PCCP

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 18551–18560

www.rsc.org/pccp

PAPER

Scrutinizing the effects of polarization in QM/MM excited state calculations

Kristian Sneskov,*[a] Tobias Schwabe,[a] Ove Christiansen[a] and Jacob Kongsted[b]

Received 24th June 2011, Accepted 26th August 2011
DOI: 10.1039/c1cp22067e

In this paper we demonstrate the importance of including polarization—especially in a fully self-consistent-field manner—when calculating excited states within linear response QM/MM methods based on correlated electronic structure methods. We perform a systematic investigation of solvent polarization effects by identifying lower order polarization reaction fields as compared to the full treatment. In the process we highlight the different nature of static and dynamic reaction fields and demonstrate—by method of example on both solvated systems as well as on biomolecules—that inclusion of both of these is mandatory for an accurate description of excited states. Ultimately, these findings reflect the importance of the change in the solvent reaction field upon electronic excitations. In light of the recent increasing interest in excited state QM/MM methods incorporating mutual polarization between subsystems as a method for treating large molecular systems, the reported investigation constitutes an important step towards dissecting the accuracy of such calculations.

1 Introduction

Combined quantum and classical mechanical[1,2] (QM/MM) methods represent a well accepted and balanced compromise between accuracy and efficiency that in a comparatively simple manner transfers the predictive power of state-of-the-art QM methods to large molecular systems like *e.g.* solvated molecules or biomolecular systems.[3,4] Such methods are of particular interest in the field of computational electronic spectroscopy[3,5–12] where theory and experiment may form a healthy alliance. However, if such hybrid approaches are to truly bridge the gap between theory and experiment it is essential to control the error inherently associated with the models. Consequently, the treatment of the artificial interface between the two subsystems is of special importance. Especially, we emphasize the reports of polarizable QM/MM methods[2,13–21] widely recognized as the most accurate QM/MM methods. Explicitly, in recent studies it has been shown that a significant increase in accuracy of excited state transition properties may be obtained by incorporating so-called explicit mutual polarization between the QM and MM subsystems.[22,23] However, inclusion of these polarization effects makes the theoretical description of the QM/MM model much more complex and, likewise, the computational effort required to solve the underlying equations increases. Consequently, the treatment of polarization effects is a common place of approximations in QM/MM methods, yet their justifications have not been fully established.[8,13,14,24–26] It is the goal of this paper to scrutinize these approximations. We stress the importance of such an investigation seen in the light of the recent increasing interest in developing and applying polarizable QM/MM methods to treat large molecular systems both in the ground and excited states.[10,27–34]

The idea of ascribing parts of the intermolecular interaction as polarization arises quite naturally from a long-range perturbational analysis.[35,36] Implying a multipole expansion of the interacting charge distributions the leading second order term is described by the dipole–dipole polarizability. Within the QM/MM framework, polarization effects are most often described *via* the resulting induced dipoles which thus reflect the response of the environment to the electronic structure of the QM subsystem. The recently reported polarizable embedding (PE) method, based on either density functional theory[22] (DFT) or coupled cluster[23] (CC) QM methods, incorporates a fully self-consistent treatment of induced dipoles for energies as well as for response property calculations, *e.g.* excitation energies. We note that other methods exist. For instance Geerke *et al.* described the MM induced dipoles using an MM charge-on-spring model,[37] while Illingworth *et al.* introduced induced charges to describe the polarization part of the Hamiltonian.[38] The Symmetry Adapted Cluster/Configuration Interaction (SAC-CI) approach is related to CC in many ways and has been applied in QM/MM schemes using point-charges for the surroundings[39,40] as well as coupled to a polarizable continuum.[41] In passing,

[a] *The Lundbeck Foundation Center for Theoretical Chemistry and Center for Oxygen Microscopy and Imaging, Department of Chemistry, University of Aarhus, Langelandsgade 140, DK-8000 Aarhus C, Denmark. E-mail: sneskov@chem.au.dk; Fax: +45 8619 6199; Tel: +45 8942 3827*
[b] *Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark*

we mention that a few other CC/MM or related approaches have been implemented mainly differing in their coupling to the external medium be it through point-charges,[42] (polarizable) force fields,[43] fragment cluster approaches,[44,45] *etc.*[46] Especially, the functional form of the PE potential resembles that of the Effective Fragment Potential (EFP) method.[16,17,47] However, we emphasize the PE method's capability to naturally incorporate polarization in a fully self-consistent fashion between the QM and MM subsystems. In related work[10,48,49] this is a frequent place for additional approximations. In the PE formulation all polarization contributions may be physically interpreted as stemming from a QM density. Besides giving rise to clear and consistent equations this formulation serves the added benefit of automatically incorporating lower-order models for the description of polarization. In fact, here we demonstrate how a hierarchy of polarization models may be defined and their relative numerical importance easily tested given that the full implementation is available. For reasons of clarity we limit ourselves to consider the CC model when deriving the intermediate polarization models. However, we demonstrate the analysis of the polarization contributions by performing calculations both within the CC as well as in the DFT framework ultimately obtaining similar conclusions. Finally, to keep the message clear we do not compare to experimental studies here. Quantitative agreement would require the inclusion of many different effects and would thus obscure the actual goal of this study: to uncover the importance of a dynamic reaction field when trying to incorporate polarization in a self-consistent fashion. For comparison to experimental solvent studies we refer to recent literature.[25,36,50,51]

The structure of this paper is as follows: in the first section we define the polarization models first in a qualitative fashion highlighting the physical concepts of the polarization contributions followed by a rigorous definition within the linear response CC framework. Then, we demonstrate the models by first elaborating on a previous study by Slipchenko[49] on solvatochromic shifts of electronic excitation energies in formaldehyde and *p*-nitro-aniline (PNA) in aqueous solution. We proceed by extending that study further to other molecules including the related *p*-nitro-phenolate (PNP), formamide, pyridazine, acetone, and acrolein. Furthermore, in the original study on PNA,[49] only small molecular clusters were examined. Given the many-body nature of the polarization effects, we find it natural to increase the size of the systems under study. Indeed, this has already been done before[28] but—perhaps surprisingly—no comments were made on the interesting effects of explicit polarization (as described by a dynamic reaction field) for accurate determination of solvatochromic shifts. Further, we also demonstrate that the conclusions also hold for full protein studies where the consequences of a changing electron density due to the absorption are even more apparent. Finally, we conclude on our findings including also advices and rules of thumb on how and when to accurately incorporate polarization effects into electronic transition properties.

2 Theory

In QM/MM one relies on the use of effective operators to describe the coupling between the QM subsystem and the MM environment. To facilitate the manipulation of the interaction between a given QM (indexed by *i*) and a MM (indexed by *a*) site—separated by a distance *R*—we introduce general *k*th order interaction tensors as[35]

$$T^{ia}_{\gamma_1\cdots\gamma_k} = \nabla_{\gamma_1}\cdots\nabla_{\gamma_k}\frac{1}{R}, \tag{1}$$

where the γ_i represent cartesian directions (*x*, *y*, or *z*). This implies that the zeroth order tensor describes the potential from a point charge, the first order tensor describes the electric field from a point charge as well as the potential from a dipole, and so forth for higher order quantities. In a second quantization formulation of the PE model the effective operator may then be written as[22,23]

$$\begin{aligned}\hat{G} &= -\sum_{s=1}^{S}\sum_{k=0}^{K}\frac{(-1)^k}{k!}\sum_{pq}\langle\phi_p(\mathbf{r}^i)|T_k^{is}|\phi_q(\mathbf{r}^i)\rangle Q_k^s\hat{E}_{pq} \\ &\quad -\sum_{a=1}^{A}\mu^{a,\mathrm{ind}}(\mathbf{D})\sum_{pq}\langle\phi_p(\mathbf{r}^i)|T_1^{ia}|\phi_q(\mathbf{r}^i)\rangle\hat{E}_{pq} \\ &= \hat{G}^{es}+\hat{G}^{\mathrm{pol}}(\mathbf{D})\end{aligned} \tag{2}$$

where Q_k^s refer to a *k*th order multipole moment at site *s* while $\mu^{a,\mathrm{ind}}(\mathbf{D})$ represents an induced dipole at polarizable site *a* stemming from a density **D** as detailed below. The *p* and *q* indices refer to molecular orbitals and the familiar singlet excitation operator $\hat{E}_{pq}$ signifies that $\hat{G}$ is a one-electron operator. In agreement with physical intuition the two contributions to the environment operator stem from electrostatic and polarization interactions, respectively. This is in contrast to most reported QM/MM calculations that account only for the electrostatic interaction treated at the point-charge level. Here, we focus exclusively on polarization effects and truncate the distributed multipole series in eqn (2) at the level of quadrupoles. Based solely on physical arguments it is clear that polarization may be incorporated at different levels of theory. At the lowest level we could simply neglect it altogether relying on a purely electrostatic environment to incorporate all relevant effects of the environment. The next level would for instance encompass the use of induced dipoles held constant throughout the calculation. However, in the context of computational spectroscopy it is important to note that such an approach neglects the effects of the changing electronic structure in the excited state as compared to the ground state. In order to set up a hierarchy of physically motivated polarization models we focus on the electronic density present as an argument in eqn (2). Especially, we may tune the amount of polarization incorporated between the subsystems by explicitly using QM densities—that ultimately gives rise to the field (from the QM system) that the induced dipoles (in the MM system) are subjected to—at different levels of theory. In turn, these induced moments result in a reaction field perturbing the QM density. Further, when modeling transitions we may differ uniquely between the reaction field stemming from a static set of ground state induced dipoles and that which has incorporated the effects of the transition itself. We call these two reaction fields static and dynamic, respectively.

2.1 Introducing the polarization models

Consider Fig. 1 where we—using double headed arrows and darker colors—indicate to which extent polarization is incorporated into the quantities on the left. Focusing on one column at a time we may describe the different models of polarization considered here as

- Model 0: $\mu^{\mathrm{ind}} = \mathbf{0}$

No MM polarization is included and the QM subsystem is only coupled to the environment using electrostatics governed by $\hat{G}^{\mathrm{es}}$. Consequently, nothing changes when traversing down the first column of Fig. 1. Most standard QM/MM applications fall into this category truncating the multipole expansion at the point-charge level.

- Model 1: $\mu^{\mathrm{ind}}(\mathbf{D} = \mathbf{D}^{\mathrm{HF}})$—static HF reaction field

The induced dipoles in the PE calculation are obtained from the reference state PE-HF calculation and kept fixed both for the ground state correlated calculation as well as when describing the electronic excitation. Essentially, this corresponds to using the PE-HF density to create a static reaction field affecting the QM system also in the correlated calculation. As a result the derived environment response operators are trivially zero and the reaction field is thus only of static nature as described by the PE-HF determined induced dipoles. The inclusion of mutual polarization is illustrated in Fig. 1 using double-sided arrows.

- Model 2: $\mu^{\mathrm{ind}}(\mathbf{D} = \mathbf{D}^{\mathrm{CC}})$—static reaction field

All PE terms are retained in the ground state optimization of the wave function and the induced dipoles are determined in a fully self-consistent manner. Thus, the field from the electrons stems from the full CC density but the derived environment response operators are artificially defined to be zero. Physically, this model may be interpreted as not allowing the induced dipoles to respond to the electronic excitation in the QM subsystem. Unlike model 1 the field at the MM sites now reflects the correlated motion of the electrons in the ground state as described by CC but the electronic excitations still do not modify the reaction field.

- Model 3: $\mu^{\mathrm{ind}}(\mathbf{D} = \mathbf{D}^{\mathrm{CC}})$—full reaction field

For the ground state optimization this model is identical to model 2. However, the description of the electronic excitation is different. The electronic transitions now modify the reaction field itself such that this model contains both static and dynamic contributions to the reaction field. Specifically, in a response picture this means that the effect of the transition enters through a derivative density and no reoptimization of the reference state is thus taking place.

Fig. 1 clearly illustrates the hierarchical nature of these polarization models.

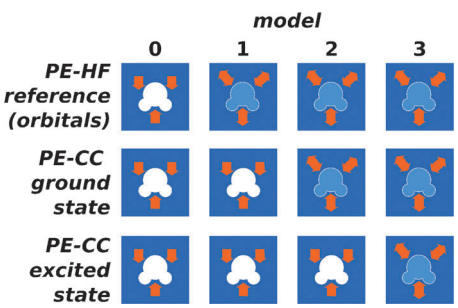

Fig. 1 Illustration of the different polarization models used in this work where double-sided arrows indicate the inclusion of mutual polarization. See the text for details.

2.2 Analysing the polarization models

To put the models on a more sound theoretical footing we briefly reflect on the method of calculating the QM/MM contribution to electronic excitation energies in a linear response formalism. We perform the analysis in the CC framework but all conclusions transfer straightforwardly to the PE-DFT method. In PE-CC theory the effects of an external field manifest themselves by the appearance of derived operators in the final response equations. These operators may be written compactly by defining the second quantized operator for the electric field at a MM site a due to an electron at $\mathbf{r}$ as

$$\hat{\varepsilon}^a = \sum_{pq} \langle \phi_p(\mathbf{r}) | T_1^{ia} | \phi_q(\mathbf{r}) \rangle \hat{E}_{pq}. \tag{3}$$

Recalling eqn (1) T_1 signifies the derivative relationship between electric field and potential. Then, we may write all operators relevant for obtaining the polarization contributions to CC excitation energies as[23]

$$\hat{G}^{\mathrm{pol}} = -\sum_{ab} R^{ab}(F^{b,\mathrm{mul}} + F^{b,\mathrm{nuc}} + \langle \Lambda | \hat{\varepsilon}^b | \mathrm{CC} \rangle) \hat{\varepsilon}^a \tag{4}$$

$$\hat{G}^{\nu_j} = \frac{\partial \hat{G}}{\partial t_{\nu_j}} = -\sum_{ab} R^{ab} \langle \Lambda | [\hat{\varepsilon}^b, \hat{\tau}_{\nu_j}] | \mathrm{CC} \rangle \hat{\varepsilon}^a \tag{5}$$

where $|\mathrm{CC}\rangle$ is the exponentially parameterized CC state while the auxiliary bra state is given as $\langle \Lambda | = (\langle \mathrm{HF}| + \sum_{\mu_i,i} \bar{t}_{\mu_i} \langle \mu_i|) \exp(-\hat{T})$. t and $\bar{t}$ represent the cluster amplitudes and multipliers, respectively. The amplitude- and multiplier-index μ_i denotes the specific excitation involving i electrons. The matrix R denotes the classical polarization response matrix containing the polarizabilities and dipole–dipole interaction tensors. Finally, F^a represents the electric field at site a with contributions due to either MM multipoles (mul) or QM nuclei (nuc). The final term in the parenthesis in eqn (4) is nothing but the electric field calculated as a CC expectation value and its relation to the CC density should be clear.

Concretely, in CC response theory electronic excitation energies are found as eigenvalues of the CC Jacobian—in turn derived from the CC Lagrangian. This idea is straightforwardly carried over into the PE domain (just like it was in its CC/MM ancestor models)[15,52] by defining the PE-CC Lagrangian as the vacuum CC Lagrangian augmented with the electrostatic and polarization interaction terms arising in a QM/MM long-range perturbational analysis. Then, the full Jacobian may be determined by differentiating the Lagrangian with respect to the CC multipliers and amplitudes which in turn may be uniquely partitioned into a vacuum and a QM/MM part

$$\mathbf{A} = \mathbf{A}^{\mathrm{vac}} + \mathbf{A}^{\mathrm{QM/MM}}. \tag{6}$$

The vacuum part does not concern us here. Rather, the elements of the QM/MM Jacobian may—in terms of the one-electron environment operator $\hat{G}$—be written as

$$A^{\mathrm{QM/MM}}_{\mu_i \nu_j} = \langle \bar{\mu}_i | [\hat{G}, \hat{\tau}_{\nu_j}] | \mathrm{CC} \rangle + \langle \bar{\mu}_i | \hat{G}^{\nu_j} | \mathrm{CC} \rangle. \tag{7}$$

with $\langle\bar{\mu}_i| = \langle\mu_i|\exp(-T)$. In line with the polarization models qualitatively defined in the previous section, we may rewrite the environment operator as

$$\hat{G} = \hat{G}^{\mathrm{es}} + \hat{G}^{\mathrm{PE\text{-}HF}}(\mathbf{D}^{\mathrm{HF}}) + \hat{G}^{\Delta}(\mathbf{D}^{\Delta}) \tag{8}$$

where Δ signifies a difference between PE-CC and the reference PE-HF, *i.e.*

$$\hat{G}^{\Delta}(\mathbf{D}^{\Delta}) = \hat{G}^{\mathrm{PE\text{-}CC}}(\mathbf{D}^{\mathrm{PE\text{-}CC}}) - \hat{G}^{\mathrm{PE\text{-}HF}}(\mathbf{D}^{\mathrm{PE\text{-}HF}}). \tag{9}$$

We note that

$$\hat{G}^{\nu_j} = \frac{\partial \hat{G}}{\partial t_{\nu_j}} = \frac{\partial \hat{G}^{\Delta}}{\partial t_{\nu_j}}. \tag{10}$$

Inserting eqn (8) into (7) we obtain

$$\begin{aligned} A^{\mathrm{QM/MM}}_{\mu_i\nu_j} &= \langle\bar{\mu}_i|[\hat{G}^{\mathrm{es}}, \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle + \langle\bar{\mu}_i|[\hat{G}^{\mathrm{PE\text{-}HF}}, \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle \\ &\quad + \langle\bar{\mu}_i|[\hat{G}^{\Delta}, \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle + \langle\bar{\mu}_i|\hat{G}^{\nu_j}|\mathrm{CC}\rangle. \end{aligned} \tag{11}$$

In essence, this is the prerequisite for the polarization models. Especially, by denoting the (biorthonormalized) left and right eigenvectors as L and R, respectively, the partitioning of the Jacobian carries over into the QM/MM contribution to the excitation energy, *i.e.*

$$\begin{aligned} \omega_k^{\mathrm{QM/MM}} = \mathbf{L}_k\mathbf{A}^{\mathrm{QM/MM}}\mathbf{R}_k &= \Delta\omega_k^{\mathrm{es}} && \text{Model 0} \\ &\quad + \Delta\omega_k^{\mathrm{hf}} && \text{Model 1} \\ &\quad + \Delta\omega_k^{\Delta} && \text{Model 2} \\ &\quad + \Delta\omega_k^{\mathrm{dynRF}} && \text{Model 3} \end{aligned}$$

where (dropping the subscript k for readability)

$$\Delta\omega^{\mathrm{es}} = \sum_{\mu_i,i}\sum_{\nu_j,j} L_{\mu_i}\langle\bar{\mu}_i|[\hat{G}^{\mathrm{es}}, \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle R_{\nu_j} \tag{12}$$

$$\Delta\omega^{\mathrm{hf}} = \sum_{\mu_i,i}\sum_{\nu_j,j} L_{\mu_i}\langle\bar{\mu}_i|[\hat{G}^{\mathrm{PE\text{-}HF}}(\mathbf{D}^{\mathrm{PE\text{-}HF}}), \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle R_{\nu_j} \tag{13}$$

$$\Delta\omega^{\Delta} = \sum_{\mu_i,i}\sum_{\nu_j,j} L_{\mu_i}\langle\bar{\mu}_i|[\hat{G}^{\Delta}(\mathbf{D}^{\Delta}), \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle R_{\nu_j} \tag{14}$$

$$\Delta\omega^{\mathrm{dynRF}} = \sum_{\mu_i,i}\sum_{\nu_j,j} L_{\mu_i}\langle\bar{\mu}_i|\hat{G}^{\nu_j}|\mathrm{CC}\rangle R_{\nu_j}. \tag{15}$$

Note the strong similarity between eqn (12)–(14). Suitable effective Jacobians may all be written in the form $\mathbf{L}\mathbf{A}^{\mathrm{eff}}\mathbf{R}$. This partitioning carries straightforwardly over into their physical interpretation as representing the incremental contribution when going from a fixed electrostatic solvent field (model 0) to a fixed HF solvent field (model 1), and further to a fixed CC field (model 2). Eqn (15) is of a different nature though. However, performing the contraction $\hat{R} = \sum_{\nu_j,j}\hat{\tau}_{\nu_j}R_{\nu_j}$ we may recast it as

$$\Delta\omega^{\mathrm{dynRF}} = \sum_{\mu_i,i} L_{\mu_i,i}\langle\bar{\mu}_i|\hat{G}^{\mathrm{dynRF}}(\mathbf{D}')|\mathrm{CC}\rangle \tag{16}$$

where

$$D'_{pq} = \langle\Lambda|[\hat{E}_{pq}, \hat{R}]|\mathrm{CC}\rangle \tag{17}$$

Thus, we see that model 3 reflects that the electronic excitation changes the QM density and thereby the dynamic reaction field. This ultimately results in a correction to the excitation energy itself which in this paper is shown to be non-negligible.

However, a shortcut presents itself if we simply neglect (some of) the terms entirely or if we *e.g.* define eqn (4) independent of the electronic parameters. While potentially computationally more efficient methods may be derived following such a strategy, it is of utmost importance to assess the additional numerical errors introduced as a consequence of such approximations.

Note that in this work our primary focus is on investigating physically consistent models for polarization. However, it is clear that a wealth of intermediate models exist defined by which terms are retained in the response equations (or simply which field (*i.e.* density) that enters in eqn (4) and (16)). Indeed, one might speculate in applying a dynamic HF reaction field within a PE-CC response calculation. Such an approach is not rigorously defined in a CC linear response framework and we thus refrain from it here. However, cost-efficient ways of incorporating polarization may be designed in this manner.

The above analysis was performed for fixed left and right eigenvectors. In practice, neglecting specific terms of the Jacobian will give different eigenvectors. All numbers presented correspond to the appropriately converged excitation energies and eigenvectors. The analysis nevertheless illustrates the physical difference between the different contributions.

A similar analysis of polarization contributions can be performed within the PE-DFT framework. One minor detail is the lack of a model 1 and 2 partitioning since DFT is a one-level calculation as opposed to the two-level CC calculations. We denote the single DFT correspondent model as model 1.

2.3 Extension to higher-order properties

For properties other than the excitation energies we may straightforwardly generalize the introduced concepts. Up to the level of quadratic response theory the only computational difference will be the emergence of higher-order differentiated environment operators such as[23]

$$^{\nu_j}\hat{G} = \frac{\partial \hat{G}}{\partial \bar{t}_{\nu_j}} = -\sum_{ab} R^{ab}\langle\nu_j|\hat{\varepsilon}^b|\mathrm{CC}\rangle\hat{\varepsilon}^a \tag{18}$$

$$\hat{G}^{\mu_i\nu_j} = \frac{\partial^2 \hat{G}}{\partial t_{\mu_i}\partial t_{\nu_j}} = -\sum_{ab} R^{ab}\langle\Lambda|[[\hat{\varepsilon}^b, \hat{\tau}_{\mu_i}], \hat{\tau}_{\nu_j}]|\mathrm{CC}\rangle\hat{\varepsilon}^a. \tag{19}$$

The strategy for their evaluation follows the same logic as in the preceding section: in the actual response vector transformation we identify suitable intermediate densities that will give rise to the important dynamic reaction field contribution to the given property. In the current PE implementation we may in the DFT framework go a step further and even obtain properties from the cubic response function with the dynamic reaction field contribution incorporated.[22]

3 Computational details

The molecular systems examined in the following consists of formaldehyde and *p*-nitro-aniline (PNA)–water clusters from a recent study,[49] as well as PNA, *p*-nitro-phenolate (PNP), acrolein, formamide, pyridazine, and acetone all in aqueous solution and derived from MD simulations. Finally, the PYP model studied is that derived from the experimental structure[53] suitably modified for the PE calculation.[25]

The smaller molecular PNA and formaldehyde clusters considered in this work were the same as those used by Slipchenko.[49] All other structures were obtained from MD simulations applying the NVT ensemble with $T = 298$ K. These were obtained similar to related studies[51] where we first perform an equilibration run of 0.4 ns followed by a production run of 1.2 ns. In order to obtain statistically uncorrelated configurations we dump configurations every 10 ps. For the PNP simulations a sodium cation was considered as a counter-ion. The cation does not introduce any significant effect. In the subsequent QM/MM calculations we describe PNA and formaldehyde—for ease of comparison—using the same 6-31+G* basis as in the original study. PNP, being an anion, was described using the 6-31++G** basis, while acrolein, formamide, pyridazine, and acetone were described using basis sets previously shown to give good results.[50,51,54] The PYP chromophore was—similar to our original study—also described at the 6-31+G* level. All CC calculations are performed at the CCSD level using a frozen core for the heavy atoms. In the DFT calculations the CAM-B3LYP functional was used.

In order to dissect the environment contributions to the shift in a transparent manner, we do not consider the change in geometry when passing from vacuum to solution. Thus, unless otherwise noted the QM molecular structure is considered to be the same in vacuum and in the QM/MM computations. Finally, for the MD simulations we applied the MOLSIM[55] program, while a locally modified version of DALTON[56] has been used for the actual QM/MM calculations. The QM/MM input was prepared using the WHIRLPOOL program.[57] Finally, due to the recent reports of the full PE-DFT[22] and PE-CC[23] implementations the lower-order polarization models were easily implemented by skipping the relevant terms.

4 Results

As defined in Table 1, ΔE^i represents the specific contribution due to model i such that the total shift in excitation energy due to the environment is given as $\Delta E^{\mathrm{tot}} = \Delta E^0 + \Delta E^1 + \Delta E^2 + \Delta E^3$ in the PE-CC case and as $\Delta E^{\mathrm{tot}} = \Delta E^0 + \Delta E^1 + \Delta E^3$ in the PE-DFT case. In Table 2 we show the results of applying the PE-CC method to investigate the formaldehyde and PNA water systems as previously considered by Slipchenko.[49] Not surprisingly, we reproduce the trends observed previously namely that the additional effects due to model 3 are negligible compared to both models 0 and 1. Also, as noted by Slipchenko, the effects of including a dynamic reaction field are typically not exceeding 0.001 eV. To investigate if these conclusions may be regarded as general we report in Fig. 2 a similar analysis of a PNA–water molecular system where we gradually increase the number of surrounding water molecules by artificially making a cut-off sphere with the origin at the PNA center of mass. By this we clearly observe a relative increase in the polarization effects obtained with model 3 as compared to model 1; indeed when incorporating all water molecules within 12 Å the size of the two contributions is the same. This is in contrast to the smaller cluster systems where the many-body effects obviously are smaller.

The configurations examined so far are all single configurations arbitrarily chosen from the MD simulations. For a more rigorous analysis we now include sampling of the phase space and thus account for the motion of the nuclei. Also, we

Table 1 Overview of the different polarization models. In the case of PE-DFT there is no model 2 and the only relevant density is the DFT density. The concept of static and dynamic reaction fields remains equally valid. See the text for details

Polarization model	Polarization density	Polarization reaction field	Shift
0	$\mathbf{0}$	—	$\Delta E^0 = E^0 - E^{\mathrm{vac}}$
1	$\mathbf{D}^{\mathrm{HF}}$	Static	$\Delta E^1 = E^1 - E^0$
2	$\mathbf{D}^{\mathrm{CC}}$	Static	$\Delta E^2 = E^2 - E^1$
3	$\mathbf{D}^{\mathrm{CC}}$	Static + dynamic	$\Delta E^3 = E^3 - E^2$

Table 2 Analysis of environmental effects on the excitation energy in the formaldehyde–water and PNA–water clusters from ref. 49. For ease of comparison we also include the results from the previous analysis. Note that the sum of models 2 and 3 is comparable to the "direct" polarization effects reported in ref. 49. All numbers are in eV

		This study			Ref. 49		
		Cluster size			Cluster size		
	Polarization model	2	4	6	2	4	6
Formaldehyde	0	+0.193	+0.131	+0.269	+0.173	+0.125	+0.262
	1	+0.046	+0.040	+0.051	+0.034	+0.032	+0.047
	2	−0.004	−0.003	−0.005	(−0.009)	(−0.008)	(−0.015)
	3	−0.001	−0.001	−0.003			
PNA	0	−0.233	−0.134	−0.165	−0.213	−0.145	−0.177
	1	−0.038	−0.035	−0.052	−0.049	−0.040	−0.043
	2	+0.003	+0.003	+0.005	(−0.008)	(−0.009)	(−0.019)
	3	−0.009	−0.009	−0.020			

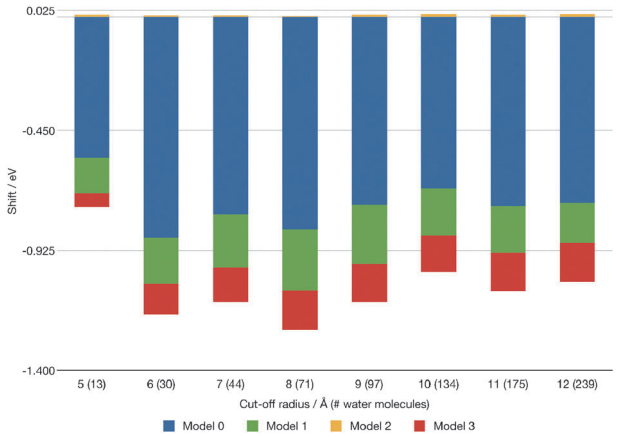

Fig. 2 Effect of including an increasing number of solvent molecules based on a cut-off radius. Results are based on one (arbitrarily chosen) MD configuration. Note that the increment due to model 2 is in the opposite direction and very small.

explicitly show that the picture is—not surprisingly—exactly the same within the DFT framework. The same partitioning of the solvatochromic shift as shown above is demonstrated in Fig. 3 (PE-CC) and Fig. 4 (PE-DFT) for 100 uncorrelated configurations incorporating all water molecules within 12 Å of the PNA molecule. As expected, the total shift is dominated by the electrostatic contribution, but—more importantly in this context—the contributions from model 1 and 3 are of similar magnitude and it is thus important to include both. In the PE-CC case, the additional accuracy obtained by applying model 2 as compared to model 1 is negligible and scarcely visible in the figure. This means that the effect of electron correlation to the static reaction field is negligible for the excitation energy—however, the dynamical component of the reaction field generally provides a sizable contribution to the solvatochromic shift and should be included.

To demonstrate that these conclusions do not apply uniquely to PNA we also consider solvated PNP. In Fig. 5 and 6 we provide a similar dissection of the solvatochromic shift as considered for PNA in Fig. 3 and 4. For this electronic transition the effects of model 3 are even more pronounced dominating the total shift completely and to neglect these is generally not advisable. Indeed, for some configurations the use of model 3 even results in an opposing effect to the shift. Finally, in Fig. 7 we report an analysis using PE-CC for the low-lying transitions for different molecular systems all based on structures from molecular dynamics simulations. We focus on the effects making up the full shift (for clarity skipping the negligible model 2 partitioning) such that 100% comprises the total solvatochromic shift. The nature of the $\pi \rightarrow \pi^*$ transition makes it much more demanding to describe compared to the $n \rightarrow \pi^*$ transition. Consequently, as expected, the model 3 contribution is negligible for the $n \rightarrow \pi^*$ transition but this is indeed not so for the $\pi \rightarrow \pi^*$ transition.

Finally, to demonstrate the flexibility of the approach we consider now the photoactive yellow protein. Recently,[25] we presented an analysis of the lowest intense excitation energy responsible for the unique properties of this protein. Essentially, we explained the experimental findings[58] of a very similar excitation energy of the chromophore under vacuum and within the protein as stemming from two oppositely directed effects. Here, we may take this analysis one step further by dissecting the polarization contributions using the models defined above. The results are shown in Fig. 8 where one bar shows the

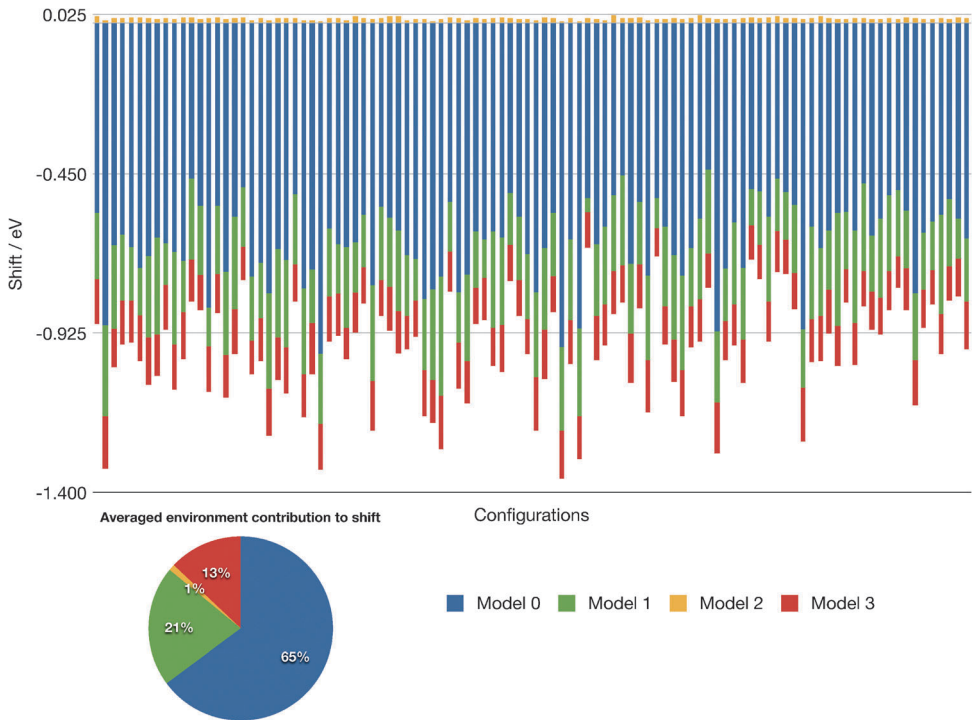

Fig. 3 Analysis of solvatochromic shifts calculated using PE-CC for several PNA–water configurations obtained from a MD simulation. The averaged contributions are shown in the pie chart.

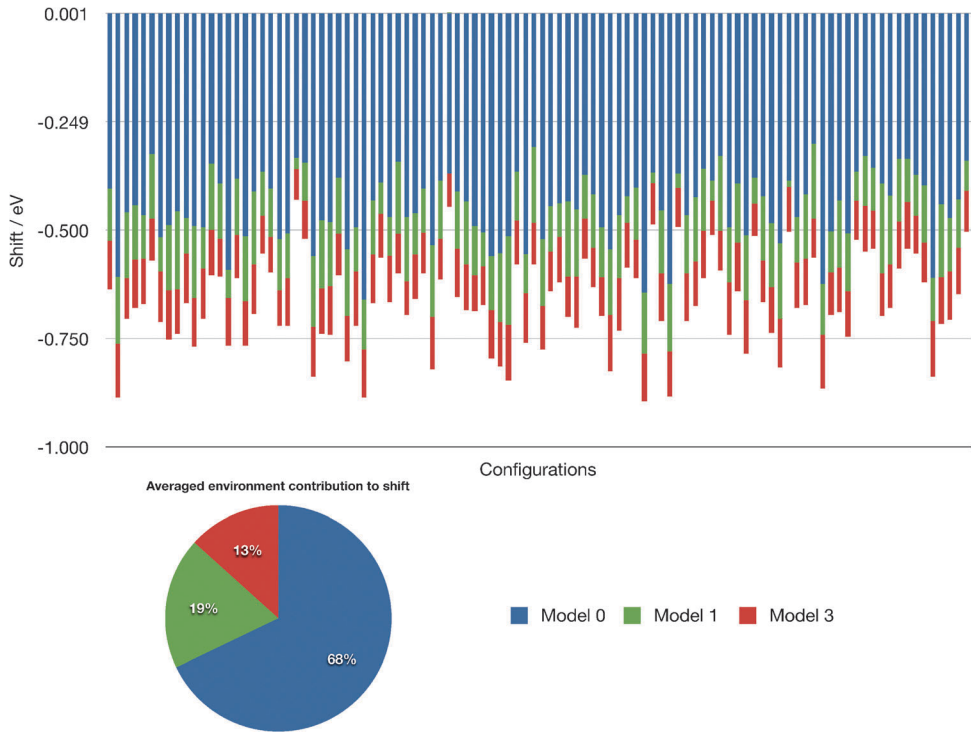

Fig. 4 Analysis of solvatochromic shifts calculated using PE-DFT for several (100) PNA–water configurations obtained from a MD simulation. The averaged contributions are shown in the pie chart.

incremental shifts due to the listed environmental effects while the other indicates the shift relative to the vacuum excitation energy of 3.33 eV. Notice the very small total residual shift when going from the vacuum to the full protein described by a polarizable force field. However, much more important in this context is the effect of using a dynamic reaction field *versus* a static. The largest incremental shift is that of incorporating the dynamic reaction field effects as it serves to cancel the effects of using only a ground state static polarization environment. In fact, in this particular case the pure electrostatic force field is in

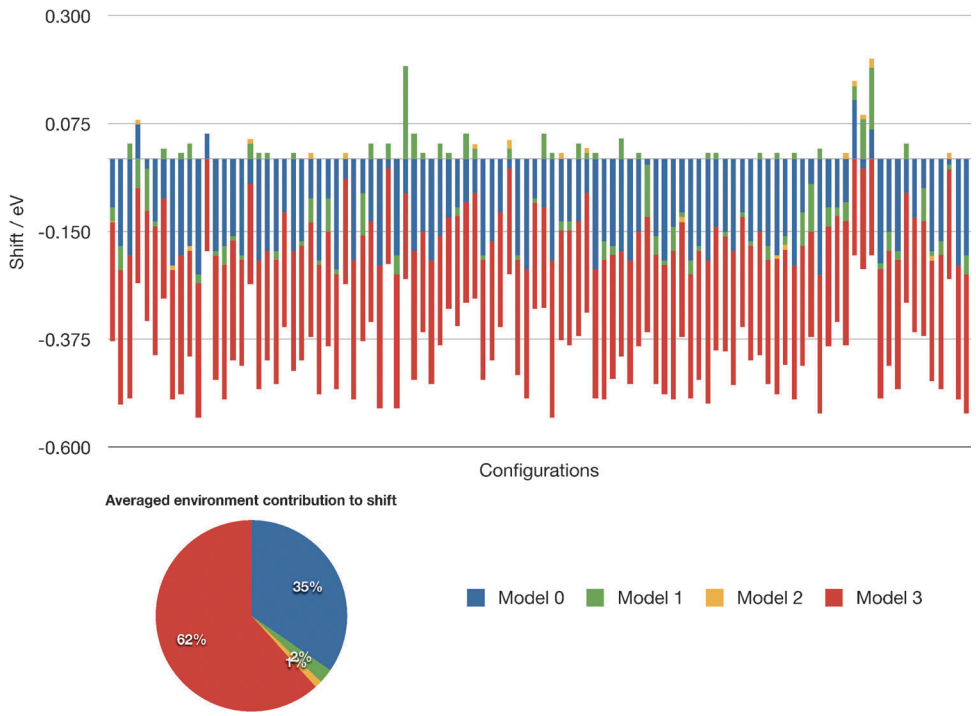

Fig. 5 Analysis of solvatochromic shifts for several PNP–water configurations obtained from a MD simulation. The averaged contributions are shown in the pie chart.

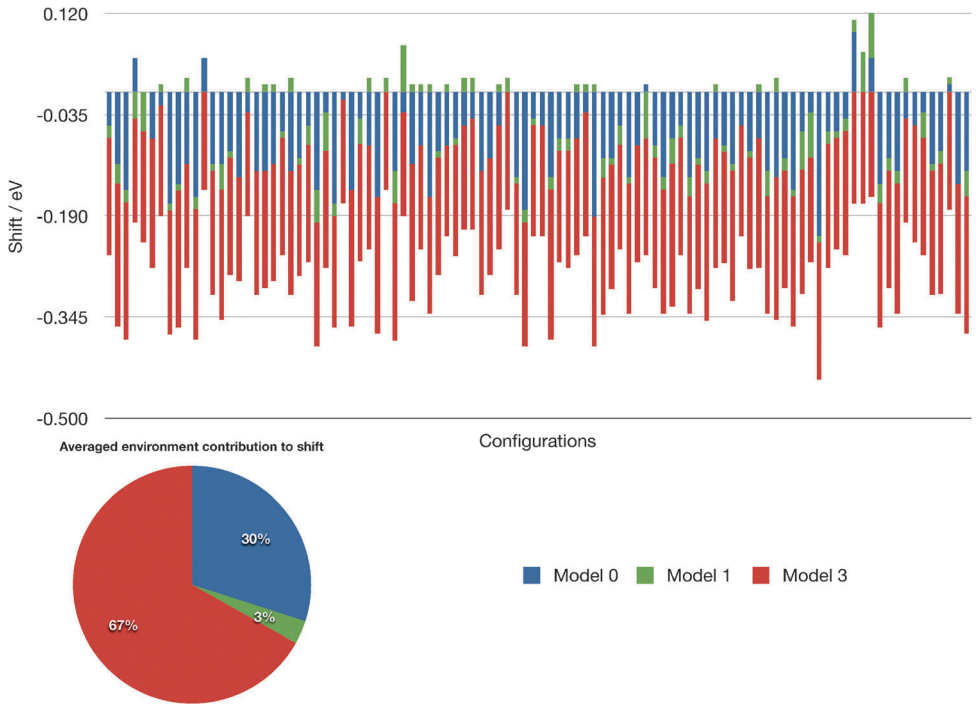

Fig. 6 Analysis of solvatochromic shifts calculated using PE-DFT for several (100) PNP–water configurations obtained from a MD simulation. The averaged contributions are shown in the pie chart.

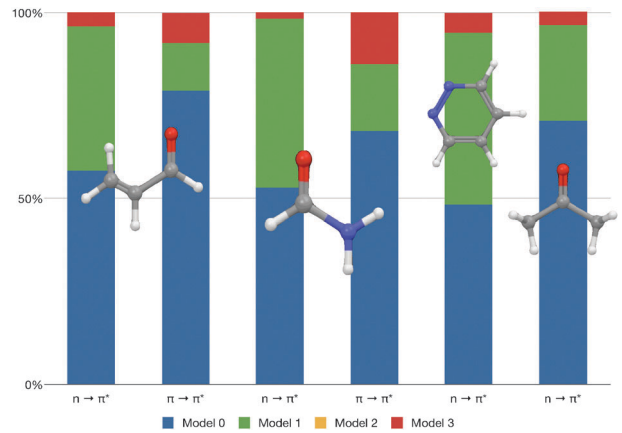

Fig. 7 Relative partitioning of different contributions to the total solvatochromic shift for acrolein, formamide, pyridazine, and acetone in aqueous solution. Results are based on one (arbitrarily chosen) MD configuration.

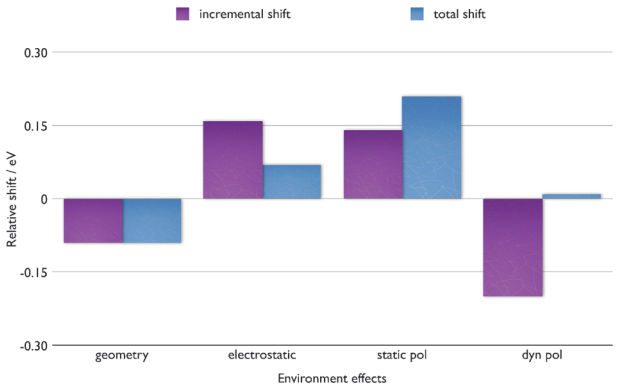

Fig. 8 Dissection of the environment contributions to the lowest intense excitation energy of PYP using PE-DFT with the CAM-B3LYP functional.

better agreement with experiment than when including only static polarization effects. This might lead to erroneous conclusions about the necessity of including polarization in the first place when the opposite is in fact true: it is only when incorporating the full effects of polarization that the physics in this system is unravelled. The total shift is constituted of a delicate balance between many different effects.

5 Conclusions

In conclusion we find that in general the model 3—representing the polarization effects of a changing electronic density upon absorption—contribution to solvatochromic shifts should be included in polarizable QM/MM approaches if high accuracy is strived for. We demonstrate that this conclusion is not restricted to solvation studies but remains equally valid when describing biomolecular systems.[8,25] Indeed, when considering a biomolecule with a very anisotropic environment it seems even more important to incorporate the dynamic reaction field contribution. Note that it is exactly when large systems are considered that the special effects of a dynamical reaction field are important in agreement with the fact that polarization is a many-body effect.

The physical interpretation behind the importance of the dynamic reaction field contribution is clear: when the electronic structure of the excited state is significantly different from that of the ground state a large effect is expected. This is often the case when treating transitions with charge-transfer

character or when dealing with intense excitations such as a typical $\pi \rightarrow \pi^*$ transition. Indeed, recently it was reported[59] that in the extreme case of ionization it may amount to as much as 0.7 eV. However this study shows that it is not only in such special cases the contribution is relatively large. Finally, we stress that the polarization model 3 is the natural implementation of polarization effects within a linear response picture in the PE method.

To what extent properties other than excitation energies are affected is at present unknown. Still, a wide area of research topics are subjects of QM/MM studies nowadays: photoactive proteins, light harvesting, organic solar cells, fluorescence in solution, proton transfer in the self-healing process of the DNA, and so on. Especially for the dynamic simulations of excited states, an accurate description of polarization effects is most likely crucial and should not be neglected. Arguably, our study also indicates that in some cases the effect of model 3 polarization is minor. However, such a conclusion is not necessarily easily drawn from the outset. A first indication that one should be wary might be the relative change in the associated dipole moment. This may straightforwardly be checked on the isolated QM system and could thus constitute an easy first test. However, the environment most likely influences this, too. And when incorporating effects of dynamics this is even a less reliable indicator due to the coupling of the dipole moment to structural changes.

Similar polarizable QM/MM schemes—be they based on a CC, DFT, or another QM method—must include a correction to the static determined HF induced dipoles if they are to be termed accurate. We emphasize that the PE method retains all polarization contributions (both static and dynamic) also to higher-order properties. At present this encompasses quadratic response for the PE-CC method and cubic response for PE-DFT.

Acknowledgements

This work has been supported by the Lundbeck Foundation and DCSC (Danish Center for Scientific Computing). O.C. acknowledges support from the Danish national research foundation, the Lundbeck Foundation and EUROHORCs through a EURYI award. T.S. thanks the Humboldt Foundation for a Feodor-Lynen fellowship. J.K. thanks The Danish Councils for Independent Research, The Villum Foundation and the Lundbeck Foundation for financial support.

References

1 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
2 A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612–5625.
3 H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
4 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185.
5 J. Neugebauer, *ChemPhysChem*, 2009, **10**, 3148.
6 A. Pedone, M. Biczysko and V. Barone, *ChemPhysChem*, 2010, **11**, 1812.
7 A. S. T. Ribeiro and R. B. de Alencastro, *Int. J. Quantum Chem.*, 2011, **111**, 1252–1255.
8 P. Söderhjelm, C. Husberg, A. Strambi, M. Olivucci and U. Ryde, *J. Chem. Theor. Comput.*, 2009, **5**, 649–658.
9 J. W. Kaminski, S. Gusarov, T. A. Wesolowski and A. Kovalenko, *J. Phys. Chem. A*, 2010, **114**, 6082–6096.
10 P. Arora, L. V. Slipchenko, S. P. Webb, A. DeFusco and M. S. Gordon, *J. Phys. Chem. A*, 2010, **114**, 6742–6750.
11 J. Neugebauer, C. Curutchet, A. Muñoz-Losa and B. Mennucci, *J. Chem. Theor. Comput.*, 2010, **6**, 1843.
12 C. König and J. Neugebauer, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10475–10490.
13 M. A. Thompson and G. K. Schenter, *J. Phys. Chem.*, 1995, **99**, 6374–6386.
14 Y. Lin and J. Gao, *J. Chem. Theor. Comput.*, 2007, **3**, 1484.
15 J. Kongsted, A. Osted, K. V. Mikkelsen and O. Christiansen, *J. Phys. Chem. A*, 2003, **107**, 2578.
16 M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys and W. J. Stevens, *J. Phys. Chem. A*, 2001, **105**, 293.
17 M. S. Gordon, L. V. Slipchenko, H. Li and J. H. Jensen, *Annu. Rep. Comput. Chem.*, 2007, **3**, 177.
18 P. Huang and E. Carter, *Annu. Rev. Phys. Chem.*, 2008, **59**, 261.
19 L. Seijo and Z. Barandiaran, *Computational Chemistry: Review of Current Trends*, World Scientific, Singapore, 1999, vol. 4, pp. 55–152.
20 N. Govind, P. V. Sushko, W. P. Hess, M. Valiev and K. Kowalski, *Chem. Phys. Lett.*, 2009, **470**, 353.
21 J. Gao, *J. Comput. Chem.*, 1997, **18**, 1061–1071.
22 J. M. Olsen, K. Aidas and J. Kongsted, *J. Chem. Theor. Comput.*, 2010, **6**, 3721–3734.
23 K. Sneskov, T. Schwabe, J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2011, **134**, 104108.
24 J. Gao and K. Byun, *Theor. Chem. Acc.*, 1997, **96**, 151–156.
25 T. Rocha-Rinza, K. Sneskov, O. Christiansen, U. Ryde and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2011, **13**, 1585–1589.
26 A. Öhrn and G. Karlström, *Theor. Chem. Acc.*, 2007, **117**, 441–449.
27 O. Acevedo and W. L. Jorgensen, *J. Phys. Chem. B*, 2010, **114**, 8425–8430.
28 D. Kosenkov and L. Slipchenko, *J. Phys. Chem. A*, 2011, **115**, 392–401.
29 N. Minezawa, N. D. Silva, F. Zahariev and M. S. Gordon, *J. Chem. Phys.*, 2011, **134**, 54111.
30 P. Söderhjelm and U. Ryde, *J. Phys. Chem. A*, 2009, **113**, 617.
31 J. Kongsted and B. Mennucci, *J. Phys. Chem. A*, 2007, **111**, 9890.
32 G. Fradelos and T. A. Wesolowski, *J. Chem. Theor. Comput.*, 2011, **7**, 213–222.
33 W. Xie and J. Gao, *J. Chem. Theor. Comput.*, 2007, **3**, 1890–1900.
34 W. Xie, M. Orozco, D. Truhlar and J. Gao, *J. Chem. Theor. Comput.*, 2009, **5**, 459–467.
35 A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, 1st edn, 1997.
36 J. M. Olsen and J. Kongsted, *Adv. Quantum Chem.*, 2011, **61**, 107.
37 D. P. Geerke, S. Thiel, W. Thiel and W. F. van Gunsteren, *J. Chem. Theor. Comput.*, 2007, **3**, 1499.
38 C. J. R. Illingworth, S. R. Gooding, P. J. Winn, G. A. Jones, G. G. Ferenczy and C. A. Reynolds, *J. Phys. Chem. A*, 2006, **110**, 6487.
39 J. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara and H. Nakatsuji, *J. Comput. Chem.*, 2007, **28**, 2443.
40 K. Fujimoto, S. Hayashi, J. Hasegawa and H. Nakatsuji, *J. Chem. Theor. Comput.*, 2007, **3**, 605.
41 R. Cammi, R. Fukuda, M. Ehara and H. Nakatsuji, *J. Chem. Phys.*, 2010, **133**, 24104.
42 M. Valiev and K. Kowalski, *J. Chem. Phys.*, 2006, **125**, 211101.
43 R. Cammi, *J. Chem. Phys.*, 2009, **131**, 164104.
44 S. Hirata, M. Valiev, M. Dupuis, S. Xantheas, S. Sugiki and H. Sekino, *Mol. Phys.*, 2005, **103**, 2255.
45 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2006, **124**, 79904.
46 R. A. Mata, *Mol. Phys.*, 2010, **108**, 381.
47 P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch and D. Cohen, *J. Chem. Phys.*, 1996, **105**, 1968.
48 S. Yoo, F. Zahariev, S. Sok and M. S. Gordon, *J. Chem. Phys.*, 2008, **129**, 144112.
49 L. Slipchenko, *J. Phys. Chem. A*, 2010, **114**, 8824–8830.
50 K. Sneskov, E. Matito, J. Kongsted and O. Christiansen, *J. Chem. Theor. Comput.*, 2010, **6**, 839–850.

51 T. Schwabe, J. M. Olsen, K. Sneskov, J. Kongsted and O. Christiansen, *J. Chem. Theor. Comput.*, 2011, **7**, 2209–2217.
52 J. Kongsted, A. Osted, K. V. Mikkelsen and O. Christiansen, *J. Chem. Phys.*, 2003, **118**, 1620.
53 U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Canestrelli and E. D. Getzoff, *Nature*, 1998, **392**, 206.
54 K. Aidas, A. Møgenhøj, E. J. Nilsson, M. S. Johnson, K. V. Mikkelsen, O. Christiansen, P. Söderhjelm and J. Kongsted, *J. Chem. Phys.*, 2008, **128**, 194503.
55 P. Linse, *MOLSIM 3.1*, Lund University, Sweden, 2000.
56 DALTON, a molecular electronic structure program, Release 2.0, 2005, see http://www.kjemi.uio.no/software/dalton/dalton.html.
57 K. Aidas, *Whirlpool—a QM/MM analysis program*, 2010.
58 I. B. Nielsen, S. Boye-Peronne, M. O. A. E. Ghazaly, M. Kristensen, S. B. Nielsen and L. H. Andersen, *Biophys. J.*, 2005, **89**, 2597.
59 D. Ghosh, O. Isayev, L. Slipchenko and A. Krylov, *J. Phys. Chem. A*, 2011, **115**, 6028–6038.

