Check for updates

Decomposition of molecular properties

Hans Ågren, [ID] *[ab] Ignat Harczuk[a] and Olav Vahtras [ID] [a]

We review recent work on property decomposition techniques using quantum chemical methods and discuss some topical applications in terms of quantum mechanics-molecular mechanics calculations and the constructing of properties of large molecules and clusters. Starting out from the so-called LoProp decomposition scheme [Gagliardi *et al.*, *J. Chem. Phys.*, 2004, **121**, 4994] for extracting atomic and inter-atomic contributions to molecular properties we show how this method can be generalized to localized frequency-dependent polarizabilities, to localized hyperpolarizabilities and to localized dispersion coefficients. Some applications of the generalized decomposition technique are reviewed – calculations of frequency-dependent polarizabilities, Rayleigh scattering of large clusters, and calculations of hyper-polarizabilities of proteins.

1 Introduction

The materials modelling community has during recent years increasingly turned its attention to multi-scale modelling. As time has passed, the replacement of the idea of one theory covering everything by the concept of bridging theories (or models) that each was limited to a certain scale in space and time, has become progressively more accepted. The most important variant of contemporary multi-scale modelling is given by the combination of quantum mechanics and classical physics, which in a sense gives the possibility to join the accuracy and rigour of the former with the applicability of the latter and which gives a possibility to find working approaches that efficiently can address the nanoscale. Thus, the joining of quantum mechanics (QM) with molecular mechanics (MM) with an expedient classical force field description of atoms has become a most important and popular area in contemporary *in silico* research with a wide variety of applied research areas. In the QM/MM approach the major types of intermolecular interactions are in principle accounted for both within and in-between the shells; electrostatic, polarization, van der Waals and short range (often though a QM/MM potential) and are mostly atomically decomposed. More recent development has concerned, in addition to structure and dynamics, also the "3rd dimension", namely general properties, thus including optical properties, and also the development of QM/MM programs including multi QM cores in homogeneous and heterogeneous environments. Applications cover various types of spectroscopy and linear and non-linear properties of molecules in solution, on surfaces, in confined biological environments or in combinations of such environments. A key formulation of the response theory in a QM/MM setting is the work by Nielsen *et al.*,[1] which has been diversified to various linear and non-linear electric and magnetic properties,[2,3] to QM/MM/PCM,[4] where a third layer constituting a dielectric continuum is added, and to resonant convergent response property formulations.[5] The QM/MM response property theory has been implemented for different types of QM cores, like multi-configurational self-consistent field (MCSCF), coupled cluster and DFT. The various QM/MM methods can also be distinguished from one another by the way in which they describe the coupling between the QM and MM regions. The mainstream implementations are of DFT/MM type where originally the DFT to MM coupling has been limited to the description of the electrostatic interaction between the electron density of the QM region and point charges of the MM region.[6] A more sophisticated description of the interaction between the MM and QM regions is achieved in the so-called effective fragment potential (EFP) method in which the coupling between the QM and MM regions includes electrostatic, polarization/induction, exchange repulsion, and a charge transfer interaction terms.[7–9] For general response properties, the more recent research has focused on the DFT/MM response formalism which is capable of computing arbitrary molecular properties expressed *via* linear, quadratic and cubic response functions and their residues,[10] and which includes full electrostatic, polarization/induction and dispersion terms in the description of the ground state interaction between a solute and its environment and the polarization/induction interaction in the generation of the response functions.[1] It allows to consider a ladder of sophistication for the coupling between the QM and MM regions, referring to truncating the MM

[a] *KTH Royal Institute of Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, Department of Theoretical Chemistry and Biology, SE-106 91 Stockholm, Sweden. E-mail: hagren@kth.se*
[b] *Tomsk State University, 36 Lenin Avenue, 634050, Tomsk, Russian Federation*

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | 2251

molecule or fragment electrostatic potential at different orders of the multipole expansion[11] and by selecting different forms of the polarizability tensor for the MM molecule or the fragment.[12–14] Various extensions from this outset have been made, for instance, to spin Hamiltonian parameters and paramagnetic open shell system,[2,15] allowing for systematic studies of, electronic g tensors and hyperfine coupling constants of various radicals and spin labels.[15] One can here also mention the recent extension of the QM/MM capacitance–polarization force field for embedding metal surfaces and/or nanoparticles in the metallic MM region in complex environments[16]—the so-called quantum mechanics/capacitance molecular mechanics (QM/CMM) model which provides a practical approach for theoretical modelling of properties of physisorbed chromophores on metal surfaces or nanoparticles. Here, the heterogeneous MM part is split into metallic and non-metallic parts, assuming a capacitance–polarization model for the electrostatics and polarization of the metallic part and distributed charges for the non-metallic part.[17,18] The implementation of a heterogeneous MM region captures the essential physical features of the composite containing metal surfaces and/or nanoparticles and physisorbed chromophores in vacuum or solutions.

Force fields, either used in the preceding classical MD simulation or in the MM part of QM/MM, constitute crucial quantities as they dictate the precision and the cost, and therefore the limitations and applicability of the QM/MM calculations. It follows that force fields have been the subject of ever ongoing efforts of refinement and analysis where both the results of more accurate models, *i.e.* quantum mechanics based, and of fitting to experimental data are employed. Largely, such force fields divide into electrostatic, polarizable and reactive force fields regulating charge transfer in the molecular mechanics medium. With the implementation of QM/MM models the quality of force fields are also scrutinized in terms of the interaction between the quantum and classical parts. One has during recent time witnessed new techniques to generate force fields for QM/MM calculations and techniques for integrating dynamics into the property calculations through QM/MM—the so-called integrated approach.[19] An ultimate aim is to use the same QM/MM parameterization for the dynamics as for the property calculations, a next level in ambition is to use the same force field parameterization in classical molecular dynamics as the one used in the MM part of the QM/MM property calculation. None of these two ambitions, are, however, fulfilled in common calculations published in literature. In some sense the requirement of the MM force field in QM/MM is higher than for the force field in pure MD, especially the paramount role of polarization and the need for a fine granulation of the force field. Thus even for the water molecule there is need to granulate the charges of the molecule as well as to granulate the polarizabilities. This puts additional demands in that the quality of the force fields that should balance the precision of the quantum part for the evaluation of the structural dynamics or properties. More recent implementations of properties and spectroscopy in the QM/MM framework often demand full granulation of such embeddings in terms of separate atomic

contributions, sometimes even covering atom–bond partitioning and multipole expansions of the charges. An often attended vehicle to fulfil this demand is the LoProp method[20] for extracting atomic and interatomic contributions to molecular properties. It was originally formulated as a sequence of transformations of the atomic overlap matrix, giving a localized orthonormal basis that depends only on molecular structure and the initial atomic orbital basis set. The localized properties are then obtained by forming partial traces over basis functions associated with one chosen centre or a pair of centres. Thus by means of LoProp several systematically connected models for the solvent interaction can be formulated, ranging from a simple molecular point charge to bond- and atom-centred multipoles and polarizabilities. The precision needed, and thus the level of decomposition, is to a large extent dictated by the property that is to be calculated. This level also has a bearing on the standard problem how to split the QM and MM parts of the system in a QM/MM calculation, for instance, the need to bring in water molecules in the QM part is often relieved by a careful LoProp MM parameterization of water.[3] To the known limitations of the LoProp method one can mention the tendency to "overpolarize" the interaction, especially to the QM part close to the QM/MM interface but also between closely encountering MM molecules. The (*ad hoc*) introduction of distributed point charges and polarizabilities by Gaussian broadenings remedies the problem to some extent.[16] These localized properties have indeed been of much use as granulated force fields in QM/MM calculations of properties, ranging from NMR, EPR over to optical properties and X-ray spectroscopy.[1,5,21–24]

In a previous work[25] we showed how a LoProp polarizable force field can be derived quantum mechanically for extended environments by using analytical response theory. The theory was recently extended to construct the localized polarizable force fields to be frequency dependent.[26] The possibility to do so follows from the fact that analytic response theory can determine the first-order response for frequencies which directly match those of the externally perturbing field. This effectively produces polarizable force fields which include the quantum mechanical density perturbation caused by an external field. This extension of the MM region in typical QM/MM calculations allows for a more physical description of properties in the QM region, since the interaction between the MM and the QM regions now includes the frequency dependence of the classical region. Further development in terms of non-linear hyperpolarizable force fields followed, which, *e.g.* can be used to reconstruct hyperpolarizable tensors, and properties derived therefrom, in large molecular clusters and proteins.[27] Further recent development has concerned the decomposition of obtain complex frequency dependent polarizabilities and dispersion coefficients. The centring of such polarizabilities to bonds is also possible, and, more importantly, to higher order dispersion coefficients, higher than C_6, is doable if the underlying property integrals are at hand. It could allow the development of chemically dependent local dispersion interaction force fields to be introduced in molecular dynamics. In addition to the use in QM/MM, the LoProp technique has recently been shown to be powerful

2252 | *Phys. Chem. Chem. Phys.*, 2019, **21**, 2251–2270

This journal is © the Owner Societies 2019

tool for building up "classical" properties of large molecular clusters, like Rayleigh scattering of atmospheric nanoparticles,[28] and hyperpolarizabilities (second harmonic generation) of proteins.[29] In the present work review recent development of force field decomposition and local property calculations with quantum chemical methods and discuss some topical applications both in terms of QM/MM and for building up properties of large molecules and clusters.

2 Theory

2.1 The LoProp transformation

Gagliardi *et al.*[20] introduced the LoProp method for extracting atomic and interatomic contributions to molecular properties. It was formulated as a sequence of transformations of the atomic overlap matrix, giving a localized orthonormal basis that depends only on molecular structure and the initial atomic orbital basis set. The localized properties were obtained by forming partial traces over basis functions associated with one chosen center or a pair of centers. The aim was to obtain physically sound and transferable local properties which sum up to the molecular property and which are reasonably constant for a given atom in chemically similar systems. The implications of this is that these localized properties can be used as force fields in QM/MM calculations.

The requirement on the initial basis set is that the atomic orbitals can be grouped into occupied and virtual orbitals, *e.g.* if they resemble the actual orbitals in atomic states. Given this division, the sequence of transformations can be briefly summarized as: for the combined atomic-orbital basis set of the molecule, orthonormalize first within atomic subblocks, second within the combined occupied AO space. Then project out any occupied components in the virtual AO space, and finally orthonormalize the virtual space. This leads to an orthonormal localized basis set, the LoProp basis, where each orbital is associated with an atomic center.

These steps generate a total matrix T_l^μ describing a transformation between the AO:s (ξ_μ) and the LoProp (ϕ_l) basis functions

$$\phi_l = \sum_\mu \xi_\mu T_l^\mu \qquad (1)$$

For integral representations of one-electron operators

$$O_{\mu\nu} = \int dV \xi_\mu(\vec{r}) O(\vec{r}) \xi_\nu(\vec{r}) \quad O_{lm} = \int dV \chi_l(\vec{r}) O(\vec{r}) \chi_m(\vec{r}) \quad (2)$$

we thus have the transformation rule

$$O_{lm} = \sum_{\mu\nu} O_{\mu\nu} T_l^\mu T_m^\nu \qquad (3)$$

while for the density matrix we have the inverse transformation

$$D^{\mu\nu} = \sum_{lm} D^{lm} T_l^\mu T_m^\nu \qquad (4)$$

Next the atom and "bond" contributions are given by grouping the summations involved in the trace operations into partial traces over the atomic subsets, denoted by $A,B\ldots$

$$\text{tr}(OD) = \sum_{lm} O_{lm} D^{lm} = \sum_{AB} \sum_{l\in A, m\in B} O_{lm} D^{lm} \equiv \sum_{AB} O_{AB} D^{AB} \quad (5)$$

The localized charges and isotropic polarizabilities are the parameters which form the force fields for the QM/MM calculations in this work.

The decomposition of the total electronic charge according to eqn (5) gives

$$Q_{el} = -\sum_{AB} \sum_{l\in Am, m\in B} S_{lm} D^{lm} \qquad (6)$$

where S is the overlap matrix. As the LoProp basis is orthonormal ($S_{lm} = \delta_{lm}$) as well as localized this becomes

$$Q_{el} = -\sum_A \sum_{l\in A} D^{ll} \qquad (7)$$

i.e. there are no bond contributions to the total charge and the localized charge associated with an atomic site A with nuclear charge Z_A, is

$$Q_A = Z_A - \sum_{l\in A} D^{ll} \qquad (8)$$

2.2 Analytical local polarizabilities

We present here an alternative formulation and implementation of the LoProp approach. The original implementation was based on finite field perturbation theory applied to localized dipole moments. Here we apply analytical response theory where the change of a wave function in an external field is represented by a unitary transformation. The theory derives from the classical work by Olsen and Jørgensen[30] for MCSCF theory and later generalized for DFT by Sałek *et al.*[31] Following the notation of the latter, we have that a wave function $|0\rangle$ is transformed by a unitary operator as a result of the perturbation

$$|\tilde{0}\rangle \equiv e^{-\hat{\kappa}}|0\rangle \qquad (9)$$

The operator κ has a real anti-symmetric matrix representation which in second quantization is given by

$$\hat{\kappa} = \kappa_{pq} a_p^\dagger a_q \qquad (10)$$

If we restrict to static perturbations, the Ehrenfest theorem gives that the first-order properties can be derived from

$$\delta\langle 0|[\hat{Q}, e^{\kappa} H e^{-\kappa}]|0\rangle = 0 \qquad (11)$$

for an arbitrary operator $\hat{Q}$. By choosing the same excitation operators that are used to expand $\hat{\kappa}$ we obtain a system of equations for the unknowns, the matrix elements κ_{pq}. As the details of these derivations are quite lengthy we refer the interested reader to ref. 31.

After solving the linear response eqn (11), some matrix algebra allows us to express the first-order change of an expectation value in terms of a trace of the product of the operator matrix and the first-order density matrix

$$\delta\langle e^{\kappa} A e^{-\kappa}\rangle = \langle[\delta\kappa, A]\rangle = \sum_{pq} A_{pq} \delta D^{pq} \qquad (12)$$

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2253**

where the first-order change in the density matrix can be expressed as

$$\delta D^{pq} = [\delta\kappa^T, D]^{pq} \tag{13}$$

all in MO basis. It is now straightforward to transform the same quantities to AO (with MO-coefficients C_p^μ) or LoProp basis (with LoProp transformation T_l^μ)

$$\delta D^{\mu\nu} = \sum_{pq} \delta D^{pq} C_p^\mu C_q^\nu = \sum_{lm} \delta D^{lm} T_l^\mu T_l^\nu \tag{14}$$

With the first-order densities at hand we are now in a position to proceed to evaluate the localized polarizabilities in the LoProp approach. These were introduced in ref. 20 by considering the electric field derivatives of the molecular dipole moment with respect to an arbitrary gauge origin C

$$\langle -\vec{r}_C \rangle = -\sum_{AB} \vec{r}_{AB} D^{AB} + \sum_A Q_A \left(\vec{R}_A - \vec{R}_C \right) \tag{15}$$

$\vec{r}_{AB}$ is the electronic coordinate with respect to the midpoint between atoms A and B, if $A \neq B$, or with respect to an atomic site $\vec{R}_A$ if $A = B$. Considering at first a first-order variation of the dipole moment

$$\delta \langle -\vec{r}_C \rangle = \sum_{AB} -\vec{r}_{AB} \delta D^{AB} + \sum_A \delta Q_A \vec{R}_A, \tag{16}$$

it is seen that the last term in (15) containing C vanishes due to charge conservation. The fact that the individual terms in (16) are origin dependent were solved by Gagliardi by introducing charge transfer between atomic sites – an anti-symmetric charge transfer matrix ΔQ_{AB} satisfying

$$\delta Q_A = \sum_B \Delta Q_{AB} \tag{17}$$

where the charge shift on a site A is

$$\delta Q_A = -\sum_{l \in A} \delta D^{ll} \tag{18}$$

This leads to a localized polarizability of the form

$$\alpha_{AB} = \delta \langle -\vec{r} \rangle_{AB} = -\vec{r}_{AB} \delta D^{AB} + \Delta Q_{AB}(\vec{R}_A - \vec{R}_B) \tag{19}$$

The charge transfer matrix is not uniquely defined but it solves the problem with gauge-dependence and one gets a physically reasonable interpretation of the local polarizability – as a change of the local dipole moment and, in the case of bond contributions, a charge shift between the two sites A and B. In our calculations we redistribute the bond contributions to their atomic sites and only consider local atomic polarizabilities.

To summarize, the main difference between this approach and the original implementation is that it is based on calculations on analytical response theory rather than finite-field perturbation theory. An advantage of the formulation is that it is straightforward to generalize the LoProp concept to local dynamical polarizabilities and to higher order polarizabilities. The analytical LoProp code is a post-processing tool for Dalton[32] written in the Python programming language.

2.3 A frequency-dependent LoProp approach

In a previous paper[25] we demonstrated an alternative derivation of the LoProp approach which was based on analytical response theory. As response theory provides a formulation of perturbation theory which handles time-independent and time-dependent cases within the same formalism, we then investigated the effects of projecting the dynamical polarizability onto atomic contributions. This can be expected to be meaningful provided that the frequencies involved are well below the first resonance. To outline the steps in single-determinant time-dependent response theory (see e.g. Salek et al.[33]) we have a state $|\tilde{0}\rangle$ that evolves in time

$$|\tilde{0}\rangle \equiv e^{-\hat{\kappa}(t)}|0\rangle \tag{20}$$

where the exponential operator is a parameterized time-evolution operator, $\hat{\kappa}$ a real anti-hermitian operator

$$\hat{\kappa} = \kappa_{pq}(t)a_p^\dagger a_q \tag{21}$$

of which the matrix elements $\{\kappa_{pq}\}$ form the parameters of the theory. A time-dependent variational principle based on the Ehrenfest theorem provides a solution to the time-independent Schrödinger equation in this space of parameters $\{\kappa_{pq}\}$, i.e. arbitrary static operators $\hat{O}$ that satisfy

$$\delta \left\langle 0 \left| \left[\hat{O}, e^{\hat{\kappa}(t)} \left(H - i\frac{\partial}{\partial t} \right) e^{-\hat{\kappa}(t)} \right] \right| 0 \right\rangle = 0 \tag{22}$$

A well-chosen set of operators $\{\hat{O}\}$ provides linear systems of equations for the parameters $\{\kappa_{pq}\}$ to various orders in the perturbation. If we consider monochromatic perturbations of frequency ω we can extract the dynamical dipole polarizability from the linear response function

$$\delta \langle e^{\hat{\kappa}} \vec{r} e^{-\hat{\kappa}} \rangle(\omega) = \langle [\delta\kappa, \vec{r}] \rangle(\omega) = \sum_{pq} \vec{r}_{pq} \delta D^{pq}(\omega) \tag{23}$$

$$\delta D^{pq}(\omega) = [\delta\kappa^T(\omega), D]^{pq} \tag{24}$$

The dipole moment is expanded in terms of atomic- and bond contributions

$$\langle -\vec{r}_C \rangle = -\sum_{AB} \sum_{\substack{l \in A \\ m \in B}} (\vec{r}_{AB})_{lm} D^{lm} + \sum_A Q_A \left(\vec{R}_A - \vec{R}_C \right) \tag{25}$$

with the localized LoProp basis, where $\vec{r}_{AB}$ is the electronic coordinate with respect to the midpoint of the bond A–B and the frequency-dependent variation can be written

$$\delta \langle -\vec{r}_C \rangle(\omega) = \sum_{AB} \sum_{\substack{l \in A \\ m \in B}} -(\vec{r}_{AB})_{lm} \delta D^{lm}(\omega) + \sum_{AB} \Delta Q_{AB}(\omega) \left(\vec{R}_A - \vec{R}_B \right), \tag{26}$$

where ΔQ_{AB} is a charge transfer matrix satisfying

$$\sum_B \Delta Q_{AB}(\omega) = \delta Q_A(\omega) \tag{27}$$

and where the local charge response to the external perturbation is

$$\delta Q_A(\omega) = -\sum_{l \in A} \delta D^{ll}(\omega) \tag{28}$$

2254 | Phys. Chem. Chem. Phys., 2019, 21, 2251–2270

This journal is © the Owner Societies 2019

This leads to a localized dynamical polarizability of the form

$$\alpha_{AB}(\omega) = \delta\langle -\vec{r}\rangle_{AB} = -\vec{r}_{AB}\delta D^{AB}(\omega) + \Delta Q_{AB}(\omega)(\vec{R}_A - \vec{R}_B) \tag{29}$$

completely analogous to the dynamic case.

Most applications of classical models for the calculation of polarizabilities and hyperpolarizabilities are semi-empirical in nature, *e.g.* in the work of Applequist *et al.* the atomic polarizabilities are parameters which are optimized for a test set of molecules. The approach reviewed above, on the other hand, is fully *ab initio* in nature in that quantum mechanical values for a monomer are used in the classical models for a cluster of molecules. For clusters of moderate size the validity of the model can be tested with a full quantum mechanical calculation for smaller oligomers. It can be expected that these models hold for well separated systems, while breaking down at close distances where the exchange interaction between the overlapping electron densities is non-negligible. After highlighting the basic point dipole models, we derive the equations necessary for the extraction of atomic hyperpolarizabilities.

2.4 Point dipole models

2.4.1 Polarizability: the Silberstein–Applequist model reviewed.
The point-dipole model associates with every particle i a dipole moment $\mathbf{p}_i$, that consists of a static (permanent) part $\mathbf{p}_i^0$ and an induced part which for weak fields depends linearly on the local electric field.

$$\mathbf{p}_i = \mathbf{p}_i^0 + \alpha_i \cdot \mathbf{E}_i \tag{30}$$

$\mathbf{E}_i = \mathbf{E}(\vec{r}_i)$ is the electric field experienced by particle i located at $\vec{r}_i$ due to external sources and other dipoles. Let the laboratory field be $\mathbf{F}$. Typically it will be uniform over the system, but as a notation we may write $\mathbf{F}_i = \mathbf{F}(\vec{r}_i)$ to emphasize what particle we are referring to. Then

$$\mathbf{E}_i = \mathbf{F}_i + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \mathbf{p}_j \tag{31}$$

where $\mathbf{T}_{ij}$ is the dipole dyadic tensor coupling particles i and j ($i \neq j$),

$$\mathbf{T}_{ij} = \frac{3\vec{r}_{ij}\vec{r}_{ij} - r_{ij}^2\mathbf{1}}{r_{ij}^5} \tag{32}$$

The components of $\vec{r}_{ij}$ in eqn (32) is for the vector pointing from the dipole at j to the dipole at i. The change of an individual dipole due to variations in the external field is

$$\delta\mathbf{p}_i = \alpha_i \cdot \left(\delta\mathbf{F}_i + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \delta\mathbf{p}_j\right) \tag{33}$$

and gives a system of equations for the dipole shifts:

$$\delta\mathbf{p}_i - \alpha_i \cdot \sum_{j \neq i} \mathbf{T}_{ij} \cdot \delta\mathbf{p}_j = \alpha_i \cdot \delta\mathbf{F}_i \tag{34}$$

Defining $\mathbf{T}_{kk} = 0$ we may write

$$\sum_j \left(\delta_{ij}\mathbf{1} - \alpha_i \cdot \mathbf{T}_{ij}\right) \cdot \delta\mathbf{p}_j = \alpha_i \cdot \delta\mathbf{F}_i \tag{35}$$

a 3N-system of equations for the induced dipoles as a function of the change in the external field. Inverting the coefficient matrix of eqn (35) gives the individual dipole shifts

$$\delta\mathbf{p}_i = \sum_j \mathbf{R}_{ij} \cdot \delta\mathbf{F} \tag{36}$$

where

$$\mathbf{R}_{ij} = \left(\delta_{ij}\mathbf{1} - \alpha_i\mathbf{T}_{ij}\right)^{-1} \cdot \alpha_j \tag{37}$$

is a classical response function, often called relay matrix, as it relays a field change at site j to a dipole shift at site i.[34]

$$\mathbf{R}_{ij} = \frac{\delta\mathbf{p}_i}{\delta\mathbf{F}_j} \tag{38}$$

The total molecular dipole shift is obtained by summing up the individual atomic dipole shifts, which for a uniform external field gives

$$\delta\mathbf{P} = \sum_i \delta\mathbf{p}_i = \sum_{ij} \mathbf{R}_{ij} \cdot \delta\mathbf{F}_i = \left(\sum_{ij} \mathbf{R}_{ij}\right) \cdot \delta\mathbf{F} \tag{39}$$

i.e. the total molecular polarizability of the system is

$$\alpha_m = \sum_{ij} \mathbf{R}_{ij} \tag{40}$$

It can be noted that for separated charges the elements of $\mathbf{T}$ are small and the total polarizability is approximately the sum of the individual polarizabilities. This is the Silberstein model[35,36] generalized firstly by Applequist[34,37,38] and later used by several others.[39–48]

2.4.2 Extension: polarizability of a system of hyperpolarizable particles.
Now we extend the Silberstein–Applequist model to a system of particles with additional structure, with local polarizabilities as well as hyperpolarizabilities. With a local hyperpolarizability tensor β_i the local dipole moments are

$$\mathbf{p}_i = \mathbf{p}_i^0 + \alpha_i \cdot \mathbf{E}_i + \frac{1}{2}\beta_i : \mathbf{E}_i\mathbf{E}_i \tag{41}$$

and consequently the induced components are given by

$$\delta\mathbf{p}_i = \alpha_i \cdot \delta\mathbf{E}_i + \beta_i : \mathbf{E}_i\delta\mathbf{E}_i \tag{42}$$

The first question displays how this affects the definition of the total polarizability α

$$\delta\mathbf{p}_i = \alpha_i \cdot \left(\delta\mathbf{F}_i + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \delta\mathbf{p}_j\right)$$
$$+ \beta_i : \mathbf{E}_i\left(\delta\mathbf{F}_i + \sum_{j \neq i} \mathbf{T}_{ij} \cdot \delta\mathbf{p}_j\right) \tag{43}$$

As before, collecting induced dipoles to the left we obtain

$$\sum_{j \neq i} \left(\delta_{ij}\mathbf{1} - (\alpha_i + \beta_i \cdot \mathbf{E}_i) \cdot \mathbf{T}_{ij}\right) \cdot \delta\mathbf{p}_j = (\alpha_i + \beta_i \cdot \mathbf{E}_i) \cdot \delta\mathbf{F}_i \tag{44}$$

We now note that the structure of this equation is the same as eqn (35) with a corrected local polarizability

$$\tilde{\alpha}_i = \alpha_i + \beta_i \cdot \mathbf{E}_i \tag{45}$$

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2255**

This equation is though different from eqn (35) in the sense that it refers to the explicit value of the local electric field at each site. As we are considering the effect of an external electric field on the total system we assume these local fields to be the stationary value in the absence of an external field, *i.e.*

$$\mathbf{E}_i^0 = \sum_j \mathbf{T}_{ij} \cdot \mathbf{p}_i[\mathbf{F} = 0] \tag{46}$$

Nevertheless, it is now not sufficient to invert a matrix – we also have to solve for the local fields iteratively in order to set up the linear systems of equations for the dipole shifts. This gives an Applequist equation with corrected local polarizabilities

$$\left(\delta_{ij}\mathbf{1} - \sum_{j \neq i} \tilde{\alpha}_i \cdot \mathbf{T}_{ij}\right) \cdot \delta\mathbf{p}_j = \tilde{\alpha}_i \cdot \delta\mathbf{F} \tag{47}$$

or

$$\delta\mathbf{p}_i = \sum_j \tilde{\mathbf{R}}_{ij} \cdot \delta\mathbf{F} \tag{48}$$

where

$$\tilde{\mathbf{R}}_{ij} = (\mathbf{1}\delta_{ij} - \tilde{\alpha}_i \cdot \mathbf{T}_{ij})^{-1} \cdot \tilde{\alpha}_j \tag{49}$$

2.5 Hyperpolarizability of a system of hyperpolarizable particles

The next question displays how an equivalent model can be defined for at total hyperpolarizability. Taking the same argument from the beginning and considering the change in the individual dipole moments we obtain

$$\mathbf{p}_i = \mathbf{p}_i^0 + \alpha_i \cdot \mathbf{E}_i + \frac{1}{2}\beta_i : \mathbf{E}_i\mathbf{E}_i$$

$$\delta\mathbf{p}_i = \alpha_i \cdot \delta\mathbf{E}_i + \beta_i : \mathbf{E}_i\delta\mathbf{E}_i$$

$$\delta^2\mathbf{p}_i = \alpha_i \cdot \delta^2\mathbf{E}_i + \beta_i : \delta\mathbf{E}_i\delta\mathbf{E}_i + \beta_i : \mathbf{E}_i\delta^2\mathbf{E}_i \tag{50}$$

$$= (\alpha_i + \beta \cdot \mathbf{E}_i) \cdot \delta^2\mathbf{E}_i + \beta_i : \delta\mathbf{E}_i\delta\mathbf{E}_i$$

$$= \tilde{\alpha}_i \cdot \delta^2\mathbf{E}_i + \beta_i : \delta\mathbf{E}_i\delta\mathbf{E}_i$$

Now we are only looking for terms quadratic in the first-order field so we can set $\delta^2\mathbf{F} = 0$. The second-order effect in the local field is

$$\delta^2\mathbf{E}_i = \sum_{j \neq i} \mathbf{T}_{ij} \cdot \delta^2\mathbf{p}_j \tag{51}$$

giving

$$\delta^2\mathbf{p}_i = \tilde{\alpha}_i \cdot \sum_{i \neq j} \mathbf{T}_{ij}\delta^2\mathbf{p}_j + \beta_i : \delta\mathbf{E}_i\delta\mathbf{E}_i \tag{52}$$

The equation for the second-order dipole moments are thus

$$(\delta_{ij}\mathbf{1} - \tilde{\alpha}_i\mathbf{T}_{ij}) \cdot \delta^2\mathbf{p}_j = \beta_i : \delta\mathbf{E}_i\delta\mathbf{E}_i \tag{53}$$

From the linear problem we have

$$\delta\mathbf{E}_i = \delta\mathbf{F}_i + \sum_j \mathbf{T}_{ij} \cdot \delta\mathbf{p}_j$$

$$= \delta\mathbf{F}_i + \sum_{jk} \mathbf{T}_{ij} \cdot \tilde{\mathbf{R}}_{jk} \cdot \delta\mathbf{F}_k \tag{54}$$

$$= \sum_k \left(\delta_{ik}\mathbf{1} + \sum_j \mathbf{T}_{ij} \cdot \tilde{\mathbf{R}}_{jk}\right) \cdot \delta\mathbf{F}_k$$

which in the case of a uniform external field simplifies to

$$\delta\mathbf{E}_i = \left(\mathbf{1} + \sum_{jk} \mathbf{T}_{ij} \cdot \tilde{\mathbf{R}}_{jk}\right) \cdot \delta\mathbf{F} \tag{55}$$

Analogously to the Applequist relations, the molecular hyperpolarizability is now the sum of local second-order dipole shifts

$$\beta_m = \sum_i \frac{\delta^2\mathbf{p}_i}{\delta\mathbf{F}^2} = \sum_{ij} \tilde{\mathbf{R}}_{ij} \cdot \beta_j : \frac{\delta\mathbf{E}_j}{\delta\mathbf{F}}\frac{\delta\mathbf{E}_j}{\delta\mathbf{F}}$$

$$= \sum_{ij} \tilde{\mathbf{R}}_{ij} \cdot \beta_j : \left(\mathbf{1} + \sum_{kl} \mathbf{T}_{jk} \cdot \tilde{\mathbf{R}}_{kl}\right)\left(\mathbf{1} + \sum_{kl} \mathbf{T}_{jk} \cdot \tilde{\mathbf{R}}_{kl}\right) \tag{56}$$

2.6 LoProp hyperpolarizability

In a previous section we showed how the localized dynamic polarizabilities of atoms can be extracted by analytical response theory at static or dynamic frequencies for a molecule.[26] We now extend the formulation to include the contribution of a higher-order perturbation to the electronic density, to obtain localized hyperpolarizabilities.

As for the LoProp frequency dependent polarizability, see Section 2.3, we start out from response theory for single-determinant wave-functions,[31] where the time-dependence of an electronic state can be written as the time-evolution operator acting on a stationary state

$$|\tilde{0}\rangle \equiv e^{-\kappa(t)}|0\rangle \tag{57}$$

where the exponential operator is a time-evolution operator and the operator $\kappa(t)$ is the anti-hermitian operator with matrix elements

$$\kappa(t) = \kappa_{pq}(t)a_{p\sigma}^{\dagger}a_{q\sigma} \tag{58}$$

$$\kappa_{qp} = -\kappa_{pq}^{\star} \tag{59}$$

The elements of $\kappa(t)$ forms the parameters of the theory. Furthermore, a time-dependent variational theory based on the Ehrenfest theorem can be expanded to give a solution to the time-independent Schrödinger equation in terms of the parameters $\kappa(t)_{pq}$

$$\delta\left\langle 0 \left| \left[\hat{O}, e^{\kappa(t)}\left(\hat{H}_0 + V(t) - i\frac{\partial}{\partial t}\right)e^{-\kappa(t)}\right] \right| 0\right\rangle = 0 \tag{60}$$

where the operator $\hat{O}$ can be any time-independent property operator. Choosing $\hat{O}$ as the non-redundant orbital-excitations gives a linear system of equations that determines the parameters κ_{rs} uniquely.[30] Choosing $\hat{O}$ as well as the perturbation V

to be dipole moment operators $\hat{\vec{r}}$, (the latter corresponding to the dipole interaction between the molecule and an external electric field with frequency ω) we can extract the polarizability as the linear response function

$$\delta\left\langle e^{\kappa}\hat{\vec{r}}e^{-\kappa}\right\rangle(\omega) = \left\langle\left[\delta\hat{\kappa},\hat{\vec{r}}\right]\right\rangle(\omega) = \sum_{pq}\hat{\vec{r}}_{pq}\delta D^{pq}(\omega) \tag{61}$$

$$\delta D^{pq}(\omega) = \left[\delta\kappa^T(\omega),\delta D\right]^{pq} \tag{62}$$

and the hyperpolarizability as the quadratic response function

$$\delta^2\left\langle e^{\kappa}\hat{\vec{r}}e^{-\kappa}\right\rangle(\omega_1,\omega_2) = \left\langle\left[\delta^2\hat{\kappa},\hat{\vec{r}}\right] + \frac{1}{2}\left[\delta\hat{\kappa},\left[\delta\hat{\kappa},\hat{\vec{r}}\right]\right]\right\rangle(\omega_1,\omega_2)$$

$$= \sum_{pq}\hat{\vec{r}}_{pq}\delta^2 D^{pq}(\omega_1,\omega_2) \tag{63}$$

where

$$\delta^2 D^{pq}(\omega_1,\omega_2) = \left[\delta^2\kappa^T(\omega_1,\omega_2),D\right]^{pq}$$
$$+ \frac{1}{2}\big(\left[\delta\kappa^T(\omega_1),\left[\delta\kappa^T(\omega_2),D\right]\right] \tag{64}$$
$$+ \left[\delta\kappa^T(\omega_2),\left[\delta\kappa^T(\omega_1),D\right]\right]\big)^{pq}$$

To obtain localized properties, the LoProp transformation is then applied to yield a localized orthonormal basis, where the molecular dipole moment can be written as the sum of atom and bond contributions

$$\delta\langle-\vec{r}_C\rangle(\omega) = \sum_{AB}\sum_{\substack{l\in A\\m\in B}}-(\vec{r}_{AB})_{lm}\delta D^{lm}(\omega)$$
$$+ \sum_{AB}\Delta Q_{AB}(\omega)\left(\vec{R}_A - \vec{R}_B\right)$$

$$\delta^2\langle-\vec{r}_C\rangle(\omega) = \sum_{AB}\sum_{\substack{l\in A\\m\in B}}-(\vec{r}_{AB})_{lm}\delta^2 D^{lm}(\omega_1,\omega_2)$$
$$+ \sum_{AB}\Delta^2 Q_{AB}(\omega_1,\omega_2)\left(\vec{R}_A - \vec{R}_B\right)$$

where $\vec{r}_{AB}$ is the electronic coordinate with respect to the bond-center between the atoms A and B. The first two terms are the contributions from the polarizability, and the last two terms are from the hyperpolarizability. $\Delta Q_{AB}(\omega)$ and $\Delta^2 Q_{AB}(\omega)$ are the atomic charge-transfer matrices, both individually satisfying

$$\sum_B \Delta Q_{AB}(\omega) = \delta Q_A(\omega) \tag{65}$$

$$\sum_B \Delta^2 Q_{AB}(\omega_1,\omega_2) = \delta^2 Q_A(\omega_1,\omega_2) \tag{66}$$

where

$$\delta Q_A(\omega) = -\sum_{l\in A}\delta D^{ll}(\omega) \tag{67}$$

$$\delta^2 Q_A(\omega_1,\omega_2) = -\sum_{l\in A}\delta^2 D^{ll}(\omega_1,\omega_2) \tag{68}$$

are the first, and second-order local response with respect to the external field. Having the second-order electronic density perturbation, the final form of the localized hyperpolarizabilities reads

$$\beta_{AB}(\omega_1,\omega_2) = -\sum_{\substack{l\in A\\m\in B}}\vec{r}_{lm}\delta^2 D^{lm}(\omega_1,\omega_2)$$
$$+ \Delta^2 Q_{AB}(\omega_1,\omega_2)\left(\vec{R}_A - \vec{R}_B\right) \tag{69}$$

We emphasize that in the original work static hyperpolarizabilities ($\omega_1 = \omega_2 = 0$) were considered. These relations have been implemented as a plugin for Dalton[32] available in the public domain.[49]

2.7 Dispersion interaction

We here summarize the relations between the linear response functions and dispersion energies. The detailed derivations are based on the work by Magnasco *et al.*[50] and are available in ref. 51.

Given the density operator in second quantization, the density–density response function of a molecular system can be written as

$$\left\langle\left\langle\hat{\rho}(\mathbf{r}_1);\hat{\rho}\left(\mathbf{r}_1'\right)\right\rangle\right\rangle_\omega^{(1)} = \sum_{n_1>0}\frac{\left\langle\Psi_0^{(1)}|\hat{\rho}(\mathbf{r}_1)|\Psi_{n_1}^{(1)}\right\rangle\left\langle\Psi_{n_1}^{(1)}|\hat{\rho}(\mathbf{r}_1')|\Psi_0^{(1)}\right\rangle}{\hbar\left(\omega-\omega_{n_1}^{(1)}\right)}$$
$$- \frac{\left\langle\Psi_0^{(1)}|\hat{\rho}(\mathbf{r}_1')|\Psi_{n_1}^{(1)}\right\rangle\left\langle\Psi_{n_1}^{(1)}|\hat{\rho}(\mathbf{r}_1)|\Psi_0^{(1)}\right\rangle}{\hbar\left(\omega+\omega_{n_1}^{(1)}\right)} \tag{70}$$

It is well established that the dispersion energy can be expressed as an integral over polarizabilities over imaginary frequencies. This gives

$$E_{\text{disp}} = -\frac{\hbar}{2\pi}\iint d\mathbf{r}_1 d\mathbf{r}_2\frac{1}{r_{12}}\iint d\mathbf{r}_1' d\mathbf{r}_2'\frac{1}{r_{12'}}$$
$$\times \int_0^\infty d\omega\left\langle\left\langle\hat{\rho}(\mathbf{r}_1);\hat{\rho}\left(\mathbf{r}_1'\right)\right\rangle\right\rangle_{i\omega}^{(1)}\left\langle\left\langle\hat{\rho}(\mathbf{r}_2);\hat{\rho}\left(\mathbf{r}_2'\right)\right\rangle\right\rangle_{i\omega}^{(1)} \tag{71}$$

A Taylor expansion with respect to intermolecular distance yield to lowest order an interaction that decays as R^{-6}

$$E_{\text{disp}} \approx -\frac{\hbar}{2\pi}\int_0^\infty d\omega\sum_{ijkl}\alpha_{ij}^{(1)}(i\omega)\alpha_{kl}^{(2)}(i\omega)T_{ik}T_{jl} \tag{72}$$

where the dipolar coupling tensor is

$$T_{ik} = \frac{\partial^2}{\partial X_{12}^i\partial X_{12}^k}\frac{1}{R_{12}} = \frac{3X_{12}^i X_{12}^k - \delta_{ik}R_{12}^2}{R_{12}^5} \tag{73}$$

The isotropic expression is obtained by a rotational averaging procedure[52]

$$E_{\text{disp}}^{\text{iso}} \approx -\frac{\hbar}{2\pi}\int_0^\infty d\omega\sum_{ijkl}\sum_{\mu\nu\sigma\tau}\alpha_{\mu\nu}^{(1)}(i\omega)\alpha_{\sigma\tau}^{(2)}(i\omega)\overline{C_{\mu i}C_{\nu j}}$$
$$\times \overline{C_{\sigma k}C_{\tau l}}T_{ik}T_{jl} \tag{74}$$

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2257**

which involves orientation averaging of coupling coefficients between body-fix and space-fix coordinate axes:

$$\overline{C_{\mu i} C_{\nu j}} = \frac{1}{3} \delta_{\mu\nu} \delta_{ij} \qquad (75)$$

Substituting with the isotropic polarizabilities

$$\bar{\alpha} = \frac{1}{3} \sum_k \alpha_{kk} \qquad (76)$$

we obtain the final expression

$$C_6^{ij} = \frac{3\hbar}{\pi} \int_0^\infty \bar{\alpha}_i^{(1)}(i\omega) \bar{\alpha}_j^{(2)}(i\omega) d\omega \qquad (77)$$

The LoProp approach[20] to obtain frequency dependent polarizabilities is based on diagonalizing the atomic overlap matrix S. This is done as described in Section 2.1 in sequential orthonormalization steps which gives a transformation matrix used to transform the property integrals and perturbed electronic densities. By summing the transformed integrals over the subspace of atom-pairs, the properties in the LoProp basis are obtained. It was shown above how the frequency-dependent polarizability and hyperpolarizability could be obtained from linear response theory.[26,27] Here, for the dispersion interaction, the natural extension to the extraction of the response vectors from the CPP code is made in order to obtain the distributed real polarizabilities of atoms at imaginary frequencies.

2.8 Damping

Due to very large molecular polarizabilities caused by the induced dipoles at short inter-atomic distances, the original Applequist[34,37] equations were modified by Thole.[53] The main difference in the Tholes model derives from the introduction of the so-called damping parameters, which depend on the scaled inter-atomic distances. The equation used to calculate the damping reads

$$u = \frac{\vec{r}_{ij}}{\left(\alpha_i^{iso} \alpha_j^{iso}\right)^{1/6}} \qquad (78)$$

which gives the following damping parameters for the potential, field, and field-gradient, respectively

$$f_V = 1 - \left(\frac{1}{2}v + 1\right) e^{-v} \qquad (79)$$

$$f_E = f_V - \left(\frac{1}{2}v^2 + \frac{1}{2}v\right) e^{-v} \qquad (80)$$

$$f_T = f_E - \frac{1}{6}v^3 e^{-v} \qquad (81)$$

where $v = au$, and a is an arbitrary damping parameter. The parameter derived most accurately using the largest set of data[54] is $a = 2.1304$, and is the one also used in this review.

In the damped model, the dyadic tensor now reads

$$\mathbf{T}_{ij} = \frac{3\vec{r}_{ij}\vec{r}_{ij}f_T - f_E r_{ij}^2 \mathbf{1}}{r_{ij}^5} \qquad (82)$$
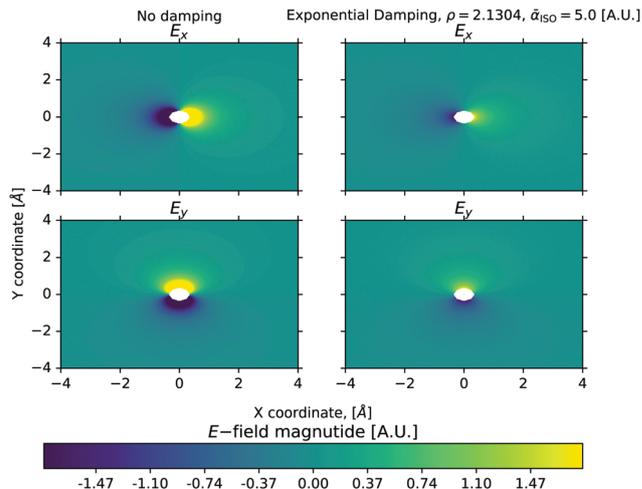

Fig. 1 The E_x and E_y components due to an ideal point charge at origo with $q = 1$ a.u. The right column illustrates how Tholes exponential damping gradually decreases the electric field near the charge source. From ref. 29. Copyright ACS Publishing.

As an illustration of the damped model, Fig. 1 shows the E_x and E_y electric-field components in the xy-plane stemming from a position in origo, with (left column) and without (right column) damping.

3 Decomposition of frequency dependent polarizabilities

The polarizability is an example of a property for which both classical and quantum models have played important roles for its calculation. In the original works by Applequist, Carl, and Fung[37] and Silberstein[35,36] the concept of the polarizable point-dipole model was introduced in which the molecular polarizability is obtained as a result of atomic interactions governed by the localized polarizability tensors of each point-dipole (see Theory section for mathematical details). In that formulation, the induced dipoles are determined solely by the atomic polarizabilities which in turn can be directly calculated by solving a system of linear equations. The atomic polarizabilities were then often fitted to yield molecular polarizabilities as closely as possible to molecular polarizabilities obtained from experimental refractive index measurements.[55] As shown already in the original works, short interatomic distances could give rise to polarization catastrophes since the induced dipoles then would increase indefinitely. Several models have subsequently been put forward to fix the problems arising from short interatomic distances, for example, Thole[53] tested two modified dipole-interaction models, based on charge distributions, also called the linear- and exponential models, and re-parameterized the atomic polarizabilities to fit the new interaction equations: Ren and Ponder included higher-order multipole induction terms, in order to study the intra- and intermolecular interactions[56] of molecules, with successful extensions to the polarizable atomic multipole water model.[57] Wang et al.[58,59] further extended the

point-dipole model with new sets of experimental data available to molecules.[60] Jensen *et al.*,[39] and others, used quantum mechanical calculated molecular polarizabilities to fit atomic polarizabilities. Here a quantum approach was used to evaluate frequency dependent molecular polarizabilities of amino acids and proteins,[40] and also later the second hyperpolarizability.[41]

Ref. 26 presented a way of incorporating quantum mechanical granulation of a molecular mechanics environment through the localization of frequency dependent polarizabilities. The work was motivated by that such frequency dependent molecular mechanics force fields allow for an instantaneous, optical, polarization in the molecular mechanics part by the electronic degrees of freedom in the quantum part of a quantum mechanics/molecular mechanics (QM/MM) model and make it possible to extend the QM/MM properties to cover the full QM/MM simulation volume. This is a generalization of the QM/MM technology where the embedding MM part acts as a perturber of the property determined by the QM. Furthermore, these force fields allow for calculations of frequency dependent properties for very large clusters that are evaluated classically. The construction of frequency dependent localized polarizable force fields within analytic response theory, makes it possible, in contrast to the finite field algorithm, to determine the first order response for frequencies which directly match those of the externally perturbing field. That includes the implementation of the modified Silberstein–Applequist mode, alluded to above, for many interacting induced dipoles for the linear frequency dependent polarizability. The performance and precision of the method was evaluated in ref. 26 through studying a few selected cases. The purely classical intermolecular polarizability using *ab initio* derived properties was also studied, and, specifically, the effect caused by frequency dependency in the MM environment in a QM/MM study of water clusters, and the effect of the dynamic local polarizability in a larger delocalized aromatic system.

The performance of the frequency dependent LoProp method was made in ref. 26 on the static and dynamic polarizability of the TIP3P water model and within a study of the size dependent behavior and convergence of the mean polarizability for different ensembles of water clusters. Two different cases were studied – the error in the mean polarizability using the modified Silberstein–Applequist model for water in gas phase was calculated by means of QM response theory, and the QM/MM response method was then used to obtain the absolute mean polarizability. That was done for the static, frequency independent case, and for frequency dispersion. Also the role of the actual decomposition scheme was explored, thus employing different water models for the polarizability, as well as the parameterization of the underlying response theory for the numerical outcome and convergence behavior of the polarizabilities. The LoProp scheme was used for construction of decomposed force fields with five main types typically used in QM/MM applications:[61] (i) atom distributed charges; (ii) atom distributed charges with a central polarizability; (iii) atom distributed charges and polarizabilities; (iv) atom and bond distributed charges and polarizabilities; (v) atom and bond distributed charges and

polarizabilities with multipoles expanded at each atomic site (normally truncating at octopoles). These different MM models have been particularly well tested in the case of the water solvent. The precision of the more refined models is associated with a computational cost, and is also directed by the type of property considered in the QM/MM calculations. It has been noted on several occasions that a good MM parameterization of water (and possibly any other solvent), relieves the need to include waters in the QM box, thereby making the QM/MM partitioning remain well-defined over simulation time, and substantially reducing cost.[62]

It was found in ref. 26 for water that the variation in the frequency dependent polarizabilities gives only a small perturbation of the QM wave function in QM/MM calculations and, indeed, that the distributed LoProp point dipoles outperform the molecule-centered ones by 10%. Furthermore, it was shown that the bulk frequency dependent $\alpha(\omega)$ prediction is 25% better for properties obtained at the corresponding frequency in comparison with the properties obtained in the static limit.

QM/MM results for the water clusters are summarized in Fig. 2 and 3. It is seen that the effect of including frequency dependent polarizabilities in the MM region does not affect the $\bar{\alpha}^{\text{Cluster}}$ with any significance. If the amount of water molecules in the MM region is small, then the mean polarizability for one QM molecule is larger than for the gas phase, but as the MM region is systematically expanded, the mean polarizability approaches the gas-phase value for the oxygen-centered type, and only a slight underestimation of the gas-phase value is seen. For one water molecule in the QM region, the QM/MM interaction stagnates at about 30 molecules in the MM region, which shows that the QM/MM interaction is only sensitive to the most close-lying MM segments.

As an example of a bonded polymer the frequency dependent mean polarizability for a tryptophan residue, embedded in an MM polarizable environment, was investigated in ref. 26. In Fig. 4 the frequency dependence of the isotropic $\alpha^{\text{LoProp}}(\omega)$ for all the atoms in tryptophan is summarized. The atom or group contribution was given to the total polarizability over a number of snapshots.
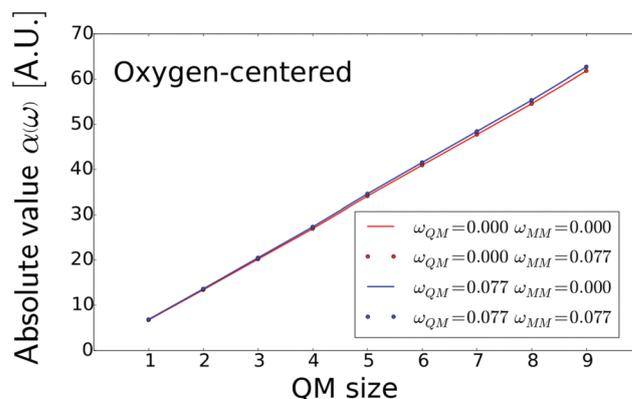

Fig. 2 $\bar{\alpha}(\omega)$ obtained by varying the size of the QM region (*x*-axis denotes number of QM waters) for in total 100 water molecules. From ref. 26. Copyright Royal Academy of Sciences.

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2259**
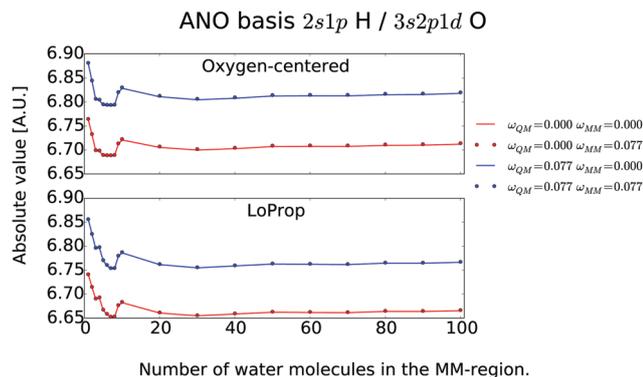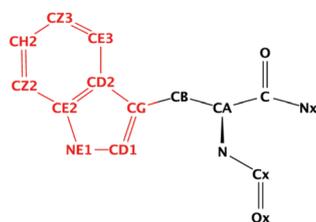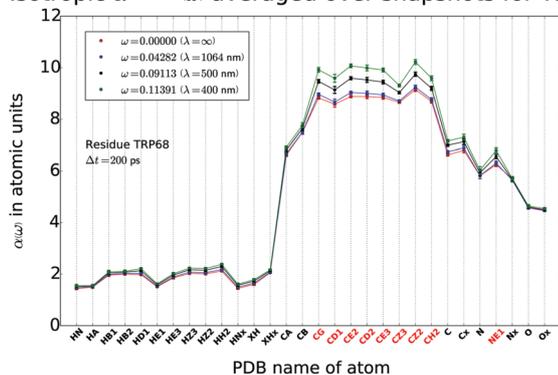
ANO basis $2s1p$ H / $3s2p1d$ O

Fig. 3 $\bar{\alpha}(\omega)$ obtained by varying the size of the MM region for one QM water molecule. From ref. 26. Copyright Royal Academy of Sciences.

(a) The tryptophan residue capped with $HCO-$ and $-NH_2$ groups.

Isotropic $\alpha^{LoProp}(\omega)$ averaged over snapshots for TRP68

(b) 200 ps spacing.

Fig. 4 $\alpha^{LoProp}(\omega)$ averaged over 10 snapshots for the residue TRP68. From ref. 26. Copyright Royal Academy of Sciences.

For larger systems, and especially polarizable systems, the effect of frequency dependence in the MM environment is expected to become non-negligible, especially if the frequence close up to resonances. The decomposition can directly be implemented with any quantum mechanical wave function and basis set, an extension of the studies above for any kind of systems is in principle possible and the frequency independent force fields in QM/MM calculations can routinely be replaced by frequency dependent ones. In a recent paper by Norby et al.[63] it was shown that the introduction of a frequency-dependent embedding potential leads to further model complications upon solving the central QM/MM equations defining specific molecular properties. It was also shown that from a numerical point of view that the consequences of using such a frequency-dependent

embedding potential is in general small except in cases if the absorption bands of the environment are close to that of the region treated using quantum mechanics.

4 Decomposition of hyperpolarizabilities

The LoProp transformation of charges, multipole moments and multipole polarizabilities presented above, thus offers a way to go beyond numerical fitting of parameters to physical observables—that holds also for non-linear properties, including the first order hyperpolarizability. In the work of Harczuk et al.[27] hyperpolarizabilities were obtained through combining the LoProp transformation and the second-order Applequist formalism, in which the point dipoles are additionally described by their non-linear polarization properties, i.e. also including the first hyperpolarizability to the original Applequist equations (for mathematical details see Theory section). The equations for the LoProp hyperpolarizabilities were extracted from the second order response of a molecule subjected to an external field using analytical response theory, where the higher order terms to the dipole moment induction thus were obtained from the non-linear contribution of the first hyperpolarizability. The theory builds on the afore mentioned original formulation of Silberstein, later developed by Applequist et al., where the interaction of point dipoles is described by point polarizabilities. The work presented in ref. 27 is the first known classical formulation of non-linear induction of point dipoles and is as such a natural extension of the earlier many-interacting polarizable point-dipoles models. As the formulation describes the molecular and cluster polarizability and hyperpolarizability of interacting molecules through their atomic site properties, it can so be used to evaluate macroscopic first order hyperpolarizabilities, something which is computationally expensive and most often unfeasible for large systems using purely ab initio methods.

For clusters of moderate size the validity of the model can be tested with a full quantum mechanical calculation for smaller oligomers. As for the corresponding polarizability, it can be expected that these models hold for well separated systems, while breaking down at close distances where the exchange interaction between the overlapping electron densities is non-negligible. In the work of Harczuk[27] applications were made on water clusters because of its anomalously varying first order hyperpolarizability with respect to phase, originally observed by Ward and Miller[64] and by Levine and Bethea,[65] and later calculated for the first solvation shell by Mikkelsen et al.[66,67] The water dimer and pentamer systems and clusters up to 25 water molecules were used as a fist benchmark of the methodology.[27] A large-scale calculation on a water cluster containing 500 water molecules was finally performed, which demonstrated that the proposed method can lead to opportunities in property predictions of large, previously unattainable, systems. At small intermolecular distances, the overlap between the electronic wave functions of the water molecules is non-negligible, and thus the

2260 | Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270
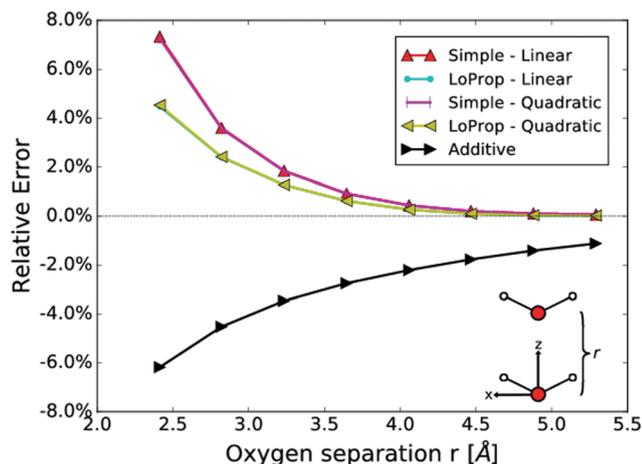
This journal is © the Owner Societies 2019

Fig. 5 The relative error of the total induced dipole moment for two water molecules as a function of the intermolecular distance. The additive method includes no polarization of the point-dipoles. From ref. 27. Copyright Royal Academy of Sciences.

classical model breaks down, failing to describe response properties in an accurate manner. However, at distances when the classical model is expected to give large errors, $r < 2.5$ Å, the oxygen in the water molecules, responsible for most of the polarization due to an external field, rarely encounter each other, which can be seen by the oxygen–oxygen radial distribution function obtained from X-ray diffraction[68] of liquid water at room temperature and pressure. It may thus indicate that it is still valid to compute the optical response using a classical formalism for liquid water.

In Fig. 5, the relative error of the total dipole moment for two water molecules in the same plane at a variable separation is plotted *versus* intermolecular distance. The relative error in the total induced dipole moment, referring to properties situated at the centre of the water molecules, is seen to be largest for the simple model, while the LoProp model decreases the error significantly. With the resolution given in the figure, the two LoProp curves overlap and the two "Simple" curves overlap almost complete. It is notable that the asymptotic behaviour is the same for all models and approaches the QM result at large distance. The zeroth model, which does not include any interactions, perfectly aligns with the additive method, which is a summation of the molecular dipole moments as defined by the LoProp basis.

In Fig. 6 the absolute value of the isotropic $\beta_{\parallel}$ component, calculated with the TDHF/TDDFT methods, is plotted against the number of molecules in water clusters. The TDHF and the TDDFT methods give quite different results for the QM part, but share the characteristics that the value of $\beta_{\parallel}$ decreases for $N < 16$ and then rapidly increases by a large magnitude. As the water cluster configuration is taken from a random MD snapshot, the properties will depend on the unique geometry, but the difference in QM methods is still quite noticeable. We see in Fig. 6 that the LoProp model agrees better with TDDFT using the B3LYP functional than with TDHF. The latter method greatly underestimates the first hyperpolarizability in conjugated systems,
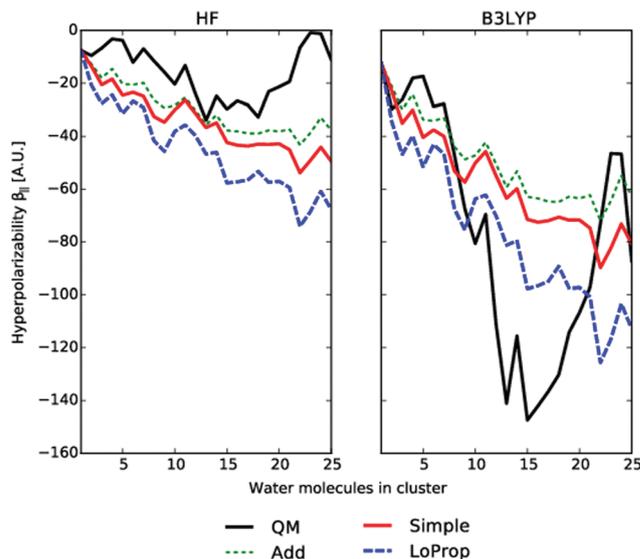

Fig. 6 The absolute value of the parallel component of the first hyper-polarizability calculated for clusters containing varying amounts of water molecules. The basis set used is the 2s (H)/4s3p1d (O) ANO type. From ref. 27. Copyright Royal Academy of Sciences.

while TDDFT often yields too large hyperpolarizabilities due to its underestimation of the excitation energy levels. These trends are well known and refers in general to overestimation, respectively, underestimation of the HOMO–LUMO gaps. As the hyperpolarizability changes its sign for the small but non-zero β_{yyz} component, it will affect the induced polarizabilities in eqn (45) differently from the TDHF. The reverse sign of the β_{yyz} component for TDDFT could thus be an additional reason why TDDFT performs differently from TDHF in this case. Likewise, since all the polarizability and hyperpolarizability components increase in magnitude from TDHF to TDDFT, the quadratic Applequist model yields larger values of the local fields, leading to an increased hyperpolarizability $\beta_{\parallel}$.

In Fig. 7, results are shown for calculations of all properties using the quadratic Applequist formalism for a cluster consisting of 500 water molecules. It can be seen that the polarizability is not so heavily affected by the long-range interactions in this model, as it converges fairly quickly, dropping 2.9% from the case where only the closest lying neighbours affect each other. For the dipole moment, the situation is different, as including the interaction of more distant lying particles can change the value quite drastically. For the hyperpolarizability, the situation is similar to the dipole moment but with the difference that it oscillates more strongly in absolute value and even changes its sign at intermediate values of the cut-off screening. This phenomenon was previously confirmed by multi-configuration self-consistent reaction field theory,[66,67] where the observed sign change was shown to occur for the water pentamer. In the quadratic Applequist method, this intermediate sign change occurs due to the lack of inclusion of long range fields for the inducible point dipoles. At full interaction, we see that all properties converge, and that the first hyperpolarizability takes a negative value.
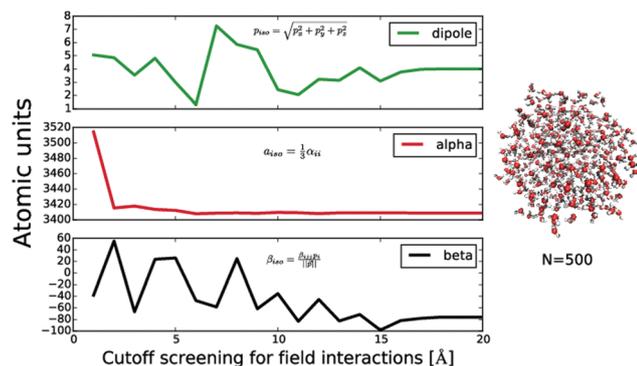
This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2261**

Fig. 7 Convergence of the isotropic analogues of the dipole moment, polarizability, and hyperpolarizability of a snapshot cluster with 500 water molecules. The *x*-axis represents the range of the field that the inducible point dipoles are affected by. Furthermost to the right corresponds to all particles inducing each other. From ref. 27. Copyright Royal Academy of Sciences.

The first hyperpolarizability is enhanced in systems with delocalised electrons, and in particular, push–pull conjugated systems. The possibility to compute cluster hyperpolarizabilities is affected by a few factors, first of all by the computationally demanding quadratic response theory calculation, which must be performed to provide LoProp hyperpolarizabilities. Also the convergence of the iterative calculations of the induced dipole moments of all the point dipoles could be demanding. This process scales approximately as $\mathbb{O}(N^2)$ with respect to the amount of interacting point dipoles. Finally solving the linear system of equations may take time depending on the scaling techniques applied.

5 Decomposition of dispersion interaction

As shown in the previous sections the LoProp scheme can be applied to different kinds of properties and intermolecular forces that describe how molecules attract or repel each other. One such force is the attractive long-range intermolecular dispersion force which in general is very subtle and weak at short ranges compared to the electrostatic, polarization or short-range terms, but which can dominate in systems which are non-polar or even lack static multipole moments. Most of the additive force fields used today[69–71] include electrostatic interaction, some include polarization but most neglect the dispersion, although it is clearly essential in many circumstances, especially for weakly bound systems.[72] The long-range contributions to the dispersion energy can be attributed to the vacuum fluctuation of the quantized field, as predicted by the so-called by the Casimir–Polder integral[73,74] and can be viewed as "instantaneous and mutual polarization" between two particles at a distance. In practice it is almost always modelled by a Leonard-Jones potential, where the sole distant dependent parameter often is obtained by fitting to heats of vaporization or solvation energies taken from experiments. The dispersion energy is inherently included in most energy correlated *ab initio* electronic structure

methods and various types of many-body interaction schemes, like Møller–Plesset perturbation theory and coupled cluster theory. More recently, various density functional schemes have been applied, notably the one of Stone and Misquitta that implement intermolecular contributions including the dispersion energy by the means of the SAPT(DFT) (Symmetry Adapted Perturbation Theory-Density Functional Theory). This method gives energy contributions of monomer–monomer perturbations to various degrees, including those of dispersion origin, and has been applied for a large variety of systems.[75–79]

In ref. 51 a new way was described to compute the two-body contribution to the dispersion energy using *ab initio* theory and using the philosophy of the LoProp scheme. By combining the complex polarization propagator method and the LoProp transformation, local contributions to the Casimir–Polder interaction could be obtained. The anisotropic as well as isotropic models of the dispersion energy could thereby be analyzed in dimer systems using the decomposition scheme for the dipole–dipole polarizability. We recapitulate that the linear dipole–dipole term defines the well-known C_6 coefficient for the dispersion interaction:

$$C_n \propto \int_0^\infty \alpha_{tt'}^{(1)}(\mathrm{i}\omega)\alpha_{uu'}^{(2)}(\mathrm{i}\omega)\mathrm{d}\omega \qquad (83)$$

where $\alpha_{tt'}^{(1)}(\mathrm{i}\omega)$ is the polarizability for site 1 calculated at an imaginary frequency iω. The lower indices of α are angular momentum labels and the order n in the C_n coefficient is defined as $n = t + t' + u + u' + 2$, giving the C_6 term from the linear dipole–dipole polarizability ($t = t' = u = u' = 1$). For the derivation of this equation we refer to the Theory section. The second ingredient in the scheme is the complex polarization propagator (CPP) approach by Norman and coworkers[80,81] which implements analytical response theory to calculate the isotropic C_6 dispersion coefficient. In this approach the expression of the polarizability is transformed as a function of real frequencies only to generate the real and imaginary parts *via* the Cauchy moments,[80,82,83] which is possible because the general complex polarizability is a function of complex frequencies. Harzcuk *et al.* thus applied the CPP approach to calculate the Casimir–Polder integral of the isotropic C_6 coefficient where the response vectors from the CPP code were used in order to obtain the distributed real polarizabilities of atoms at imaginary frequencies. That made it possible to decompose the molecular C_6 coefficients into pairs of atomic sites, giving rise to a total dispersion in a dimer system at various relative orientations. The so obtained LoProp decomposed polarizabilities are additive and sum up to the molecular polarizability, and the sum of all LoProp C_{ij6} elements is thus equal to the molecular C_6 coefficient. The LoProp dispersion calculated that way could be compared to the dispersion calculated between the centers of mass of the dimers, and with high level of theory, like the above mentioned SAPT(DFT) method.

Fig. 8 illustrates basis set convergence of LoProp decomposed atomic and total molecular C_6 coefficients.

Results for dispersion interaction energies calculated with the anisotropic models were presented and analyzed in ref. 51
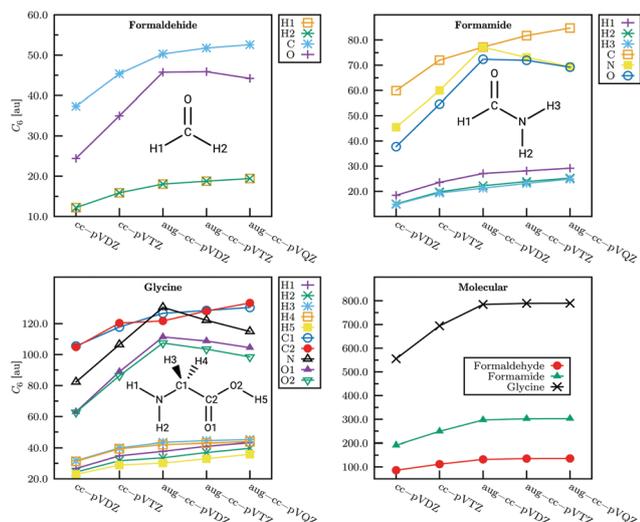
2262 | *Phys. Chem. Chem. Phys.*, 2019, **21**, 2251–2270

This journal is © the Owner Societies 2019

Fig. 8 Basis set convergence of LoProp decomposed atomic and total molecular C_6 coefficients. From ref. 51. Copyright Royal Academy of Sciences.

for different orientations of H2, N2, CO, methane, pyridine and benzene dimers. For the diatomic dimers the LoProp results agree well with the reference values—for some (unplausible) orientations the interaction is notably underestimated. By stacking the dimers the energy is higher than for the SAPT(DFT) method but agree better at larger distances (Fig. 10 and 11).
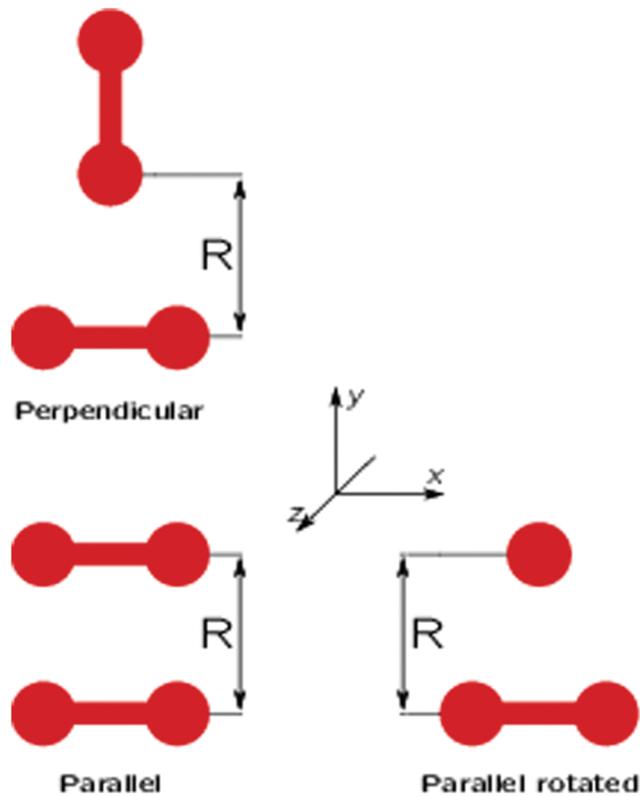
Fig. 9 Schematic representation of the relative orientations for the N2–N2 dimer system. From ref. 51. Copyright Royal Academy of Sciences.
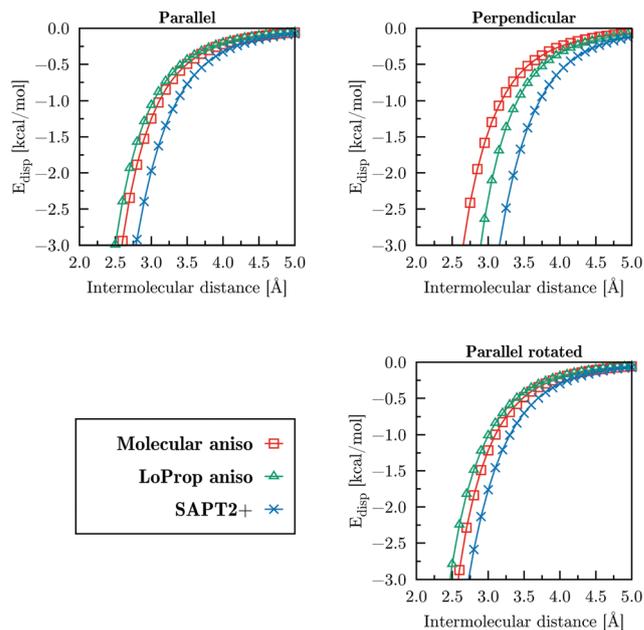

Fig. 10 Comparison of SAPT2+ results with the anisotropic dispersion energy models computed for the nitrogen dimer. Fig. 9 defines the structures. From ref. 51. Copyright Royal Academy of Sciences.

The results indicate that the leading energy term calculated from the dipole–dipole polarizability tends to underestimate the reference data for the molecular dimers, and display scatter for the methane and pyridine dimers. The convergence of the isotropic atomic and summed molecular C_6 coefficients as a function of the basis is in general found satisfactory, see Fig. 8. It can be seen in the figure that the augmentation is also necessary for dispersion coefficients and consequently, polarizabilities.

Some of the remaining discrepancies can probably be assigned to higher order multipole interactions, providing C_8 and C_{10} energy terms, the inclusion of which into the decomposition scheme is a rather straightforward generalization that would improve the performance. Also the selected interaction points can be generalized to bond points. Corresponding decomposition of charges and polarizabilities have shown in some cases the importance of such a fine granulation of the interaction. Since the contribution of the dispersion energy contribution to the two-body potential is significant for large systems, and since the LoProp transformation gives transferable properties between similar groups between different systems, these results could be useful for fitting of new accurate force fields with applications to, *e.g.*, biological DNA/protein systems. In all, the LoProp decomposition procedure as applied to dispersion interaction in ref. 51 is a promising way to design new force fields derived from *ab initio* reference data and to introduce localized dispersion interaction into quantum-classical hybrid methods like QM/MM methods. It may be argued that the incorporation of atomic C_6 coefficients in new atomic force fields will have important ramifications in molecular dynamics studies of biomolecular systems.

This journal is © the Owner Societies 2019

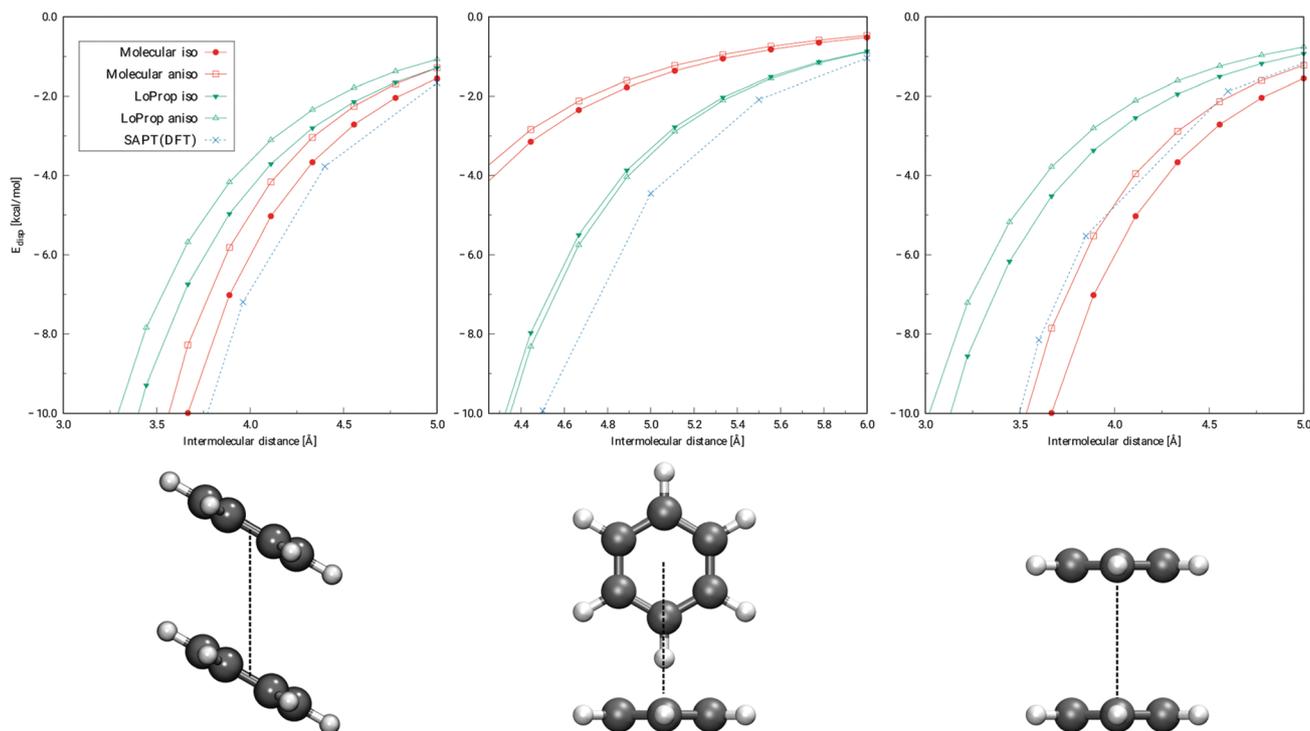*Phys. Chem. Chem. Phys.*, 2019, **21**, 2251–2270 | **2263**

Fig. 11 Dispersion energies for benzene dimers M1, S3, and S8, from left to right, respectively. SAPT(DFT) reference points were taken from ref. 84. Other results from ref. 51. Copyright Royal Academy of Sciences.

6 Hyperpolarisability of collagen

The generalization of the decomposition technique to include non-linear properties has a strong motivation in the fact that non-linear optical properties of materials have progressively become more accessible and attended during recent years. That includes also biostructures and biomolecules like peptides and proteins. For such systems the inherent requirement of electron correlation in quantum mechanical methods still sets limits on the size of the systems that can be investigated at high levels of accuracy, thus promoting the need for coarser but more large-scale applicable methods to be developed.[85,86] An example is the above mentioned point dipole model, originally proposed by Silberstein,[35,36] and later used by Applequist et al.,[37] which has found applications in prediction of molecular polarisabilities, inter-molecular potentials,[87] second order susceptibilities $\chi^{(2)}$,[88] and third order susceptibilities $\chi^{(3)}$,[89] which are examples of relevant development work. It has furthermore recently been shown that the distribution of molecular properties based on the so-called "quantum theory of atoms in molecules" (QTAIM) method[90] could be used for the calculation of $\chi^{(2)}$ in crystal packings.[91] Recently, Harczuk et al. showed that their derivation of decomposed non-linear polarisabilities with the LoProp-Applequist method derived in ref. 27 provides a vehicle to predict the non-linear optical characteristics of large molecular peptides and proteins.[29] The capability of the method was demonstrated for the collagen triple helix, the most abundant structural protein in the human body, which in recent years has drawn attention due to its high first order

non-linear response and for which second harmonic generation (SHG) imaging have been used for heart disease diagnostics and other biosensor applications.[92–96] A procedure with molecular fractionation using conjugate caps in order to determine the atomic and bond contributions to the net β tensor of the collagen $[(PPG)_{10}]_3$ triple-helix was then employed from which the intensity of the β_{HRS} signal and the depolarization ratios could be derived. By furthermore using Thole's exponential damping modification[53,54,97] to the dyadic tensor in the Applequist equations, the hyperpolarisability could be predicted and compared with experiment. The LoProp transformation was thus used for the atom and bond decomposition of the molecular properties of each residue, in combination with the so-called MFCC[98,99] procedure for cutting at the peptide bonds, which account for the properties of the helix without any long-range interactions between the residues. The second order quadratic point-dipole model is then used to calculate the inter, and intra-chain hyperpolarisation, between and inside the chains. The total scattering intensity and depolarization ratio, which are rotationally invariant and measurable, were then computed from the resulting $\beta^{Collagen}$ tensor and compared to previous experiments and calculations. This provided, for the first time, the non-linear extension of Applequist's equations to covalently bonded fragments.

Measurements of the hyper-Rayleigh scattering in collagen[100] together with additive approximate models,[101,102] and more advanced models as the so-called ONIOM model,[103] have related the large value of β to the stacking of the peptide bonds between the (hydroxy)-proline and glycine residues. The specific rigidness

2264 | Phys. Chem. Chem. Phys., 2019, 21, 2251–2270
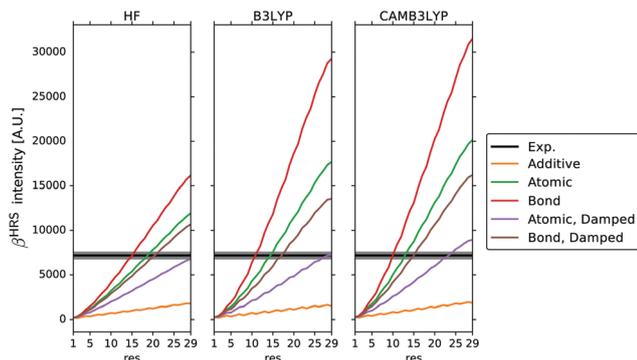
This journal is © the Owner Societies 2019

Fig. 12 The total static scattering intensity β_{HRS} for a growing collagen triple helix. For 29 residues, the rat-tail collagen is obtained. From ref. 29. Copyright ACS Publishing.
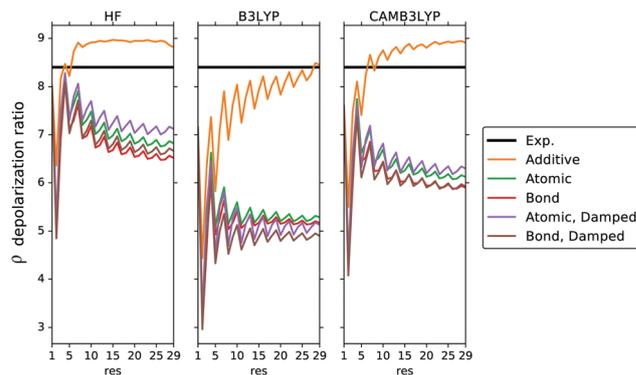

Fig. 13 The depolarization ratio ρ for the growing collagen triple helix. For 29 residues, the rat-tail collagen is obtained. From ref. 29. Copyright ACS Publishing.

caused by the [PPG] subunit has been used to explain the build up of the tensor component in the longitudinal direction of the chains, something which could not be found when modelling the [GGG] trimer. The results from an additive scheme was included, in which all the components of the dipole coupling tensor are set to zero, and the Applequist properties are the sums of the localized atomic or bond polarisabilities, and hyperpolarisabilities, respectively.

In Fig. 12, the β_∞^{HRS} scattering intensity is plotted as a function of the amount of residues of each chain. The model peptide $[(PPG)_{10}]_3$ is approximately obtained with 29 residues in each chain. This collagen model has 87 residues in total, since the X-ray structure has one glycine missing in the C terminal of each chain.

The first hyperpolarisability for the amino acid constituents of the collagen triple helix was calculated employing time-dependent Hartree–Fock and density functional theory. The atomic decomposition stemming from TD-DFT/B3LYP using Thole's damping shows the best agreement with experiment. The corresponding value is for TDHF, and TD-DFT/CAMB3LYP lower, respectively, higher, see Fig. 12. The results thus show that when damping is introduced the atom + bond polarisability scheme not necessarily is better than the pure atom polarisability scheme for which the damping is parametrised. Further reasons for this are given in ref. 29. Another trend seen in Fig. 12 is that the scattering intensity grows in a slight sigmoidal shape for all the models, whereas the additive scattering, *i.e.* just taking the sum of the localized hyperpolarisabilities, grows linearly. Adding more residues would thus not necessarily produce a growth in intensity, but might converge with respect to the chain length. It is clear that when damping is included, the intensity is not as sensitive to the computational method used to obtain the properties.

Fig. 13 shows all the methods tested and all the models used for the depolarization ratio. It can be seen that the depolarization oscillates more wildly than the scattering intensity. The experimental value of 8.4 hints to the fact that the collagen helix is highly dipolar, with no significant octupole character. The reason why the quadratic Applequist model underestimates the dipolar nature of this collagen model might be due to that some components such as β_{xxx} and β_{yyy} in the collagen tensor are over-induced.

It is notable that TD-B3LYP underperforms with respect to TDHF and TD-CAMB3LYP which can be attributed to the long-range exchange being important for the calculation of the non-diagonal hyperpolarizability components using the quadratic Applequist equations. Calculations in ref. 101 and 102 elucidated that the successive stacking of the π-character peptide bonds was the main reason behind the large second-order response in collagen, something that also could be confirmed in the work of Harczuk *et al.*[29]

A motivation for the collagen work[29] can be found in the general ambition to derive models to efficiently predict properties of large biostructures with an accuracy that allows to optimize them for a particular property by structural modifications like substitutions or residual mutation.

7 Rayleigh scattering of atmospheric nanoparticles

In paper[28] yet another application of the atomic decomposition scheme for polarisabilities was presented, namely for Rayleigh scattering of naturally polarised light by systems with atmospheric relevance, in particular by growing water clusters containing foreign compounds. The Rayleigh scattering intensity relates directly to the dynamical polarisability tensors of the clusters and could so be obtained by the combined LoProp-Applequist algorithm reviewed in the Theory section and in the original work.[26] The study found motivation in that various particles act as cloud condensation nuclei in the atmosphere with important climate implications, like green house effect and earth's albedo,[104,105] where elastic scattering of light in the atmosphere leads to cooling.

The Rayleigh scattering intensity can be obtained[106] for naturally polarised light as

$$\mathscr{R}_n = 45(\bar{\alpha}^2) + 13(\Delta\alpha^2) \tag{84}$$

where the isotropic ($\bar{\alpha}$) and anisotropic ($\Delta\alpha$) polarisabilities are defined as

$$\bar{\alpha} = \frac{1}{3}\alpha_{ii} \tag{85}$$

$$\Delta\alpha = \sqrt{\frac{3\alpha_{ij}\alpha_{ij} - \alpha_{ii}\alpha_{jj}}{2}} \quad (86)$$

For a spherical aerosol particle interacting with light of wavelengths approximately equal to the radius of the particle, Mie scattering constitutes the predominant scattering process, while when light interacts with aerosol particles with characteristic sizes of less than one tenth of the wavelength of the light, Rayleigh scattering guides the predominant elastic scattering effect. In Rayleigh theory, the scattering is proportional to the frequency of the incident light to the fourth power, manifesting the blue colour saturation of the sky. Thus the frequency dependent polarisability of atmospheric clusters relates to the Rayleigh scattering tensor, something that is addressable directly with the LoProp-Applequist models. It could form, as investigated in ref. 26, an excellent approach for Rayleigh scattering of particles in the nanoscale covering the scattering intensities of various incident light frequencies. Going to the other limit, to very small clusters, quantum chemistry methods can be used both for the scattering problem and for the nucleation of the clusters, a recent example being the work of Elm *et al.*[106,107] who used quantum mechanical response theory and a combinatorial sampling approach to study the Rayleigh and hyper-Rayleigh scattering properties of some binary and ternary water clusters containing sulfuric acid and ammonia. The molecular Rayleigh, and hyper-Rayleigh, scattering intensities were expressed from the total dipole polarisability α and hyperpolarisability β tensors allowing for the exploration of the effect of cluster morphology on the scattering properties. It was found that the Rayleigh scattering intensity depends quadratically on the number of water molecules in the cluster and that a single foreign molecule (ammonia) could generate high anisotropy, which further increased the scattering intensity. The hyper Rayleigh scattering activities were found to be extremely low. That work thus gave a molecular level understanding of scattering properties of small clusters, and on the impact of foreign constituents on water clusters on Rayleigh scattering properties. As the atmosphere contains a large body of neutral molecular clusters with sizes in the nanometre scale[108–111] for which the Rayleigh scattering is the dominating mechanism, it is desirable to reach the nanoscale for Rayleigh scattering simulations as achievable by the LoProp-Applequist approach.[26,37] The choice in ref. 26 of atmospheric water droplets formed by adsorbed *cis*-pinonic acid adsorption (CPA) was motivated by the atmospheric abundance of this compound and that it is known to decrease the water droplet surface tension, leading to smaller droplet formations, causing a larger density of small droplet sizes in the clouds, which increase the overall albedo and gives a net cooling of the Earth.[112] Using the LoProp-Applequist model it was then possible to scrutinise details with respect to contributions of molecular interaction, cluster size, mass constituent and dispersion on the total Rayleigh scattering. Studies were made for cluster diameters up to 5 nm, and for frequencies below the first resonance frequency, thus within the eligibility of Rayleigh scattering theory.

In Fig. 14, the increase in scattering with respect to cluster size given by the Applequist model is compared to the polarisability
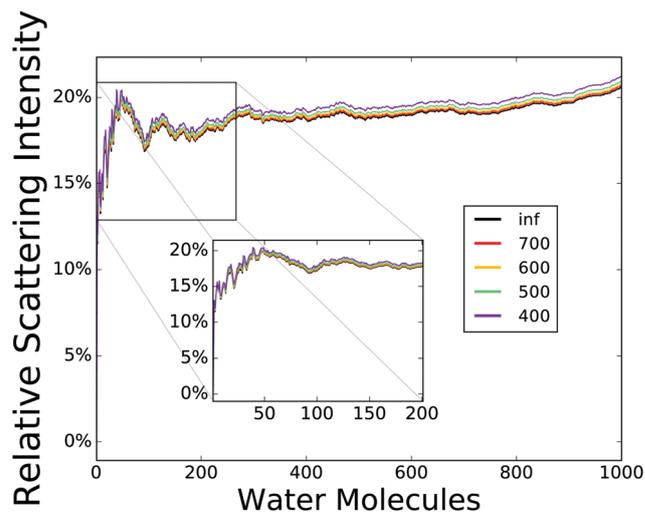

Fig. 14 The relative increase in scattering intensity from the Applequist equations for a growing droplet with different amounts of water molecules in the cluster. From ref. 28. Copyright ACS publications.

obtained as a summation of the polarisability components of each molecule, not taking interaction into account. All the wavelengths follow the same growth pattern, with 400 nm giving the largest increases. For the absolute magnitude of the scattering intensity the total scattering intensity for the different wavelengths depends on the cluster growth of water molecules—here the scattering intensity for all the wavelengths grow quadratically. The difference between the wavelengths also increases quadratically with respect to the static frequency. This shows that the dynamic polarisability contribution to the total scattering grows quadratically with respect to the frequency of the external field, but that the major contribution is due to the static polarisability. As shown in Fig. 15, the scattering exhibits a weak quadratic increase with respect to the concentration of CPA in the water droplet. The weak increase
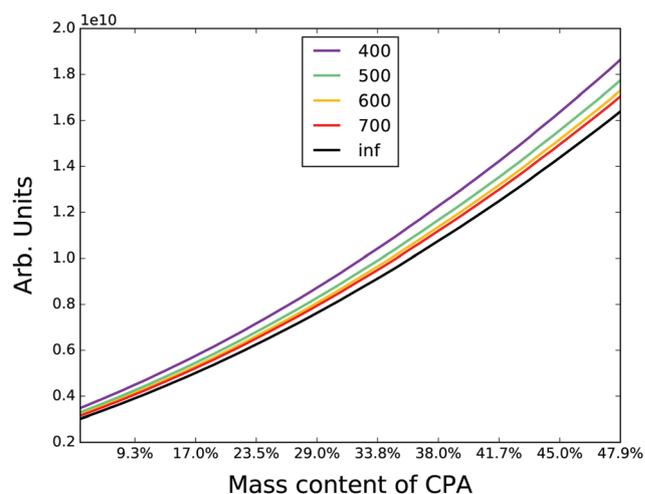

Fig. 15 Absolute Rayleigh-scattering intensity for different wavelengths as a function of *cis*-pinonic acid mass content. From ref. 28. Copyright ACS publications.
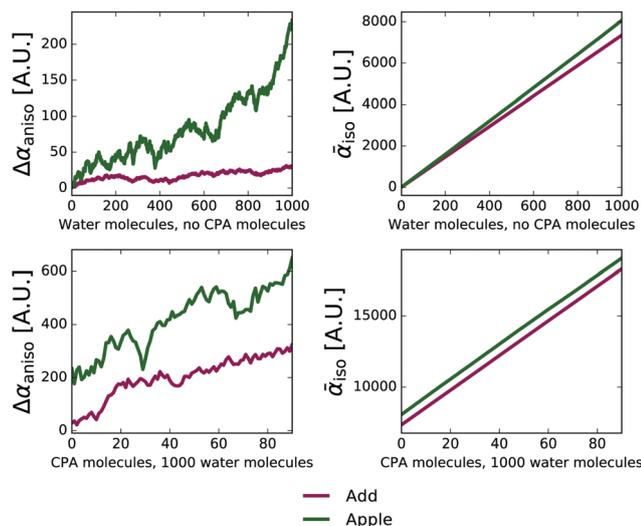
2266 | *Phys. Chem. Chem. Phys.*, 2019, **21**, 2251–2270

This journal is © the Owner Societies 2019

Fig. 16 The static isotropic and anisotropic component of the polarisability at static external field. From ref. 28. Copyright ACS publications.

could be attributed to CPA avoiding the water, something which also causes it being pushed out of the water droplet sphere, residing at the surface. Close-range interactions will cause large induced polarisation, which could explain why the change in scattering intensity is almost linear with the addition of CPA to the water droplet.

In the Fig. 16, the isotropic and anisotropic contributions to the net scattering are plotted for the water only cluster, and the water cluster interacting with CPA molecules, respectively, at static frequency. It can be seen that the major contribution to the Rayleigh scattering of the pure water clusters is due to the isotropic component, which is greatly enhanced in the range of 1–1000 water molecular clusters. It is furthermore gradually increased using the Applequist equations, as compared to the non-interacting model. The anisotropic component is rather stable for the water cluster growth, but increases sharply for clusters of sizes larger than 700 water molecules. For the clusters containing 1000 water molecules and adsorbed CPA molecules, also the anisotropic component grows steadily. The Rayleigh scattering of naturally polarised light was found to increase smoothly with the increase of the overall cluster size, with the increase of mass constituent of *cis*-pinonic acid molecules, and with dispersion (increase of wavelength). The calculations predict that weakly adsorbed molecules may cause a lowering of the net scattering intensities compared to strongly adsorbed ones, overall affecting the albedo of clouds consisting of these types of cloud condensation nuclei.

The Applequist interaction was found to yield scattering intensities 20% larger for a cluster consisting of 1000 water molecules compared to the method where all the polarisabilities of molecules are added without interactions. It was confirmed that the scattering intensity depends quadratically on the number of water molecules in the cluster, and that it also increases quadratically with increase for the mass constituent of the foreign substance. The adsorption of the *cis*-pinonic acid was found to increase the contribution to the scattering intensity

stemming from the anisotropic polarisability, as compared to the isotropic contribution. Paper[26] indicates that large clusters with organic molecules in water is a realistic proposition for hundreds of molecules (organic and water) in the clusters and the method is shown to be both sufficiently accurate and general for any type of cluster of molecules up to size of 10 nm, which transcends already the Rayleigh scattering region and goes into Mie scattering region. That was probably the first time large clusters can be reached by a molecular-level based calculation of the Rayleigh scattering, well beyond the small ones reachable by quantum chemistry.

8 Conclusion

In the present work we have reviewed recent development of local decomposition schemes for localizing molecular properties, and highlighted their use either in QM/MM (Quantum Mechanics Molecular Mechanics) calculations or for the purpose of generating "classical" properties of large clusters. A few sample applications of such decomposition have been recapitulated and briefly discussed. The highlighted LoProp procedure was originally developed to obtain granulated solute–solvent interactions using a scheme for decomposing the quantum chemical overlap matrix and to use the resulting transformation matrix to localize a given property into a predefined level of localization. Recent extensions, highlighted in this review, of the technique comprise locally decomposed hyperpolarisabilities,[27] frequency dependent polarisabilities,[26] and locally decomposed frequency dependent complex polarisabilities.[51] The LoProp scheme has become instrumental for many solvent interaction models, and in particular for implementations of QM/MM (quantum mechanics molecular mechanics) methods. In addition, the LoProp technique has recently been shown to be a powerful tool for building up properties of large molecular clusters, like for Rayleigh scattering of atmospheric nanoparticles[28] and hyperpolarisabilities (second harmonic generation) of proteins.[29] The flexibility of the model to address localized interactions with atomically decomposed charges and polarisabilities has thus been critical for the accuracy of many structure and property calculations by means of QM/MM. The flexibility not only concerns the position of the interaction points but also the multipolar expansion of charges and polarisabilities. By means of LoProp several systematically connected models for the solvent interaction can be formulated, ranging from simple molecular point charges to bond and atom centred multipole expanded charges and polarisabilities. The precision needed, and thus the level of decomposition, is to a large extent dictated by the property that is to be calculated. The level of decomposition also has a bearing on the standard problem how to split up the QM and MM parts of the system in a QM/MM calculation, for instance, the need to bring in water molecules in the QM part is often relieved by a careful LoProp MM parametrisation of water.[3] To the known limitations of the LoProp method one can mention the tendency to "overpolarise" the interaction, especially to the QM part close to the QM-MM interface but also between closely encountering MM molecules.

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | **2267**

The (*ad hoc*) introduction of distributed point charges and polarisabilities by Gaussian broadenings remedies the problem to some extent.[16,53] It goes without saying that the LoProp model does not handle charge transfer unless special parametrization for that is derived (*e.g.* in terms of capacitance parameters as in the QMCMM model mentioned above).

The flexibility of the LoProp method makes it open ended for further fine-tuning of accuracy, thus to multipole expand charges and polarisabilities, and include higher order dispersion coefficients, which is straightforwardly doable if the underlying property integrals are at hand. Such extensions would be the most natural measure of development of the present implementations in order to reach the ultimate precision of such a decomposition scheme. These implementations are in principle possible to introduce into QM/MM, although, to our knowledge, such a QM/MM has not yet been theoretically formulated for localized dispersion. Importantly, the presented scheme would allow to develop chemically dependent local dispersion interaction force fields in molecular dynamics, improving the commonly used force fields of atomically dependent Lennard-Jones interactions. It is desirable to "homogenize" the force fields utilized in the so-called integrated approach that combines MD with QM/MM, thus that the force fields in the two types of calculations are identical. Implementation of the schemes reviewed here into general MD simulation software would certainly not be straightforward "add-on" efforts, as current MD programs often are heavily parametrised towards certain reference data. The field of decomposition of molecular properties is thus wide open for further research, and important new applications for cluster properties, QM/MM and molecular dynamics can be envisaged.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 C. B. Nielsen, O. Christiansen, K. V. Mikkelsen and J. Kongsted, *J. Chem. Phys.*, 2007, **126**, 154112.

2 Z. Rinkevicius, N. A. Murugan, J. Kongsted, B. Frecus, A. H. Steindal and H. Agren, *J. Chem. Theory Comput.*, 2011, **7**, 3261–3271.

3 Z. Rinkevicius, N. A. Murugan, J. Kongsted, K. Aidas, A. H. Steindal and H. Agren, *J. Phys. Chem. B*, 2011, **115**, 4350–4358.

4 A. H. Steindal, K. Ruud, L. Frediani, K. Aidas and J. Kongsted, *J. Phys. Chem. B*, 2011, **115**, 3027–3037.

5 M. N. Pedersen, E. D. Hedegård, J. M. H. Olsen, J. Kauczor, P. Norman and J. Kongsted, *J. Chem. Theory Comput.*, 2014, **10**, 1164–1171.

6 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.

7 L. Adamowic, M. A. Freitag and M. S. Gordon, *J. Chem. Phys.*, 2003, **118**, 6725–6732.

8 I. V. R. Claudia, F. Jonathan and L. V. Slipchenko, *J. Chem. Phys.*, 2018, **149**, 094103.

9 L. V. Slipchenko, M. S. Gordon and K. Ruedenberg, *J. Phys. Chem. A*, 2017, **121**, 9495–9507.

10 Z. Rinkevicius, X. Li, J. A. R. Sandberg and H. Ågren, *Phys. Chem. Chem. Phys.*, 2014, **16**, 661–665.

11 A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128–1132.

12 E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, *J. Chem. Theory Comput.*, 2018, **14**, 4309–4319.

13 M. Scheurer, M. F. Herbst, P. Reinholdt, J. M. Haugaard Olsen, A. Dreuw and J. Kongsted, *J. Chem. Theory Comput.*, 2018, **14**, 4870–4883.

14 M. Scheurer, M. F. Herbst, P. Reinholdt, J. M. H. Olsen, A. Dreuw and J. Kongsted, *J. Chem. Theory Comput.*, 2018, **14**, 4870–4883.

15 Z. Rinkevicius, B. Frecus, O. Vahtras, J. Kongsted and H. Ågren, *J. Phys. Chem. B*, 2012, **8**, 257–263.

16 Z. Rinkevicius, X. Li, J. A. R. Sandberg, K. V. Mikkelsen and H. Ågren, *J. Chem. Theory Comput.*, 2014, **10**, 989–1003.

17 F. Iori and S. Corni, *J. Comput. Chem.*, 2008, **29**, 1656–1666.

18 F. Iori, R. Di Felice, E. Molinari and S. Corni, *J. Comput. Chem.*, 2009, **30**, 1465–1476.

19 O. Crescenzi, M. Pavone, F. De Angelis and V. Barone, *J. Phys. Chem. B*, 2004, **109**, 445–453.

20 L. Gagliardi, R. Lindh and G. Karlström, *J. Chem. Phys.*, 2004, **121**, 4494–4500.

21 N. A. Murugan, J. Kongsted, Z. Rinkevicius and H. Ågren, *Phys. Chem. Chem. Phys.*, 2012, **14**, 1107.

22 K. Aidas, A. Møgelhøj, C. B. Nielsen, K. V. Mikkelsen, K. Ruud, O. Christiansen and J. Kongsted, *J. Phys. Chem. A*, 2007, **111**, 4199–4210.

23 J. J. Eriksen, J. M. H. Olsen, K. Aidas, H. Ågren, K. V. Mikkelsen and J. Kongsted, *J. Comput. Chem.*, 2011, **32**, 2853–2864.

24 T. Löytynoja, I. Harczuk, K. Jänkälä, O. Vahtras and H. Agren, *J. Chem. Phys.*, 2017, **146**, 124902.

25 I. Harczuk, N. A. Murugan, O. Vahtras and H. Ågren, *J. Chem. Theory Comput.*, 2014, **10**, 3492–3502.

26 I. Harczuk, O. Vahtras and H. Ågren, *Phys. Chem. Chem. Phys.*, 2015, **17**, 7800–7812.

27 I. Harczuk, O. Vahtras and H. Ågren, *Phys. Chem. Chem. Phys.*, 2016, **18**, 8710–8722.

28 I. Harczuk, O. Vahtras and H. Ågren, *J. Phys. Chem. B*, 2016, **120**, 4296–4301.

29 I. Harczuk, O. Vahtras and H. Ågren, *J. Phys. Chem. Lett.*, 2016, **7**, 2132–2138.

30 J. Olsen and P. Jørgensen, *J. Chem. Phys.*, 1985, **82**, 3235–3264.

31 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, *J. Chem. Phys.*, 2002, **117**, 9630–9645.

32 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. B. O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. E. T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. F. L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I. M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. K. H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, P. Reine, Z. Rinkevicius, T. A. Ruden, R. K. Ruud, V. V. Rybkin, P. Sałek, C. C. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Stendal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 269–284.

33 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, *J. Chem. Phys.*, 2002, **117**, 9630.

34 J. Applequist, *J. Chem. Phys.*, 1985, **83**, 809–826.

35 L. Silberstein, *Philos. Mag.*, 1917, **33**, 92–128.

36 L. Silberstein, *Philos. Mag.*, 1917, **33**, 521–533.

37 J. Applequist, J. R. Carl and K.-K. Fung, *J. Am. Chem. Soc.*, 1972, **94**, 2952–2960.

38 J. Applequist, *J. Chem. Phys.*, 1993, **98**, 7664.

39 L. Jensen, P. O. Åstrand, A. Osted, J. Kongsted and K. V. Mikkelsen, *J. Chem. Phys.*, 2002, **116**, 4001–4010.

40 T. Hansen, L. Jensen, P. O. Åstrand and K. V. Mikkelsen, *J. Chem. Theory Comput.*, 2005, **1**, 626–633.

41 L. Jensen, K. O. Sylvester-Hvid, K. V. Mikkelsen and P.-O. Åstrand, *J. Phys. Chem. A*, 2003, **107**, 2270–2276.

42 L. Jensen, P.-O. Åstrand and K. V. Mikkelsen, *Nano Lett.*, 2003, **3**, 661–665.

43 C. E. Dykstra, *J. Comput. Chem.*, 1988, **9**, 476–487.

44 J. D. Augspurger and C. E. Dykstra, *J. Comput. Chem.*, 1992, **43**, 135–146.

45 J. M. Stout and C. E. Dykstra, *J. Am. Chem. Soc.*, 1995, **117**, 5127–5132.

46 C. E. Dykstra, *Chem. Phys. Lett.*, 1999, **305**, 132.

47 R. W. Munn, *Mol. Phys.*, 1988, **64**, 1.

48 M. in het Panhuis and R. W. Munn, *J. Chem. Phys.*, 2000, **113**, 10685.

49 O. Vahtras, *LoProp for Dalton*, 2014, DOI: 10.5281/zenodo. 13276.

50 V. Magnasco, G. Figari and C. Costa, *THEOCHEM*, 1992, **261**, 237–253.

51 I. Harczuk, B. Nagy, F. Jensen, O. Vahtras and H. Ågren, *Phys. Chem. Chem. Phys.*, 2017, **19**, 20241–20250.

52 A. Salam, *Appendix B: Rotational Averaging of Cartesian Tensors*, John Wiley & Sons, Inc., 2009, pp. 378–380.

53 B. Thole, *Chem. Phys.*, 1981, **59**, 341–350.

54 P. T. Van Duijnen and M. Swart, *J. Phys. Chem. A*, 1998, **102**, 2399–2407.

55 H. E. Watson and K. L. Ramaswamy, *Proc. R. Soc. London, Ser. A*, 1936, **156**, 144–157.

56 P. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.

57 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

58 J. Wang, P. Cieplak, J. Li, T. Hou, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3091–3099.

59 J. Wang, P. Cieplak, J. Li, J. Wang, Q. Cai, M. Hsieh, H. Lei, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3100–3111.

60 R. Bosque and J. Sales, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1154–1163.

61 K. Aidas, A. Møgelhøj, E. J. K. Nilsson, M. S. Johnson, K. V. Mikkelsen, O. Christiansen, P. Söderhjelm and J. Kongsted, *J. Chem. Phys.*, 2008, **128**, 194503.

62 Z. Rinkevicius, N. A. Murugan, J. Kongsted, K. Aidas, A. H. Steindal and H. Ågren, *J. Phys. Chem. B*, 2011, **115**, 4350–4358.

63 M. S. Norby, O. Vahtras, P. Norman and J. Kongsted, *Mol. Phys.*, 2017, **115**, 39–47.

64 J. F. Ward and C. K. Miller, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1979, **19**, 826.

65 B. F. Levine, *J. Chem. Phys.*, 1976, **65**, 2429.

66 K. V. Mikkelsen, Y. Luo, H. Ågren and P. Jørgensen, *J. Chem. Phys.*, 1995, **102**, 9362.

67 K. O. Sylvester-Hvid, K. V. Mikkelsen, P. Norman, D. Jonsson and H. Ågren, *J. Phys. Chem. A*, 2004, **108**, 8961–8965.

68 A. H. Narten, *J. Chem. Phys.*, 1971, **55**, 2263.

69 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *J. Comput. Chem.*, 2010, **31**, 671–690.

70 D. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossváry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. Kollman, *Amber 12*, University of California, San Fransisco, 2012.

71 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

72 J.-P. Piquemal, L. Perera, G. A. Cisneros, P. Ren, L. G. Pedersen and T. A. Darden, *J. Chem. Phys.*, 2006, **125**, 054511.

73 H. Casimir and D. Polder, *Phys. Rev.*, 1948, **73**, 360.

74 E. Power and T. Thirunamachandran, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1993, **48**, 4761.

75 A. J. Stone and A. J. Misquitta, *Int. Rev. Phys. Chem.*, 2007, **26**, 193–222.

This journal is © the Owner Societies 2019

Phys. Chem. Chem. Phys., 2019, **21**, 2251–2270 | 2269

76 R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell and M. Head-Gordon, *J. Phys. Chem. A*, 2007, **111**, 8753–8765.

77 R. J. Azar and M. Head-Gordon, *J. Chem. Phys.*, 2012, **136**, 024103.

78 A. J. Misquitta, R. Podeszwa, B. Jeziorski and K. Szalewicz, *J. Chem. Phys.*, 2005, **123**, 214103.

79 A. J. Misquitta and A. J. Stone, *Mol. Phys.*, 2008, **106**, 1631–1643.

80 P. Norman, D. M. Bishop, H. J. Aa. Jensen and J. Oddershede, *J. Chem. Phys.*, 2001, **115**, 10323–10334.

81 P. Norman, D. M. Bishop, H. J. Aa. Jensen and J. Oddershede, *J. Chem. Phys.*, 2005, **123**, 194103.

82 A. Jiemchooroj, P. Norman and B. E. Sernelius, *J. Chem. Phys.*, 2005, **123**, 124312.

83 A. Jiemchooroj, P. Norman and B. E. Sernelius, *J. Chem. Phys.*, 2006, **125**, 124306.

84 R. Podeszwa, R. Bukowski and K. Szalewicz, *J. Phys. Chem. A*, 2006, **110**, 10345–10354.

85 G. Maroulis, *J. Chem. Phys.*, 1998, **108**, 5432–5448.

86 B. Champagne, E. Botek, M. Nakano, T. Nitta and K. Yamaguchi, *J. Chem. Phys.*, 2005, **122**, 114315.

87 C. E. Dykstra, *J. Comput. Chem.*, 1988, **9**, 476–487.

88 R. Munn, *Int. J. Quantum Chem.*, 1992, **43**, 159–169.

89 H. Reis, M. Papadopoulos, P. Calaminici, K. Jug and A. Koster, *Chem. Phys.*, 2000, **261**, 359–371.

90 R. F. W. Bader and T. T. Nguyendang, *Adv. Quantum Chem.*, 1981, **14**, 63–124.

91 T. Seidler and B. Champagne, *Phys. Chem. Chem. Phys.*, 2015, **17**, 19546–19556.

92 S. Roth and I. Freund, *J. Chem. Phys.*, 1979, **70**, 1637–1643.

93 I. Freund, M. Deutsch and A. Sprecher, *Biophys. J.*, 1986, **50**, 693–712.

94 E. Brown, T. McKee, A. Pluen, B. Seed, Y. Boucher and R. K. Jain, *et al.*, *Nat. Med.*, 2003, **9**, 796–800.

95 G. Cox, E. Kable, A. Jones, I. Fraser, F. Manconi and M. D. Gorrell, *J. Struct. Biol.*, 2003, **141**, 53–62.

96 A. C. Ko, A. Ridsdale, M. S. Smith, L. B. Mostaco-Guidolin, M. D. Hewko, A. F. Pegoraro, E. K. Kohlenberg, B. Schattka, M. Shiomi and A. Stolow, *et al.*, *J. Biomed. Opt.*, 2010, **15**, 020501.

97 M. Swart and P. T. van Duijnen, *Mol. Simul.*, 2006, **32**, 471–484.

98 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599.

99 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617–627.

100 A. Deniset-Besseau, J. Duboisset, E. Benichou, F. Hache, P.-F. Brevet and M.-C. Schanne-Klein, *J. Phys. Chem. B*, 2009, **113**, 13437–13445.

101 C. Loison and D. Simon, *J. Phys. Chem. A*, 2010, **114**, 7769–7779.

102 A. E. Tuer, S. Krouglov, N. Prent, R. Cisek, D. Sandkuijl, K. Yasufuku, B. C. Wilson and V. Barzda, *J. Phys. Chem. B*, 2011, **115**, 12759–12769.

103 M. de Wergifosse, J. de Ruyck and B. Champagne, *J. Phys. Chem. C*, 2014, **118**, 8595–8602.

104 A. Dai, *Wiley Interdiscip. Rev.: Clim. Change*, 2011, **2**, 45–65.

105 V. F. McNeill, N. Sareen and A. N. Schwier, *Atmospheric and Aerosol Chemistry*, Springer, 2014, pp. 201–259.

106 J. Elm, P. Norman, M. Bilde and K. V. Mikkelsen, *Phys. Chem. Chem. Phys.*, 2014, **16**, 10883.

107 J. Elm, P. Norman and K. V. Mikkelsen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 15701–15709.

108 M. Kulmala, I. Riipinen, M. Sipila, H. E. Manninen, T. Petaja, H. Junninen, M. Dal Maso, G. Mordas, A. Mirme, M. Vana, A. Hirsikko, L. Laakso, R. M. Harrison, I. Hanson, C. Leung, K. E. J. Lehtinen and V.-M. Kerminen, *Science*, 2007, **318**, 89–92.

109 M. Sipila, K. Lehtipalo, M. Kulmala, T. Petaja, H. Junninen, P. P. Aalto, H. E. Manninen, E. M. Kyro, E. Asmi, I. Riipinen, J. Curtius, A. Kuerten, S. Borrmann and C. D. O'Dowd, *Atmos. Chem. Phys.*, 2008, **8**, 4049–4060.

110 K. Lehtipalo, M. Sipila, I. Riipinen, T. Nieminen and M. Kulmala, *Atmos. Chem. Phys.*, 2009, **9**, 4177–4184.

111 H. E. Manninen, T. Petaja, E. Asmi, I. Riipinen, T. Nieminen, J. Mikkila, U. Horrak, A. Mirme, S. Mirme, L. Laakso, V.-M. Kerminen and M. Kulmala, *Boreal Environ. Res.*, 2009, **14**, 591–605.

112 X. Li, T. Hede, Y. Tu, C. Leck and H. Ågren, *J. Phys. Chem. Lett.*, 2010, **1**, 769–773.