First hyperpolarizability of water in the bulk phase: long-range electrostatic effects included via the second hyperpolarizability[†]

Guillaume Le Breton,[a] Oriane Bonhomme,[a] Emmanuel Benichou[a] and Claire Loison [*a]

The molecular first hyperpolarizability β contributes to second-order optical non-linear signals collected from molecular liquids. For the Second Harmonic Generation (SHG) response, the first hyperpolarizability $\beta(2\omega, \omega, \omega)$ often depends on the molecular electrostatic environment. This is especially true for water, due to its large second hyperpolarizability $\gamma(2\omega, \omega, \omega, 0)$. In this study we compute the electronic $\gamma(2\omega, \omega, \omega, 0)$ and $\beta(2\omega, \omega, \omega)$ for water molecules in their bulk phase using QM/MM calculations. The average value of $\gamma(2\omega, \omega, \omega, 0)$ is smaller than the one for the gaz phase, and its standard deviation among the molecules is relatively small. In addition, we demonstrate that the average second hyperpolarizability $\langle\gamma(2\omega, \omega, \omega, 0)\rangle$ can be used to describe the electrostatic effects of the neighborhood on the $\beta(2\omega, \omega, \omega)$; but only the more distant neighbors can be accounted this way. This study paves the way towards QM/MM calculations of the first hyperpolarizability of molecules in complex environments, in which long-range electrostatic effects can be crucial, for example nearby charged interfaces.

1 Introduction

Non-Linear Optical (NLO) technics are increasingly used to probe structural properties of matter. In Second Harmonic Generation (SHG) or Hyper-Rayleigh Scattering (HRS), two photons with the same fundamental frequency interact with a nonlinear material that generates a new photon with twice the energy of the initial photons[1]. SHG-based technologies have been developed to investigate simple and complex liquids[2,3], biomimetic systems[4–6], or even biological materials[7,8]. One interesting property of non-resonant SHG is the sensitivity of the response to the electrostatic environment[9]. Historically, the prototypical application is the electric-field-induced second harmonic generation (EFISHG) of molecules in a gas phase[10]. During such experiments, a macroscopic electrostatic field $\mathbf{E}^{DC}$ is applied and the SHG response of the system is described using second and third order susceptibility tensor: $\chi^{(2)}(2\omega, \omega, \omega)$ and $\chi^{(3)}(2\omega, \omega, \omega, 0)$, respectively. For a mesoscopic unitary volume, the induced total dipole moment at the second harmonic frequency $\mathbf{P}^{2\omega}$ expressed as a sum of two terms :

$$\mathbf{P}^{2\omega} \propto \chi^{(2)}(2\omega, \omega, \omega) : \mathbf{E}^{\omega}\mathbf{E}^{\omega} + \chi^{(3)}(2\omega, \omega, \omega, 0) : \mathbf{E}^{\omega}\mathbf{E}^{\omega}\mathbf{E}^{DC}, \quad (1)$$

where $\mathbf{E}^{\omega}$ is the exciting electromagnetic field at the fundamental frequency, and $\mathbf{E}^{DC}$ is the external electrostatic field that potentially modifies (i) the orientation of dipoles, and (ii) the electronic-dependent molecular intrinsic SHG response. More recently, EFISHG has received renewed interest as a non-intrusive tool for probing electric fields in diluted media with spatial and temporal definition, using laser pulses[11].

Equation 1 was also applied to condensed phases[12,13], and the EFISHG is an established technique to determine the first and the second hyperpolarizabilities of compounds in solutions[14]. Even more, studies on liquid/solid or liquid/air interfaces have reported the evolution of the Surface-SHG (S-SHG) response when the surface charge is modulated. Formalisms based on Equation 1 for the fluid nearby the surface permit to extract a surface potential or an effective surface charge[15–19]. But the interpretation of the different terms in the S-SHG intensity generated by aqueous solutions is still the subject of many recent works[20–23] requiring theoretical calculations at the molecular level[24].

To provide an interpretation at the microscopic scale, Equation 1 can also be rewritten at the molecular level:

$$\mu^{2\omega} = \frac{1}{2!}\beta^{(2)}(2\omega, \omega, \omega) : \mathbf{e}^{\omega}\mathbf{e}^{\omega}$$

$$+ \frac{1}{2!}\gamma^{(3)}(2\omega, \omega, \omega, 0) : \mathbf{e}^{\omega}\mathbf{e}^{\omega}\mathbf{e}^{DC}, \quad (2)$$

where $\mu^{2\omega}$ is the induced dipole moment of the molecule at the second harmonic frequency, $\mathbf{e}^{\omega}$ the exciting electromagnetic field in the molecular frame, $\beta^{(2)}(2\omega, \omega, \omega)$ and $\gamma^{(3)}(2\omega, \omega, \omega, 0)$ are respectively the first and second molecular hyperpolarizabilities, noted β and γ later on. The second hyperpolarizability γ describes the variations of the first hyperpolarizability value under a static external electrostatic field $\mathbf{e}^{DC}$, present in the molecular frame.

This approach permits to link the EFISHG macroscopic measurements for the gas phase to quantum calculations of the molecular first hyperpolarizability β [25–30].

In this work, we question whether and how Equation 2 can be used to predict the hyperpolarizability of molecules in liquid phases. Indeed, this equation appears very useful to calculate the first hyperpolarizability of molecules in a complex environment, for example nearby a charged surface. Given the values of β and γ obtained without external electric field, one could include the environment effects through a static electrostatic field exerted on a molecule. This electric field describing the environment could be modeled analytically, or obtained numerically using Molecular Mechanic (MM) simulations.

However, Equation 2 is generally not used this way. Predictions of molecular first hyperpolarizabilities in liquid phase often demand expensive Quantum Mechanical/Molecular Mechanic (QM/MM) calculations [9,24,26,31,32]. In such approaches, the electrostatic effect of the environment is included within the QM calculations of a molecular hyperpolarizability β^{env} for a given electrostatic environment, so that

$$\mu^{2\omega} = \frac{1}{2!}\beta^{env} : \mathbf{e}^\omega \mathbf{e}^\omega, \tag{3}$$

where β^{env} depends on the electric field exerted by the environment $\mathbf{e}^{env}$ (T convention [33]). Liang *et al.* [9] have compared the first hyperpolarizability of water within the liquid phase β^{env} obtained by QM/MM calculations to the one predicted by Equation 2, where $\mathbf{e}^{DC}$ was the electrostatic field generated by the surrounding water molecules, noted $\mathbf{e}^{env}$, and β and γ were the gas-phase values: a quantitative difference emerged.

To investigate the physical origins of the discrepancy between approximations based on Eq. 2 and more advanced models based on QM/MM electrostatic embedding, we have calculated both β^{env} and γ^{env} of individual water molecules in the liquid phase using QM/MM approaches at an optical wavelength of 800 nm (typically used for experiments [34,35]), and have scrutinized their relationship. Section 2 describes the numerical details. In Section 3, the individual values of γ^{env} and β^{env} are first reported. Then, we show that the electrostatic field generated by the environment, namely $\mathbf{e}^{env}$, is strongly heterogeneous in space. This heterogeneity explains why the application of Eq. 2 may be problematic. Finally, based on this study of $\mathbf{e}^{env}$, we propose to separate the effects of the electric field generated by the environment into a short- and a long-range part. Using this separation, we show that, for water molecule in the bulk phase, the effect of the long-range electrostatic environment on β^{env} can be included using a linear correction proportional to γ.

2 Method

The calculations are based on a sequential QM/MM approach in two steps. First, we use a classical Molecular Dynamics (MD) simulation of bulk water to obtain typical structures of the liquid. Then, we compute the first and second hyperpolarizabilities of individual water molecules at the Density Functional Theory (DFT) level, within an electrostatic embedding framework. To investigate fluctuations due to the changing environment, statistics are performed over numerous configurations of the MD simulation.

2.1 Molecular Dynamics

LAMMPS [36], V.11.08.2017 is used to perform the MD simulation along with the rigid TIP4P/2005 water force field [37]. 15625 rigid TIP4P/2005 water molecules are placed in a simulation box (approx $7.8 \times 7.8 \times 7.8$ nm^3) to form a 3D-periodic bulk system. We have used an isothermal, isobaric ensemble (NPT) with Nose-Hoover thermostat at 300 K ($\tau_T = 0.4$ ps) and Nose-Hoover barostat at 1 atm ($\tau_P = 2$ ps) [38]. 1 ns of equilibration is performed before the 1 ns production run, both with a time step of 2 fs. Both electrostatic and Lennard-Jones intermolecular interactions are computed using the long-range Particle–Particle–Particle–Mesh (PPPM) formalism [39,40]. neighbor lists are updated every time step within a radius of 10 Å. This simulation leads to a density of about 0.996 kg/L. Such a large system was necessary to investigate environmental effects up to 40 Å.

2.2 First hyperpolarizability in the liquid phase

To compute the first hyperpolarizability of water molecules within the liquid phase, an explicit environment composed of point charges is used : the Polarizable Embedding model at the zeroth-order (PE0) [41,42], implemented in the DALTON software [43], release 2018.2 package. As in our previous recent work [32], the QM calculations were carried out on individual water molecules. Point charges represent the surrounding water molecules, and the same electrostatic description is used as for the MD (TIP4P/2005 model). The MD trajectories are used directly in the QM calculations without further optimization. We define the parameter R_c as the maximal distance upon which neighbors are included in the PE formalism. The electrostatic field generated by this environment is spatially-heterogeneous in the vicinity of the target molecule and creates a new potential for the target molecule electronic Hamiltonian.

The electrostatic description could be refined by adding dipole, quadrupole or polarizability of the environment [42,44–47], but we shall see in the results that Eq.2 becomes relevant only for long-range effects, where the point-charges electrostatic field is dominant. We have therefore chosen to keep the environment description as simple and robust as possible. To optimize parts of the QM/MM routines, the home-made software FROG is used [*].

The whole procedure is composed of the following steps: (1) Perform MD simulation using LAMMPS. (2) Read the MD trajectories using FROG to build the electrostatic environment for each water molecule defined into a sphere of radius R_c centered on the mass center of the target molecule. If a neighbor possesses at least one atom in the sphere, the whole molecule is considered in the environment. The output files for DALTON are created for 300 molecules for 8 selected MD configurations. (3) For each of the 2400 target molecules, calculate the electrostatic field, and its gradients, generated by the environment using FROG. (4) For

each of the 2400 target molecules, calculate the value of β at the fundamental wavelength of 800 nm using DFT with the functional CAM-B3LYP[48] and the basis d-aug-cc-pVTZ[49], using the quadratic response formalism as implemented in DALTON[50]. (5) Read all DALTON output files and analyze the data using FROG.

The largest inaccuracy of our approach is due to the QM method and the corresponding functional. For the first hyperpolarizability of water molecules, the values in vacuum obtained with our method (DFT/CAM-B3LYP) are typically overestimated by about 10 %[32] in comparison to the golden standard for QM/MM calculation of small molecules (CCSD[51–53]) The basis set d-aug-cc-pVTZ[49] used is widespread across the community[26,51–53] and present a good balance between accuracy and cost. We have not taken into account the nuclear degree of freedom (vibrational effects) since Beaujean *et. al.*[54] have shown that this effect is small for the second hyperpolarizability of water, especially at small wavelength.

2.3 Second Hyperpolarizability in the liquid phase

In this work, we are interested in the evolution of the first hyperpolarizability under a static and homogeneous electrostatic field, which is reflected by $\gamma(2\omega,\omega,\omega,0)$ noted here γ. For each individual molecule, the second hyperpolarizability γ can be computed from the first hyperpolarizability β using the Finite Field (FF) framework. For that, we use the PE formalism described above to compute β when an extra homogeneous electrostatic field is applied, and then compute γ according to

$$\gamma_{ijkl} = \frac{\delta\beta_{ijk}}{\delta e_l} \simeq \frac{\beta_{ijk}(e_l) - \beta_{ijk}(0)}{e_l}, \qquad (4)$$

where $\beta_{ijk}(e_l)$ is the first hyperpolarizability of a molecule on which an extra static electrostatic field e_l is applied along the l molecular direction. Numerically, the γ_{ijkl} is obtained for each molecule as the slope of a linear fit performed on seven calculations with e_l from 0 to 1.5×10^{-3} a.u. Noticeably, during the γ calculation, the target molecule is embedded in a total electrostatic field composed of a spatially-heterogeneous contribution from its PE environment plus the spatially-homogeneous e_l. More details regarding the FF procedure can be found in the Electronic Supplementary Information (ESI), Sect. S1.1. The first and second hyperpolarizabilities are given in atomic units, our conventions are defined by Eq. 2 and 4. They are expressed in the molecular frame $\{a,b,c\}$ with the axis c along the dipole moment pointing from the Oxygen towards the Hydrogens, and the atoms within the $\{a,c\}$ plane see insert in Table 1.

Frequency dispersion can be important for γ : for example, Beaujean *et.al.*[54] have shown a variation of about 25 % between an excitation wavelength of 800 nm and at the infinite wavelength limit for $\gamma(3\omega,\omega,\omega,\omega)$. Here, even if we are using a Finite Field approach, the frequency dispersion of β related to the fundamental frequency (here 800 nm) is correctly reproduced since we use the first hyperpolarizability obtained using the frequency dependent response scheme.

2.4 Embedding Environment size: R_c

One of the objectives of this study is to understand the impact of the electrostatic environment on the first hyperpolarizability, distinguishing the effects of close and far neighbors. Therefore, different values of R_c are used for β calculations. In the following, the notation $\beta^{PE}(R_c)$ denotes the hyperpolarizability obtained by a QM/MM calculation at a given R_c and $\mathbf{e}(R_c)$ the electric field generated by this PE environment. In this notation, $R_c = 0$ corresponds to a gas phase calculation. According to previous works[32,53,55], a large environment (typically between 10 and 20 Å) is necessary to obtain a good convergence of the β values. Here, for each molecule, a reference calculation is performed using the PE formalism with an extremely large radius of $R_f = 40$ Å, and we consider that $\beta^{\mathrm{env}} = \beta_{PE}(40\mathrm{AA})$. Similarly, we consider that the electrostatic field $\mathbf{e}^{\mathrm{env}} = \mathbf{e}(40\text{Å})$. Note that such a large radius still does not represent the MD liquid phase, described using periodic boundary conditions.

On the contrary, the variation of γ with the size of the environment is not investigated in details here. Therefore, the PE radius used to compute the second hyperpolarizability γ is set to 10 Å. Yet, we have verified that the γ do not evolve much using larger radius in agreement with Osted *et al.*[55] (data not shown) and have noted it γ^{env}.

2.5 Statistical averaging

The first and second hyperpolarizability in the liquid phase are obtained from 2400 configurations of the PE environment. These configurations are extracted from 8 MD snapshots, with 300 water molecules randomly selected in each frame. These frames are separated by 100 ps to ensure time decorrelation[32]. We have verified that this number of configuray-tions is sufficient to reach convergence (see ESI, Figures S2 and S3); the numerical error due to the configuration number is smaller than the one due to the DFT by at least one order of magnitude. The C_{2v} symmetry of the first and second hyperpolarizability obtained by QM/MM confirms this good convergence (see the Section 3 and ESI Sec. S2.1).

3 Results

In the following, Section 3.1 first presents the second hyperpolarizability tensor γ calculated using QM/MM approaches for single water molecules within the liquid bulk phase, and the standard deviation of the calculated values. Then, Section 3.2.1 presents some difficulties encountered when applying Eq. 2 for condensed phases. We attribute them to the properties of the electrostatic field generated by the environment, $\mathbf{e}^{\mathrm{env}}$, described in 3.2.2. Based on these results, and on Eq. 2, in Sect. 3.2.3, we propose a correction to β that permits to include long-range interactions. Finally, we show that this correction enhances the precision of standard QM/MM calculations.

3.1 QM/MM results: second hyperpolarizability of water in the bulk phase

Table 1 presents the values of selected γ components calculated either in the vacuum or in liquid phase – the other non-zero components of γ are presented in ESI, Table S4.

Table 1 Water second hyperpolarizability $\gamma^{(3)}(2\omega,\omega,\omega,0)$ components, either in the vacuum (γ^{vac}) or in its liquid phase (γ^{env}) for an exciting wavelength of 800 nm, in atomic units. For the bulk phase, the average value $\langle\gamma^{\text{env}}\rangle$ and the standard deviation $\sigma[\gamma^{\text{env}}]$ are reported. The tensors components are expressed in the molecular frame $\{a,b,c\}$, see inserted figure. Comparing with response calculations, the error on γ due to the finite field differentiation is estimated at about 20 a.u. (see ESI, Table. S1).

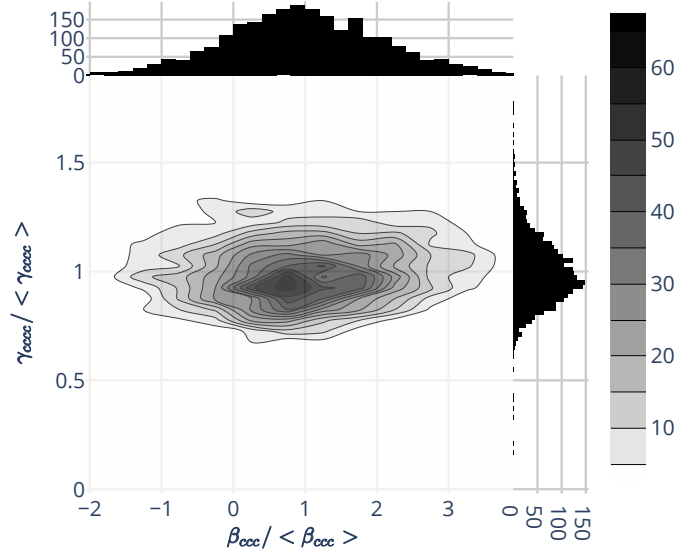|	Vacuum	Liquid	
$ijkl$	γ^{vac}	$\langle\gamma^{\text{env}}\rangle$	$\sigma[\gamma^{\text{env}}]$
$aaaa$	1110	740	140
$bbbb$	4090	2940	650
$cccc$	2000	1400	230
$aabb$	930	640	200
$bbaa$	1060	710	150
$baab$	1010	680	220
$abba$	940	660	140
$aacc$	590	370	60
$ccaa$	610	380	80
$caac$	610	380	60
$acca$	590	370	80
$bbcc$	1120	770	170
$ccbb$	1010	720	180
$cbbc$	1010	730	150
$bccb$	1060	750	200

Fig. 1 Join population of $\{\beta_{ccc}^{\text{env}}, \gamma_{cccc}^{\text{env}}\}$ obtained by QM/MM approaches for bulk liquid water. The independent ones are presented on the top and right for the β_{ccc}^{env} and $\gamma_{cccc}^{\text{env}}$ respectively. The values are normalized with respect to their average ones: $\langle\beta_{ccc}^{\text{env}}\rangle = 4.1$ a.u. and $\langle\gamma_{cccc}^{\text{env}}\rangle = 1400$ a.u.

For the vacuum phase, the molecular C_{2v} symmetry is fulfilled: the only non-vanishing components are the γ_{iiii} and γ_{iijj} for i and j in $\{a,b,c\}$, plus their permuted terms[1] (see ESI, Table S4). In the liquid phase, this symmetry is not valid for individual values, but is recovered for the average $\langle\gamma^{\text{env}}\rangle$. Our results agree with the ones by Osted et al.[55] within about 15%. We impute the discrepancies to both the difference of methods (DFT vs. CCSD, different MD models) and the difference of excitation wavelengths (800 vs. 1080 nm). More detailed comparisons with literature are provided in ESI, Sect. S1.1.

Noticeably, when transferred into the laboratory frame, the average second hyperpolarizability becomes centro-symmetric (see ESI, Table S5), indicating that our sampling of molecular orientation in the liquid phase is sufficient. Interestingly, the average value of all C_{2v}-authorized γ^{env} components are positive. Moreover, in the liquid phase, the γ^{env} components are about 30 % smaller than the ones in the vacuum – see ESI, Figure S4.

Beyond the average values, the environment induces fluctuations of γ^{env}, and of β^{env}. A typical and relevant case is illustrated by Figure 1, displaying the join probabilities of $\{\beta_{ccc}, \gamma_{cccc}\}$ obtained for the 2400 water molecules within their environment. Projections on the axes display the distributions of β_{ccc} and γ_{cccc} separately. As already discussed in the literature[32,53,55], the water first hyperpolarizability fluctuations in the liquid phase are strong, with standard deviations similar to the absolute values of the averages – for instance $\langle\beta_{ccc}^{\text{env}}\rangle = 4.1$ a.u. and $\sigma[\beta_{ccc}^{\text{env}}] = 3.1$ a.u. For individual molecules, some β components, null in average,

can be larger than the one with a net average (see ESI, Figure S5).

The second hyperpolarizability γ also fluctuates, but in a smaller extent than β: the standard deviations of the different components $\sigma[\gamma^{\text{env}}]$ are about 20 % of their average values $\langle\gamma^{\text{env}}\rangle$ – see Fig.1 for γ_{cccc}, and ESI, Table S4 for the other components.

Figure 1 moreover shows that the values of β and γ are not strongly correlated. Hence, in the following, we neglect the γ^{env} dispersion and attribute the same second hyperpolarizability $\langle\gamma^{\text{env}}\rangle$ to all the water molecules.

3.2 Using γ to calculate individual β in condensed phases

After describing the second hyperpolarizability γ^{env} in the liquid phase, we will question if it can be used to predict the first hyperpolarizability of water molecules in the liquid phase using different approximations inspired by Eq. 2. We will firstly present the limitations of the gas-phase reference model.

3.2.1 Limitations of the gas-phase reference model

Comparing Eq. 2 and Eq. 3, one straightforward approach to include the effect of the environment on the first hyperpolarizability β is to use the second hyperpolarizability γ and the field created by the environment $\mathbf{e}^{\text{env}}$:

$$\beta^{\text{GR}} = \beta^{\text{vac}} + \gamma^0 \cdot \mathbf{e}^{\text{env}}, \qquad (5)$$

where a linear perturbation is done relative to the Gas-phase Reference (GR), for which $\mathbf{e}^{\text{env}} = \mathbf{0}$ and $\beta = \beta^{\text{vac}}$, the first hyperpolarizability obtained for a single water molecule in vacuum. If

the linear proportionality factor γ^0 is known, and the same for all molecules, this approach would be very helpful: the liquid first hyperpolarizability depends only on $\mathbf{e}^{env}$. An MD simulation can provide $\mathbf{e}^{env}$ for each molecule, and the liquid phase first hyperpolarizability is calculated using Eq. 5, so that no expensive QM/MM approach is needed. Moreover, such an approximation is commonly used in the Surface-SHG community (for example Refs. 18,19).

Liang et al.[53] have reported that Eq. 5 using the vacuum value for the second hyperpolarizability $\gamma^0 = \gamma^{vac}$ is a poor estimator for β^{env}. We have thus applied Eq. 5 using the average bulk second hyperpolarizability tensor $\gamma^0 = \langle \gamma^{env} \rangle$ for all the molecules (values from Table 1 and from ESI, Table S4), and the electrostatic field generated by the environment $\mathbf{e}^{env}$. For liquid water, the distribution of the largest component of $\mathbf{e}^{env}$, along the water molecular axis c is represented on Fig. 2: it has typical values around 3×10^{-2} a.u., i.e. 1.6 V.Å^{-1}. The other components average to zero, with approximately the same standard deviations (data not shown).
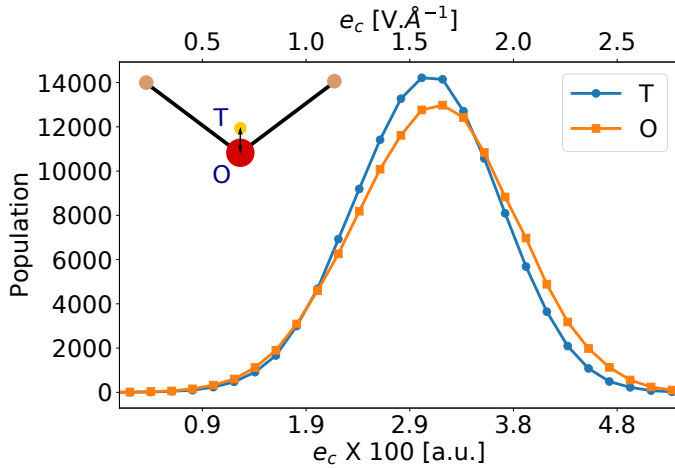

Fig. 2 Distribution of the electrostatic field generated by the environment $\mathbf{e}^{env}$ along the molecular c-axis. The electrostatic field is measured at 2 positions: at the oxygen atom position (O: orange squares), or at the negative charge one (T: blue dots) according to the TIP4P/2005 force field. The O and T positions are distant by 0.15 Å.

The predicted average values $\langle \beta^{GR} \rangle$ are compared to the QM/MM values $\langle \beta^{env} \rangle$ in Table 2. First, the addition of the linear correction in Eq. (5) allows recovering signs for β^{GR} components in agreement with β^{env}. However, the β^{GR} model largely overestimates the reference values: the correction $\gamma^0 \cdot \mathbf{e}^{env}$ should be smaller. This overshooting is even worse when the vacuum value γ^{vac} is used for γ^0, as Liang et al. had tested[9]. Thus, even if it contains some relevant physics, Eq. 5 describing a correction that is proportional to $\mathbf{e}^{env}$ is not workable: it predicts an average hyperpolarizability $\langle \beta^{GR} \rangle$ one order of magnitude too large.

Moreover, the value of $\mathbf{e}^{env}$ in Eq. 5 is ill-defined because of its strong spatial heterogeneity. To highlight this fact, Figure 2 presents the electrostatic field created by the environment along the molecular c-direction, $\mathbf{e}_c^{env}$, calculated at 2 positions within

Table 2 Comparison between relevant QM/MM components of the first hyperpolarizability, β^{env}, the vacuum ones, β^{vac}, and β^{GR} predicted by Eq. 5 with $\gamma^0 = \langle \gamma^{env} \rangle$ from Table 1. All values are in atomic units and are calculated for fundamental wavelength of 800 nm. The values of $\mathbf{e}^{env}$ obtained at the center-of charge position have been used.

ijk	$\langle \beta^{vac} \rangle$	$\langle \beta^{env} \rangle$	$\langle \beta^{GR} \rangle$
ccc	-15.3	4.1	28.5
caa	-12.5	-2.0	-0.6
aca	-12.4	-2.0	-0.8
cbb	-5.0	2.5	17.7
bcb	-7.4	2.2	16.9

the embedded molecule: either at the position of the negative charge of the TIP4P/2005 model, or at the position of the oxygen atom. While the 2 distributions look very similar, a difference on the average values of about 10^{-3} a.u. appears. Given the large value of γ_{cccc} (≈ 1400 a.u., see table 1), the component β_{ccc} predicted by β^{GR} changes by about 40%, depending on where $\mathbf{e}^{env}$ is calculated. Therefore, we attribute the failures of Eq. 5 to the strong intensity and large gradient of the embedding field $\mathbf{e}^{env}$, which are incompatible with a simple linear expansion.

3.2.2 Separation into short and long-range neighbors

The evolution of the electrostatic field $\mathbf{e}(R_c)$ generated by the environment around the target molecule of the QM/MM calculations with the environment size R_c is presented on Figure 3. The main contribution to this electrostatic field is created in the first and second solvation shells, at distances between 2 and 5 Å from the target molecule. This neighborhood region also creates the largest electrostatic field gradients, see ESI, Figure S6.
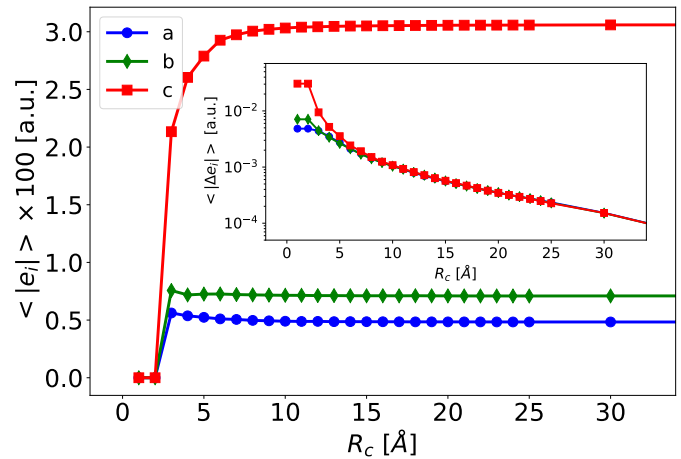

Fig. 3 Averaged norm of the electrostatic field generated by the electrostatic environment ($\mathbf{e}(R_c)$) felt at the T-position, for increasing environment size R_c. In insert, the difference $\Delta \mathbf{e}(R_c)$ relative to the value at $R_c = 40$ Å, in logarithmic scale as a function of the size R_c.

Given that the closest shells are responsible for both the intensity and the heterogeneity of $\mathbf{e}^{env}$, we divide the neighborhood

into two parts, as illustrated on Figure 4.

The *direct area* contains the closest neighbors of the target molecule (at the center), up to a distance R_c: they create a strong and heterogeneous electric field, noted $\mathbf{e}(R_c)$. The *long-range area*, due to neighbors beyond R_c from the target molecules, creates a less intense and more homogeneous electric field on the target molecule.
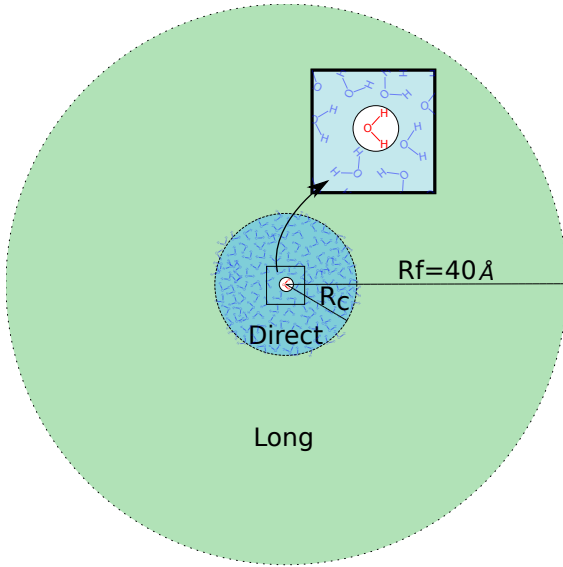

Fig. 4 Sketch of the electrostatic embedding procedure. The QM box is defined for only one molecule of water at the center. Direct area, until R_c: all the neighbors are included in a QM/MM calculation with discrete solvation procedure, such as the PE formalism. Long-range area, from R_c to R_f: the neighbors are included by the electrostatic field they generate at one specific point of the QM box.

The total embedding electrostatic field $\mathbf{e}^{\text{env}}$ is thus separated into a short-range contribution $\mathbf{e}(R_c)$, and a correction, $\Delta\mathbf{e}(R_c)$ due to the long-range electrostatic interactions:

$$\mathbf{e}^{\text{env}} = \mathbf{e}(R_c) + \Delta\mathbf{e}(R_c). \tag{6}$$

where all the fields are evaluated at the position of the negative charge in the TIP4P/2005 model (T-position, see Figure 2 inset). The evolution of $\Delta\mathbf{e}(R_c)$ as a function of R_c is available in the insert of Fig. 3. At $R_c = 10\,\text{Å}$, the electrostatic field generated by the long-range neighbors is very small compared to the one created by close neighbors: the average correction $\Delta\mathbf{e}(R_c = 10)$ is close to 10^{-3} a.u. In addition, we have investigated whether the correction $\Delta\mathbf{e}(R_c)$ is spatially homogeneous by measuring the correction to the electric field gradient $\Delta\frac{\delta\mathbf{e}}{\delta\mathbf{x}}(R_c)$ (see ESI, Fig. S6). Generally, the corrections to the gradients are small (typically 10^{-4} a.u) for neighbors further away than $10\,\text{Å}$.

Therefore, the correction $\Delta\mathbf{e}(R_c)$ is more homogeneous than $\mathbf{e}^{\text{env}}$, and hardly depends on the point on which it is calculated. In the following, we use the value of $\Delta\mathbf{e}(R_c)$ evaluated at the T-position.

3.2.3 Long-range correction to β using γ

The further the neighbor is from the target molecule, the better its contribution to $\mathbf{e}^{\text{env}}$ can be modeled as a weak and homogeneous electrostatic field. A linear γ correction to β could be tested to describe the long-range electrostatic effects.

For the first hyperpolarizability, as for $\mathbf{e}^{\text{env}}$, we describe the impact of the environment in two different ways depending on the neighbors distance. The *direct neighbors* within the sphere of radius R_c are included explicitly in the QM/MM calculation through the PE approach. The *long-range neighbors*, beyond R_c, are included implicitly via the homogeneous field they produce on the target molecule $\Delta\mathbf{e}(R_c)$ with a correction proportional to $\langle\gamma^{\text{env}}\rangle$. This approximation can be viewed as a linear expansion relative to the value of first hyperpolarizability already perturbed by the direct neighbors, $\beta^{PE}(R_c)$:

$$\beta^{\text{PE+L}}(R_c) = \beta^{PE}(R_c) + \langle\gamma^{\text{env}}\rangle \cdot \Delta\mathbf{e}(R_c). \tag{7}$$

In the following, we compare this approximation of β, noted $\beta^{\text{PE+L}}(R_c)$ and the usual QM/MM calculation where only the PE scheme up to a distance R_c is used, noted $\beta^{PE}(R_c)$. To estimate the accuracy of the two approximations PE and PE+L as a function of R_c, we compare them to our reference values β^{env}.

As a typical example, we compare on Figure 5 the β_{ccc} computed with $R_c = 5\,\text{Å}$ either using PE approximation (top) or the PE+L one (bottom) to the reference value β_{ccc}^{env}. The diagonal dashed lines represent the case where the approximation reproduces exactly the reference. For the usual calculation (β^{PE}), there is a systematic error leading to inaccuracy: at $R_c = 5\,\text{Å}$, the average value of β_{ccc}^{PE} is smaller than the one of β_{ccc}^{env}. Moreover, the error fluctuates a lot, it can reach tens of a.u., depending on the molecule. On the contrary, the PE+L approach almost quantitatively reproduces the reference, even for R_c as small as $5\,\text{Å}$

To illustrate the effect of R_c, Figure 6 depicts the evolution of $\langle\beta_{PE}\rangle$ and $\langle\beta_{PE+L}\rangle$ as a function of R_c for the *ccc* component (similar trends are obtained for other components). As expected, the inaccuracy of the PE and PE+L approaches decreases when R_c increases. When the correction is added, the convergence is reached faster ($R_c = 4\,\text{Å}$) compared to the usual β^{PE} calculation ($R_c \simeq 12\,\text{Å}$). The effect of the neighbors further than $5\,\text{Å}$ can thus be very efficiently included by a second hyperpolarizability corrective term. Beyond the description of the average first hyperpolarizability, Figure 5 also illustrates that the PE+L approach improves the precision of the individual predictions. This can be quantified through the mean absolute error (MAE), averaged over all the $N = 2400$ molecular configurations and all the 27 components, noted $\Delta\beta_T$:

$$\Delta\beta_T^X = \frac{1}{27N}\sum_{ijk}\sum_{n=1}^{N}|\beta_{ijk}^{\text{env}(n)} - \beta_{ijk}^{X(n)}| \tag{8}$$

where X stands for the approximation *PE*, or *PE + L*, and the exponent (n) indicates that the values of the *n*-th molecule are used. Figure 7 report the evolution of $\Delta\beta_T^X$ as a function of R_c.

The long-range correction provides very good results and reaches an error below 0.1 a.u. at $R_c = 12\,\text{Å}$. To obtain the same

Fig. 6 Average of the $\beta_{ccc}^X(R_c)$ component as a function of R_c for the *PE* (blue disks) and *PE*+*L* (orange squares, see Eq. 5) approximations. The dashed line is the reference value β^{env} obtained for $R_f = 40$ Å .
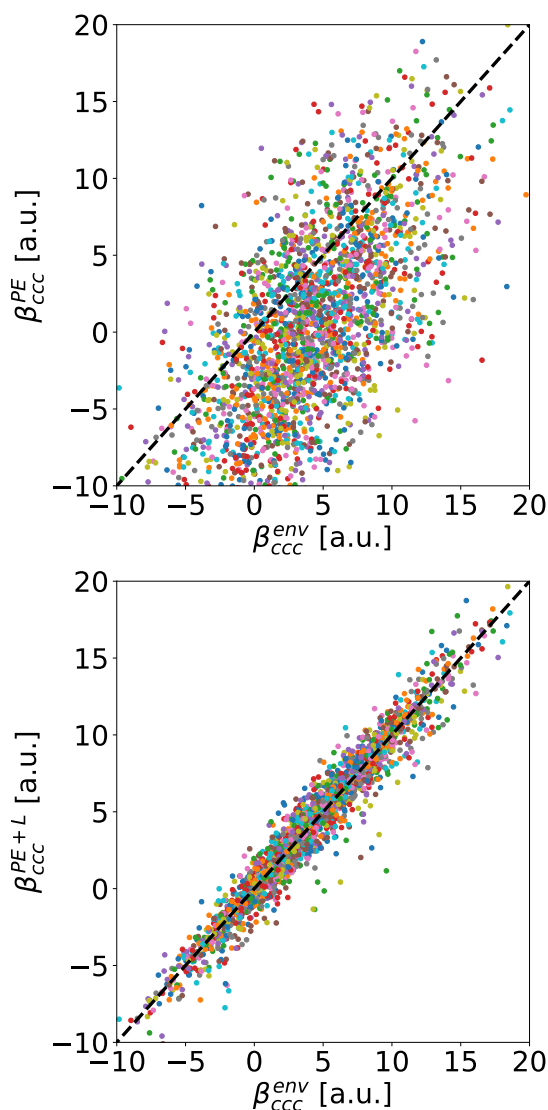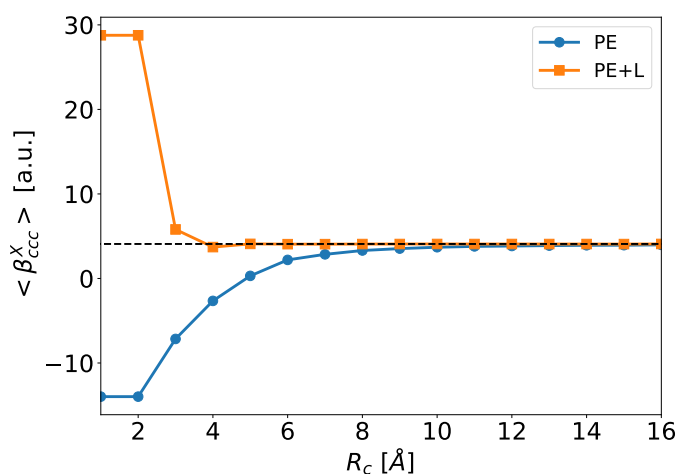
Fig. 5 Comparison of the reference individual molecular β_{ccc}^{env} component and the one obtained using *PE* (TOP) and *PE*+*L* (BOTTOM) approximation with $R_c = 5$ Å . The dashed line corresponds to the ideal case where a correction matches perfectly the expected value. Each dot represents a molecule.
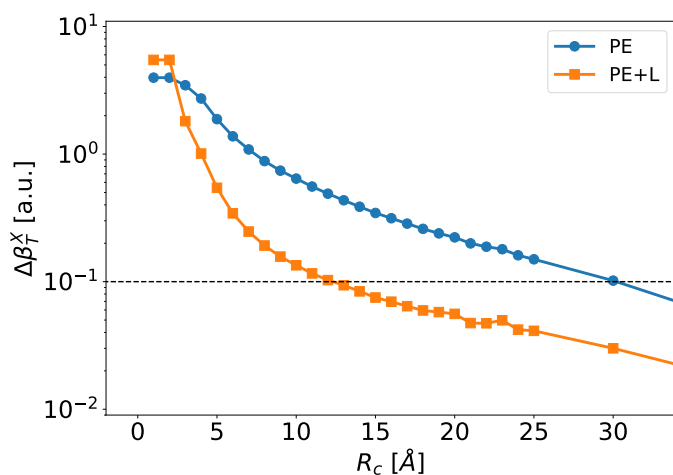
Fig. 7 Mean absolute error of the individual first hyperpolarizability tensor using Eq. 8 as a function of R_c for the *PE* (blue disks) and *PE*+*L* (orange squares) approximations. The dashed line corresponds to an error of 0.1 a.u. at which we define convergence.

degree of accuracy, one would need to include explicitly neighbors up to $R_c \simeq 30$ Å in the usual *PE* scheme without long-range correction. For R_c larger than 10 Å , the error obtained with the long range correction (*PE*+*L*) is about one order of magnitude lower than the one with the traditional *PE* approach. This shows that – once the strong short-range effects are explicitly included – it is possible to use the knowledge of γ to predict accurately the long-range electrostatic effects on β . Note that the long-range correction is proportional to the average value $\langle\gamma^{\text{env}}\rangle$, i.e. Eq. 7 neglects γ^{env} fluctuations. As discussed further in ESI, Sec. S4, including the fluctuations of γ^{env} in Eq. 7 only slightly improves the results. However, it implies to compute γ^{env} for every molecule, which greatly increases the numerical cost. Therefore, using an average γ^{env} is a very good compromise between computational

cost and accuracy, at least for bulk liquid water.

4 Conclusions

In this work, we have computed the first and second hyperpolarizability of water in the liquid phase using state-of-the-art QM/MM calculations. With this set of values, several approaches were tested to take into account the electrostatic environment in the first hyperpolarizability.

The second hyperpolarizability is high in the bulk phase. The neighborhood induces a dispersion on the second hyperpolarizability $\gamma(2\omega,\omega,\omega,0)$ values, but less than for the first hyperpolarizability.

Yet, this second hyperpolarizability obtained for water in its liquid phase cannot be used directly to predict the first hyperpolarizability from the electric field created by the whole environment.

Indeed, using a linear correction relative to the gas phase leads to overshot averages. We attribute this inaccuracy to the fact that linear expansions are valid for a weak and spatially-homogeneous electrostatic field, while the embedding electrostatic field created by the liquid surroundings is not. To compute individual molecular hyperpolarizability in the liquid phase, the molecular structure of the nearest neighbors have to be described explicitly, and QM/MM models are very relevant.

But long-range effects are more difficult to take into account using QM/MM approaches, since a very large number of neighbors would have to be included. We have proposed to consider the effect of long-range electrostatic environment on the liquid first hyperpolarizability using the average second hyperpolarizability γ. This long-term correction increases the precision of the hyperpolarizability calculation, and speeds up the convergence of the average value relative to the QM/MM environment size R_c. For pure bulk water, our approximation makes sense for a long-range part defined beyond 5 to 10 Å, depending on the quantity required. From a numerical point of view, the correction can be powerful for charged environments because the size of the explicit embedding region can be reduced drastically.

The present study is applied to pure water bulk phase at 300 K ; however, the methodology developed here should be also relevant for many other system geometries or compositions, where the long-range electrostatic effects can modify optical responses. The major drawback of this method is that it requires the calculation of $\langle \gamma^{env} \rangle$, which may be system-dependent and time-consuming. Ongoing work to overcome this numerical issue is under development in our home-made software FROG.

This work has demonstrated the promising potentialities of using the second hyperpolarizability to predict first hyperpolarizabilities in environments with long-range electrostatic effects. It opens up perspectives to model the NLO response of water in salted aqueous solutions, or nearby charged interfaces.

Author Contributions

GLB : Conceptualization, Methodology, Validation, Writing. OB: Supervision, Validation, Writing. EB: Supervision, Writing, review, editing. CL: Supervision, Validation, Writing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 P. F. Brevet, 1997.

2 M. C. Goh, J. M. Hicks, K. Kemnitz, G. R. Pinto, K. Bhattacharyya, K. B. Eisenthal and T. F. Heinz, Journal of Physical Chemistry, 1988, **92**, 5074–5075.

3 Q. Wei, D. Zhou and H. Bian, Phys. Chem. Chem. Phys., 2018, **20**, 11758–11767.

4 M. N. Nasir, E. Benichou, C. Loison, I. Russier-Antoine, F. Besson and P.-F. Brevet, Physical Chemistry Chemical Physics, 2013, **15**, 19919–19924.

5 G. Licari, J. S. Beckwith, S. Soleimanpour, S. Matile and E. Vauthey, Phys. Chem. Chem. Phys., 2018, **20**, 9328–9336.

6 E. Donohue, S. Khorsand, G. Mercado, K. M. Varney, P. T. Wilder, W. Yu, A. D. MacKerell, P. Alexander, Q. N. Van, B. Moree, A. G. Stephen, D. J. Weber, J. Salafsky and F. McCormick, Proceedings of the National Academy of Sciences of the United States of America, 2019, **116**, 17290–17297.

7 M. E. Didier, O. B. Tarun, P. Jourdain, P. Magistretti and S. Roke, Nature Communications, 2018, **9**, 1–7.

8 S. Bancelin, C. Aimé, I. Gusachenko, L. Kowalczuk, G. Latour, T. Coradin and M. C. Schanne-Klein, Nature Communications, 2014, **5**, 1–8.

9 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, Phys. Rev. B, 2017, **96**, 1–6.

10 P. Kaatz, E. A. Donley and D. P. Shelton, Journal of Chemical Physics, 1998, **108**, 849–856.

11 J. Yang, E. V. Barnat, S.-k. Im and D. B. Go, Journal of Physics D: Applied Physics, 2021.

12 B. Levine and C. Bethea, The Journal of Chemical Physics, 1976, **65**, 2429–2438.

13 T. N. Ramos, S. Canuto and B. Champagne, Journal of Chemical Information and Modeling, 2020, **60**, 4817–4826.

14 B. F. Levine and C. G. Bethea, The Journal of Chemical Physics, 1975, **63**, 2666–2682.

15 E. C. Yan, Y. Liu and K. B. Eisenthal, J. Phys.Chem. B, 1998, **102**, 6331–6336.

16 T. Joutsuka, T. Hirano, M. Sprik and A. Morita, Phys. Chem. Chem. Phys., 2018, **20**, 3040–3053.

17 S. Pezzotti, D. R. Galimberti, Y. R. Shen and M.-P. Gaigeot, Physical Chemistry Chemical Physics, 2018, **20**, 5190–5199.

18 L. Dalstein, K.-Y. Chiang and Y.-C. Wen, The Journal of Physical Chemistry Letters, 2019, **10**, 5200–5205.

19 A. Marchioro, M. Bischoff, C. Lütgebaucks, D. Biriukov, M. Předota and S. Roke, The Journal of Physical Chemistry C, 2019, **123**, 20393–20404.

20 R. Hartkamp, A. L. Biance, L. Fu, J. F. Dufrêche, O. Bonhomme and L. Joly, Current Opinion in Colloid and Interface Science, 2018, **37**, 101–114.

21 L. B. Dreier, C. Bernhard, G. Gonella, E. H. Backus and M. Bonn, The Journal of Physical Chemistry Letters, 2018, **9**, 5685–5691.

22 E. Ma, P. E. Ohno, J. Kim, Y. Liu, E. H. Lozier, T. F. Miller, H.-F. Wang and F. M. Geiger, J. Phys. Chem. Lett., 2021, **12**, 5649–5659.

23 H. Chang, P. E. Ohno, Y. Liu, E. H. Lozier, N. Dalchand and F. M. Geiger, J. Phys.Chem. B, 2020, **124**, 641–649.

24 Y. Foucaud, B. Siboulet, M. Duvail, A. Jonchere, O. Diat, R. Vuilleumier and J.-F. Dufrêche, Chem. Sci., 2021, **12**, 15134–15142.

25 G. Maroulis, Chem. Phys. Lett., 1998, **289**, 403–411.

26 J. Kongsted, A. Osted, K. V. Mikkelsen and O. Christiansen, J. Chem. Phys., 2004, **120**, 3787–3798.

27 M. Hidalgo Cardenuto and B. Champagne, Phys. Chem. Chem. Phys., 2015, **17**, 23634–23642.

28 T. Giovannini, M. Ambrosetti and C. Cappelli, Theoretical Chemistry Accounts, 2018, **137**,.

29 P. Beaujean and B. Champagne, The Journal of Chemical Physics, 2019, **151**, 064303.

30 T. N. Ramos, F. Castet and B. Champagne, J. Phys.Chem. B, 2021, **125**, 3386–3397.

31 L. Jensen and P. T. Van Duijnen, J. Chem. Phys., 2005, **123**,.

32 G. Le Breton, O. Bonhomme, P.-F. Brevet, E. Benichou and C. Loison, Phys. Chem. Chem. Phys., 2021, **23**, 24932–24941.

33 A. Willetts, J. E. Rice, D. M. Burland and D. P. Shelton, The Journal of Chemical Physics, 1992, **97**, 7590–7599.

34 O. Bonhomme, L. Sanchez, E. Benichou and P. Brevet, The Journal of Physical Chemistry B, 2021, **125**, 10876–10881.

35 A. Pardon, O. Bonhomme, C. Gaillard, P.-F. Brevet and E. Benichou, Journal of Molecular Liquids, 2021, **322**, 114976.

36 S. Plimpton, Journal of Computational Physics, 1995, **117**, 1–19.

37 J. L. Abascal and C. Vega, The Journal of Chemical Physics, 2005, **123**, 234505.

38 G. J. Martyna, D. J. Tobias and M. L. Klein, The Journal of chemical physics, 1994, **101**, 4177–4189.

39 R. E. Isele-Holder, W. Mitchell and A. E. Ismail, Journal of Chemical Physics, 2012, **137**,.

40 R. E. Isele-Holder, W. Mitchell, J. R. Hammond, A. Kohlmeyer and A. E. Ismail, Journal of Chemical Theory and Computation, 2013, **9**, 5412–5420.

41 J. M. H. Olsen and J. Kongsted, J. Kongsted, Academic Press, 2011, vol. 61, pp. 107–143.

42 E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, Journal of Chemical Theory and Computation, 2018, **14**, 4309–4319.

43 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle et al., Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014, **4**, 269–284.

44 N. H. List, H. J. A. Jensen and J. Kongsted, Phys. Chem. Chem. Phys., 2016, **18**, 10070–10080.

45 M. T. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted and J. M. H. Olsen, Journal of Chemical Theory and Computation, 2016, **12**, 1684–1695.

46 C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted and J. M. H. Olsen, International Journal of Quantum Chemistry, 2019, **119**, e25717.

47 A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted and C. Hattig, Journal of Chemical Theory and Computation, 2020, **16**, 1373–1381.

48 T. Yanai, D. P. Tew and N. C. Handy, 2008, **393**, 51–57.

49 R. A. Kendall, T. H. Dunning and R. J. Harrison, The Journal of Chemical Physics, 1992, **96**, 6796–6806.

50 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, The Journal of Chemical Physics, 2002, **117**, 9630–9645.

51 P. Besalú-Sala, S. P. Sitkiewicz, P. Salvador, E. Matito and J. M. Luis, Physical Chemistry Chemical Physics, 2020, **22**, 11871–11880.

52 P. Beaujean and B. Champagne, Theoretical Chemistry Accounts, 2018, **137**, 50.

53 C. Liang, G. Tocci, D. M. Wilkins, A. Grisafi, S. Roke and M. Ceriotti, Physical Review B, 2017, **96**, 1–6.

54 P. Beaujean and B. Champagne, The Journal of Chemical Physics, 2019, **151**, 064303.

55 A. Osted, J. Kongsted, K. V. Mikkelsen, P. O. Åstrand and O. Christiansen, Journal of Chemical Physics, 2006, **124**, 124503.