, "Amber 2021," 2021.

[73] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," *The Journal of Chemical Physics*, vol. 120, no. 20, pp. 9665–9678, 2004.

[74] P. Li, L. F. Song, and K. M. Merz, "Systematic parameterization of monovalent ions employing the nonbonded model," *Journal of Chemical Theory and Computation*, vol. 11, no. 4, pp. 1645–1657, 2015.

[75] K. S. Kjær, T. B. V. Driel, T. C. B. Harlang, K. Kunnus, E. Biasin, K. Ledbetter, R. W. Hartsock, M. E. Reinhard, S. Koroidov, L. Li, M. G. Laursen, F. B. Hansen, P. Vester, M. Christensen, K. Haldrup, M. M. Nielsen, A. O. Dohn, M. I. Pápai, K. B. Møller, P. Chabera, Y. Liu, H. Tatsuno, C. Timm, M. Jarenmark, J. Uhlig, V. Sundstöm, K. Wärnmark, P. Persson, Z. Németh, D. S. Szemes, É. Bajnóczi, G. Vankó, R. Alonso-Mori, J. M. Glownia, S. Nelson, M. Sikorski, D. Sokaras, S. E. Canton, H. T. Lemke, and K. J. Gaffney, "Finding intersections between electronic excited state potential energy surfaces with simultaneous ultrafast x-ray scattering and spectroscopy," *Chemical Science*, vol. 10, no. 22, pp. 5749–5760, 2019.

[76] H. Zulfikri, M. Pápai, and A. O. Dohn, "Simulating the solvation structure of low- and high-spin $[\mathrm{Fe(bpy)_3}]^{2+}]$: long-range dispersion and many-body effects," *Physical Chemistry Chemical Physics*, vol. 24, no. 27, pp. 16655–16670, 2022.

[77] P. Li and K. M. Merz, "MCPB.py: A python based metal center parameter builder," *Journal of Chemical Information and Modeling*, vol. 56, no. 4, pp. 599–604, 2016.

[78] H. A. Lorentz, "Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase," *Annalen der Physik*, vol. 248, no. 1, pp. 127–136, 1881.

[79] D. Berthelot, "Sur le mélange des gaz," *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, vol. 126, pp. 1703–1855, 1898.

[80] A. O. Dohn, E. O. Jónsson, G. Levi, J. J. Mortensen, O. Lopez-Acevedo, K. S. Thygesen, K. W. Jacobsen, J. Ulstrup, N. E. Henriksen, K. B. Møller, and H. Jónsson, "Grid-based projector augmented wave (GPAW) implementation of quantum mechanics/molecular mechanics

(QM/MM) electrostatic embedding and application to a solvated diplatinum complex," *Journal of Chemical Theory and Computation*, vol. 13, no. 12, pp. 6010–6022, 2017.

[81] G. Levi, M. Pápai, N. E. Henriksen, A. O. Dohn, and K. B. Møller, "Solution structure and ultrafast vibrational relaxation of the PtPOP complex revealed by δSCF-QM/MM direct dynamics simulations," *The Journal of Physical Chemistry C*, vol. 122, no. 13, pp. 7100–7119, 2018.

[82] A. O. Dohn, "Multiscale electrostatic embedding simulations for modeling structure and dynamics of molecules in solution: A tutorial review," *International Journal of Quantum Chemistry*, vol. 120, no. 21, pp. 1–22, 2020.

[83] T. S. Carlton, "Using heat capacity and compressibility to choose among two-state models of liquid water," *The Journal of Physical Chemistry B*, vol. 111, no. 47, pp. 13398–13403, 2007.

[84] H. L. Pi, J. L. Aragones, C. Vega, E. G. Noya, J. L. Abascal, M. A. Gonzalez, and C. McBride, "Anomalies in water as obtained from computer simulations of the TIP4p/2005 model: density maxima, and density, isothermal compressibility and heat capacity minima," *Molecular Simulation*, vol. 107, no. 4-6, pp. 365–374, 2009.

[85] W. Gawelda, V.-T. Pham, M. Benfatto, Y. Zaushitsyn, M. Kaiser, D. Grolimund, S. L. Johnson, R. Abela, A. Hauser, C. Bressler, and M. Chergui, "Structural determination of a short-lived excited iron(II) complex by picosecond x-ray absorption spectroscopy," *Phys. Rev. Lett.*, vol. 98, no. 5, p. 057401, 2007.

[86] W. Zhang, R. Alonso-Mori, U. Bergmann, C. Bressler, M. Chollet, A. Galler, W. Gawelda, R. Hadt, R. Hartsock, T. Kroll, K. Kjær, K. Kubiček, H. Lemke, H. Liang, D. Meyer, M. Nielsen, C. Purser, J. Robinson, E. Solomon, Z. Sun, D. Sokaras, T. Van Driel, G. Vankó, T.-C. Weng, D. Zhu, and K. Gaffney, "Tracking excited-state charge and spin dynamics in iron coordination complexes," *Nature*, vol. 509, no. 7500, pp. 345–348, 2014.

[87] R. D. B. Fraser, T. P. MacRae, and E. Suzuki, "An improved method for calculating the contribution of solvent to the x-ray diffraction pattern of biological molecules," *Journal of Applied Crystallography*, vol. 11, no. 6, pp. 693–694, 1978.

[88] O. Glatter and O. Kratky, eds., *Small angle X-ray scattering.* San Diego, CA: Academic Press, 1982.

[89] A. Guinier, G. Fournet, and K. L. Yudowitch, "Small-angle scattering of x-rays," *Structure of Matter Series*, 1955.

[90] C. J. Knight and J. S. Hub, "WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics," *Nucleic Acids Research*, vol. 43, no. W1, pp. W225–W230, 2015.

[91] E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad, and D. M. Zuckerman, "Best practices for foundations in molecular simulations [article v1.0]," *Living Journal of Computational Molecular Science*, vol. 1, no. 1, 2019.