Ionization energies in solution with the QM : QM approach

Tóth, Zsuzsanna

2020-05-21

Tóth , Z , Kubecka , J , Muchová , E & Slavicek , P 2020 , ' Ionization energies in solution with the QM : QM approach ' , Physical Chemistry Chemical Physics , vol. 22 , no. 19 , pp. 10550-10560 . https://doi.org/10.1039/c9cp06154a

http://hdl.handle.net/10138/330152
https://doi.org/10.1039/c9cp06154a

acceptedVersion

Ionization Energies in Solutions with QM:QM approach

Zsuzsanna Tóth, Jakub Kubečka[1], Eva Muchová, Petr Slavíček*

[1]*University of Chemistry and Technology Prague, Department of Physical Chemistry, Technická 5, 16628 Prague 6, Czech Republic*

Abstract

We discuss a fragment-based QM:QM scheme as a practical way how to access the energetics of vertical electronic processes in the condensed phase. In the QM:QM scheme, we decompose the large molecular system into small fragments, which interact solely electrostatically. The energies of fragments are calculated in a self-consistent field generated by other fragments and the total energy of the system is calculated as a sum of fragment energies. We show on two test cases (cytosine and a sodium cation) that the method allows to accurately simulate the shift of vertical ionization energies (VIE) while going from the gas phase to the bulk. For both examples, the predicted solvent shifts and peak widths estimated at the DFT level agree well with the experimental observations. We argue that the QM:QM approach is more suitable than either an electrostatic embedding based QM/MM approach, a full quantum description at the DFT level with generally used functional or a combination of both. We also discuss the potential scope of the applicability for other electronic processes such as the Auger decay.

1 Introduction

Quantum chemical simulations of the processes in the condensed phase are tricky and demanding. Solvent effects can be significant, such as in the case of acid-base reactions, redox reactions[1,2] or generally, charge transfer processes[3,4] and their deficient description adversely affects the results and interpretation.

Reliable theoretical strategies for processes in solutions have to faithfully reproduce polarization of the system. Both the solute and solvent adjust their positions and electronic structure in response to its environment (in fact, the distinction between solute and solvent *per se* is rather artificial). The molecules react in two ways: the molecular geometry and orientation changes in response to the environment (nuclear polarization) and the electron cloud distorts

[1] Currently at Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki FI-00014, Finland

(electron polarization). Electron and nuclear polarization have characteristic response times which have to be accounted for (non-equilibrium vs equilibrium solvation schemes[5]).

The theoretical accuracy is particularly important for vertical electronic processes. In this case, the charge distribution changes are typically large and abrupt, *e.g.* during charge transfer excitation or ionization,[6,7] and only the electronic degrees of freedom of the solvent can react. Here, missing or inaccurate electronic polarization may lead to quantitatively incorrect energetics.[8] This is particularly true for the liquid phase ionization: a charge state of an atom or a molecule suddenly changes and any non-polarizable solvent models cannot adequately capture the physics. This has lead *e.g.* to a spectacular overshooting of the calculated ionization energies of nucleic acid basis in its natural environment.[6] The problem was not immediately apparent, as there were virtually no data on photoemission from polar liquids. Fortunately, these data are nowadays accessible by the liquid microjet technique[9,10] and a direct comparison can be made.

Various techniques have been so far proposed to reduce the costs of the simulations while keeping the accuracy. In some cases, QM/MM techniques can describe the solvent effects even within a simple electrostatic embedding scheme.[11] However, the electronic polarization is in this approach completely missing because the electrostatic embedding QM/MM (EE-QM/MM) scheme employs only fixed charges. In many frequently used force fields, the electronic polarization is included implicitly by using scaled charges and [12–15] the error of, *e.g.* the thermodynamic properties can be in this way compensated. Yet for vertical ionization the use of EE-QM/MM is problematic (as we show in the text).

The missing electronic polarization can be included in several ways. The simplest approach is to apply polarizable dielectric models such as polarizable continuum model in its non-equilibrium formulation.[16–20] The information on the granularity of the solvent environment is, however, lost. Continuum models are designed to provide a configurationally sampled solvent effect, therefore we can only get the mean value of VIE but no information on its distribution. In addition, the classical continuum model cannot include specific solute-solvent interactions (charge-transfer type). This problem can be circumvented with the cluster-continuum model, yet this brings new issues with a proper selection of the cluster and its size.[21,22] An accurate solution can be provided by a polarizable embedding (*e.g.* AMOEBA force field),[23] or by other related techniques such as effective fragment potential (EFP)

method.[23–25] In this case, we most often rely on linear polarizabilities which do not have to be accurate enough when accounting for large electric fields.

Alternatively, we can perform fully quantum simulations. In this case, we typically have to rely on density functional theory (DFT) approaches, which allows for simulations of hundreds of atoms. DFT can be, however, not only expensive but also incorrect, as we will also show in this work. DFT functionals were parametrized for small molecular systems and applications to extended systems can exhibit pathologies.[26–28]

In this article, we advocate for the use of QM:QM fragment techniques. Fragment methods facilitate the calculations on large molecular systems where regular *ab initio* quantum methods are intractable because the system size.[29] In QM:QM, the super system is divided into a large number of small fragments and each fragment is calculated at the same level of *ab initio* theory. In this way, the complexity of the problem is reduced, in an optimal case to linear scaling with the system size.[30] This makes it much easier to perform quantum chemical simulations for even thousands of atoms.[31] Such a family of methods has a long and rather successful history (see partially discussed below); a broad review has been provided *e.g.* by Gordon et. al.[29] or by Herbert.[30]

The apparent easiness of doing fragment-based methods have resulted in a plethora of possible variations, the majority of them, however, is built on several keys ideas. One of them is the many-body expansion which decomposes the system energy into energies of fragments and many-body corrections. These methods based on energy fragmentation[32] are particularly well suited for liquids (a disordered, non-covalently bounded systems containing a large number of units). For example, the generalized energy-based fragment (GEBF) method of Li at al.[33,34] applies the many-body expansion formula to obtain the total energy. Calculations for fragments or fragment clusters are performed either ignoring the other fragments or replacing them by fixed charges.[34] The GEBF method was even applied for geometry optimization and vibrational spectra optimization.[35,36] Energy fragmentation approaches can be, in principle, applied also to macromolecules, in which case covalent bonds are severed and fragments are capped to restore the correct valences. This is the case of the molecular fractionation with conjugate caps (MFCC).[37,38]

Interactions between fragments can be calculated in many ways; *ab initio* calculation can be performed not only for fragment monomers, but also for fragment dimers, trimers and, in general, *n*-mers. Other methods, for example the molecular tailoring approach[39,40] (MTA)

treats the interaction between fragments by defining overlapping fragments. These approaches are particularly suitable for short-range interactions. Another way how to describe interactions between fragments is established on fragment molecular orbital method (FMO) of Kitaura *et al.*[41], which uses fragment and fragment pair Hamiltonians. Later, the FMO method has been generalized to use even trimer and *n*-mer Hamiltonians.[42] The model Hamiltonians include embedding potential derived from the density matrices of all of the fragments, leading to an iterative procedure. In the same direction, Dahlke and Truhlar[43–45] introduced electrostatically embedded many-body expansion method. Similarly to FMO, fragment monomer energies and also fragment cluster energies are calculated. A huge simplification is, that during each *ab initio* calculations the other fragments are replaced by fixed point charges.

Similarly, the effective fragment potential method (EFP)[29,46,47] was originally formulated to describe water clusters. The EFP energy expression contains electrostatic, polarization and exchange repulsion terms, all of their parameters are obtained from *ab initio* fragment calculations. The accurate representation of the electrostatic potential is obtained using multipolar expansion (or distributed multipolar analysis).[48] Recently the EFP method was applied to calculate excitation energies in the liquid phase[49]: combined with time dependent density functional theory (TDDFT),[50] combined with CIS,[51] combined with EOM-CCSD.[52]

In the present work, we studied two systems with the QM:QM and QM/MM schemes – a solvated sodium cation and solvated cytosine. Such a choice was purely pragmatic; there exist theoretical and experimental data, which allow to estimate the solvent shift.[6,7,53] Ionization of simple ions has been studied previously by Winter et al.[54] and ionization of nucleic acids has been studied in the context of oxidative stress, radiation damage or charge transfer based molecular nano-devices.[55–57] At the same time, both systems are a convenient test bed systematic characteristics of the QM:QM approach as they represent two cases with a very different solvent shift.

We compare three type of approaches: full QM treatment of a relatively large solvated system, QM/MM approach within electrostatic embedding and finally the QM:QM fragment-based approach with DFT description of the units. These approaches have been successfully used for a variety of systems; yet, to the best of our knowledge, they have never been tested in a different situation (vertical ionization). We want to mainly focus on the importance of solvent polarization for which QM:QM methods are ideally suited. We discuss charge localization and

how to effectively avoid charge "leak" which is otherwise present in DFT. We would also like to point out that QM:QM offers an easy determination of diabatic states defined by their charge and that we can also approach even energetically high-lying structures or arbitrarily strong fields.

The paper is organized as follows. We first describe the fragmentation approach used in the present work. Next, we clarify the simulation details. Then we show the results and we discuss connotations. At the end, we present the outlook with possible applicability of the fragmentation approach in radiation chemistry.

2 Methods

To study ionization processes we used a simple energy based fragmentation method in which the energies of the monomer units are calculated in the background charge of the other monomers. This QM:QM approach is analogical to the GEBF, FMO and EFP methods. A great simplification compared to the GEBF and FMO methods is that only fragment monomer energies are calculated.

The approach is schematically depicted on Figure 1.

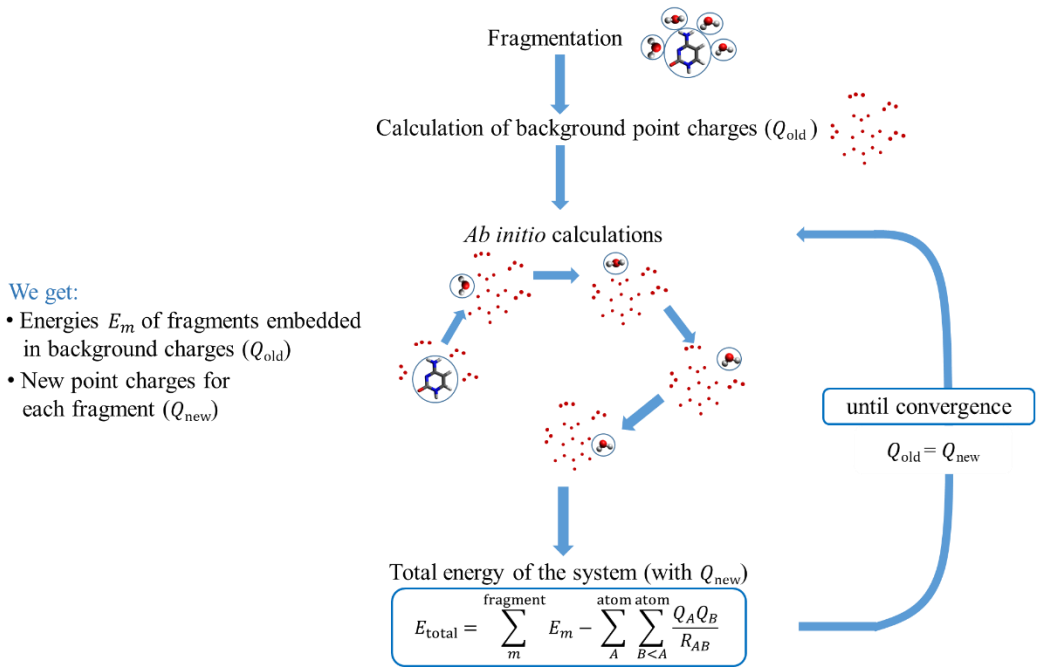

Figure 1. *A sketch showing the QM:QM method applied in this study.*

In the first step, the system is divided into fragments. In our case, each fragment coincides with a molecule. In the next step, the electronic energy of the fragments is calculated. The electronic energies were calculated at the DFT level, though in principle, any *ab initio* QM method may be used. The electrostatic potential of other fragments is taken into account by replacing the atoms outside of the given fragment with point charges. As the atomic charges of the calculated fragment depend on the charges of the environment, the charges are updated after the fragment calculations. This leads to a self-consistent iteration process. The point charges are generated by fitting the electrostatic potential according to the Merz-Singh-Kollman scheme.[58]

In each iteration step the total energy is calculated as a sum of the fragment energies

$$E_{\text{total}} = \sum_m^M E_m - \sum_A \sum_{A<B} \frac{Q_A Q_B}{R_{AB}}, \qquad (1)$$

where index *m* denotes a fragment and the maximum number of fragments is *M*. The Q_A is the point charge of atom *A* and R_{AB} is the distance between atoms *A* and *B*. The second term (here in atomic units) is added to avoid double counting of the electrostatic energy in the first term. The fragments were calculated at the BMK/6-31+g* level.[59] Such basis set is small enough to provide reliable charges. Ionization energy was calculated as a difference between the total energy of the ionized and non-ionized systems.

The initial ensemble was generated by classical molecular dynamics performed with Gromacs 5.0.5 code.[60] Water was in all cases described by the SPC/E model.[61] The force field of cytosine was generated by an automated builder[62] from a B3LYP/6-31g* optimized geometry. Bonded and van der Waals parameters were taken from the GROMOS 54A7 parameter set,[14] initial charges were estimated using the Merz-Kollman scheme.[58] Parameters for the sodium cation were taken from ref.[63]

For cytosine, two types of simulations were performed: simulation for cytosine molecule in a box filled with water molecules and simulation for cytosine molecule in water droplet. Simulation box contained 10040 water molecules, simulations of droplets contained 50, 100, 200, 400 and 600 water molecules. The time step for the production run was set to 0.5 fs and the total simulation lengths was 5 ns. In the case of box simulations 3D periodic boundary conditions were employed. The pressure of 1 bar was controlled by the Berendsen barostat with a coupling constant of 0.5 ps and constant temperature of 300 K was controlled by the velocity rescaling thermostat[64] with a coupling time of 0.5 ps (0.1 ps for sodium). Cytosine molecule was kept in centre of the droplet by the pull code implemented in Gromacs 5.0.5.[60] Spherical

water clusters containing 50, 100, 200, 400, 600 water molecules were cut out of the simulation box. The water molecules were selected based on the distance to the centre of mass of the cytosine molecule. The individual geometries were selected randomly from the MD trajectories, the total number ranged from 100 to 400, as specified in each case.

Simulations with a sodium cation contained one cation in a box of 33151water molecules (100 Å box), time step was set to 2 fs, total length of the simulation was 4 μs. From the trajectory, 200 spherical clusters containing a defined number of water molecules (see in discussion) were extracted and used for further calculations. Simulations were performed in periodic boundary condition, pressure was controlled by the Berendsen barostat and constant temperature of 300 K was controlled by the velocity rescaling thermostat.[64]

The mean value of the vertical ionization energy, $\mathrm{E}[IE]$ is calculated over the sample geometries selected from the classical molecular dynamics trajectory. The standard deviation by definition is $\sigma = \sqrt{\mathrm{E}[IE - \mathrm{E}[IE]]^2}$. The standard error of the mean value is derived from the standard deviation: $\mathrm{SEM} = \frac{\sigma}{\sqrt{n}}$, where n is the number of samples. The full width at half maximum (FWHM) was calculated approximately supposing normal distribution: $\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma$. The standard error of the FWHM can be expressed using the fourth central moment $\mu_4 = \mathrm{E}[(IE - \mathrm{E}[IE])^4]$[65]

$$\mathrm{SE(FWHM)} = \frac{1}{2\sigma}\sqrt{\frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)}. \qquad (2)$$

Because the number of points was limited, for the purpose of presentation of the calculated spectra we utilized an advanced kernel density estimation method with the Gaussian kernel in which we replaced each single point with a narrow Gaussian.[66,67] The width of the individual Gaussians was determined by the simple Silverman's rule of thumb[68] (suitable for unimodal distributions). We also fitted the spectra to a single Gaussian function and we provide the fitted values in the insets of the figures.

3 Results and Discussion

First, we discuss the performance of the various models on the test molecule of cytosine dissolved in water. The example was selected due to the biological importance of nucleic acid bases ionization as well as due to the small size of the chromophore. Note that the ionization

energy of cytosine in water has not been directly measured experimentally because of its very low solubility. However, we can make a fair comparison with a corresponding nucleoside. The first ionization energies of the base and of the nucleoside are similar and the solvent shift for the nucleoside is known.[6] The experimental value of the VIEs for cytosine in the gas phase is 8.89±0.02 eV[69] and for cytidine is 8.46 eV[70]. The experimental value for the VIE of cytidine in water is 8.1 eV,[71] *i.e.* the solvent shift is 0.36 eV. We can expect a similar ~0.3–0.4 eV solvent shift for cytosine as the ionization is localized on the cytosine unit.[6]

All the energies are calculated with the BMK functional and 6-31+g* basis set. This functional represents a global hybrid functional often used for the evaluation of energetics in the nucleic acid systems (see *e.g.* ref.[6] and references therein).

A simple estimate of the solvent shift can be done by polarizable solvent model. The calculated value of VIE for cytosine in the gas phase at the BMK/6-31+g* level is 8.73 eV and the corresponding value in the polarizable continuum is only 7.95 eV. We can see that the ionization energetics is captured relatively well yet solvent shift of 0.78 eV is significantly overestimated.

Another estimate can be done by a simple electrostatic embedding QM/MM scheme (EE-QM/MM, *e.g.* we standardly divide the system into a quantum zone and an MM zone). The SPC/E charges[61] were used for the water molecules throughout the simulations. Figure 2 shows the evolution of the VIE distributions with the system size, *i.e.* with the number of solvating molecules for a system containing 5 water molecules in the QM region. The mean value of VIE and the width of the distribution expressed as full width at the half of the maximum (FWHM) are shown in the inset of Figure 2. The corresponding numerical values of all the data are collected in the Supporting Information (Table S1). We can observe that the calculated VIEs depend only very little on the number of solvating molecules. In fact, the ionization energy is slightly higher for the solvated system when compared with the isolated cytosine (Table S1). This observation agrees well with previous studies employing electrostatic embedding scheme, where large increase in the electron binding energies were reported and later were found to be an artefact of the non-polarizable models.[24,29] We also observe that the width of the distribution is saturated quickly with the number of solvent molecules, with a value around 0.7 eV. This is close to the experimentally determined widths of nucleosides in liquid phase in the range 0.8–0.9 eV.[71] Note that we use constrained geometry of cytosine in our simulations and we therefore neglect contribution of inner sphere reorganization energy into the FWHM.

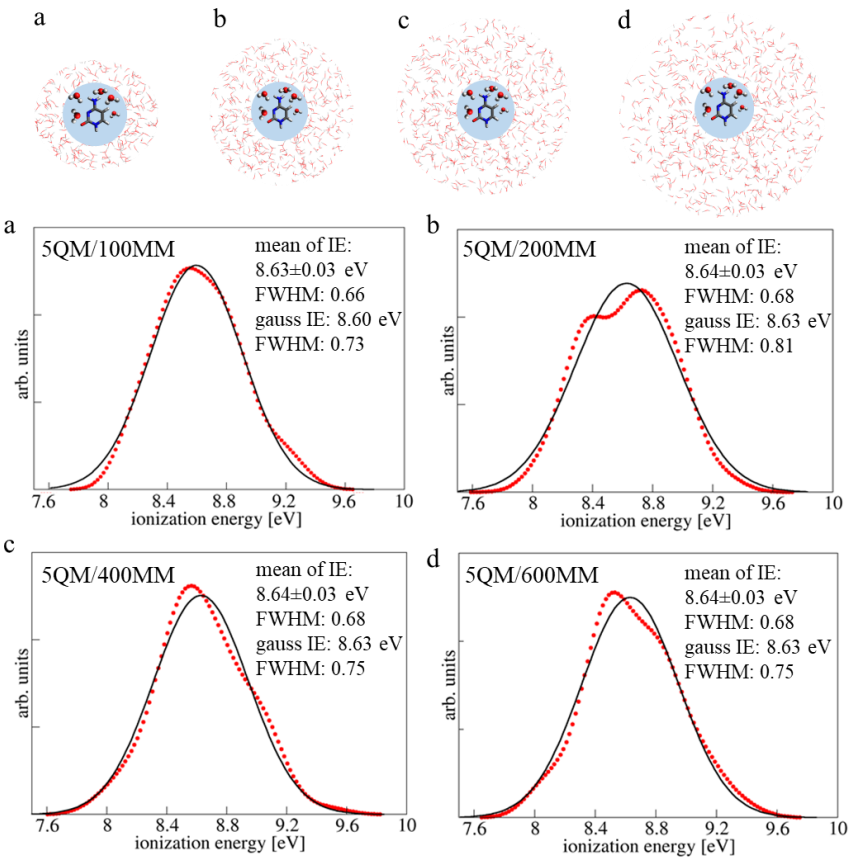

Figure 2. Ionization of cytosine within the electrostatic embedding QM/MM model. The evolution of the VIE distribution for cytosine as a function of a system size: a) 100, b) 200, c) 400, d) 600 molecules in total, 5 water molecules closest to the center of mass of cytosine were considered explicitly at the QM level, the remaining molecules were treated as point charges. Calculated mean values of the VIE and FWHM are shown in the inset of the pictures. SPC/E charges were used for water molecules. The red curve corresponds to kernel density estimation method, in which each single point was replaced with a narrow Gaussian, the black curve shows a Gaussian fit of the distribution. Data in graphs were collected from 100 independent calculations of VIE. The individual geometries were selected randomly from the MD simulation of cytosine in a box.

In the next step, we gradually extended the quantum zone in the EE-QM/MM approach. Cytosine and a defined number of adjacent water molecules (0, 5, 10, 20, 40, 60) were calculated at the DFT level (BMK/6-31+g* level in our case) and the remaining water molecules were calculated at the MM level (SPC/E). Such a combined scheme was used recently for studying ionization and electron attachment to nucleic acid components and its successful application was reported.[72] In this case, we could observe an interesting trend:

addition of explicit solvent molecules lowered the ionization energy while addition of non-polarizable solvent molecules corrected this lowering. Combination of the largest quantum zones (40, resp. 60 water molecules in QM zone and additional 560, resp. 520 MM molecules) provided an ionization energy of 8.3 eV, *e.g.* a reasonable solvent shift of 0.43 eV was obtained. The data are shown in Figure 3 and the numerical values are again collected in the Supplementary Information Table S1. The width of the distribution remains essentially constant for all the simulations.

Application of the above scheme is based on the assumption that the QM description itself is correct for all system sizes and we would ultimately converge to the experimental value within the purely *ab initio* approach. This is, however, an assumption which is not justified for the common DFT functionals. The approximate functionals were optimized on sets of small molecules and the performance of the functionals for large systems should be always carefully inspected. Therefore we investigated the performance of DFT on large clusters of cytosine and water. We gradually increased the number of solvating water molecules from 5 to 60. Table 1 shows the calculated mean value of the VIE for these clusters calculated at the HF, BMK, PBE and LC-ωPBE levels with 6-31+g* basis set. The DFT functionals were selected so that they have varying contribution of the HF exchange. We can observe that with increasing the cluster size, the ionization energy quickly drops. In fact, it drops much faster than one could infer from electrostatic considerations and the calculated value is below the experimental one for solvated cytidine. We also present an integrated charge for the cytosine unit in the ionized state in Table 1, showing immediately some warning. The BMK functional predicts that the positive hole will be to a much extent delocalized over neighbouring water molecules. This is clearly an artefact of the approximate DFT methods which can be attributed to the self-interaction error.[73] The artificial charge leak is enhanced for functionals with a lower fraction of the exact exchange as it is evident from the comparison between the HF, BMK, PBE and LC-ωPBE functionals (see Table 1): while HF method keeps the charge localized, the GGA functional PBE exhibits a huge charge leak. For BMK and LCωPBE the leak is not so pronounced. This significantly complicates convergence for large systems, because charge delocalization grows with increasing the system size. The problem can be partially limited if we add extra non-polarized water molecules around the QM zone, in this case the charge can be to some extent back-localized (see Table S2 where data for BMK and PBE are presented).

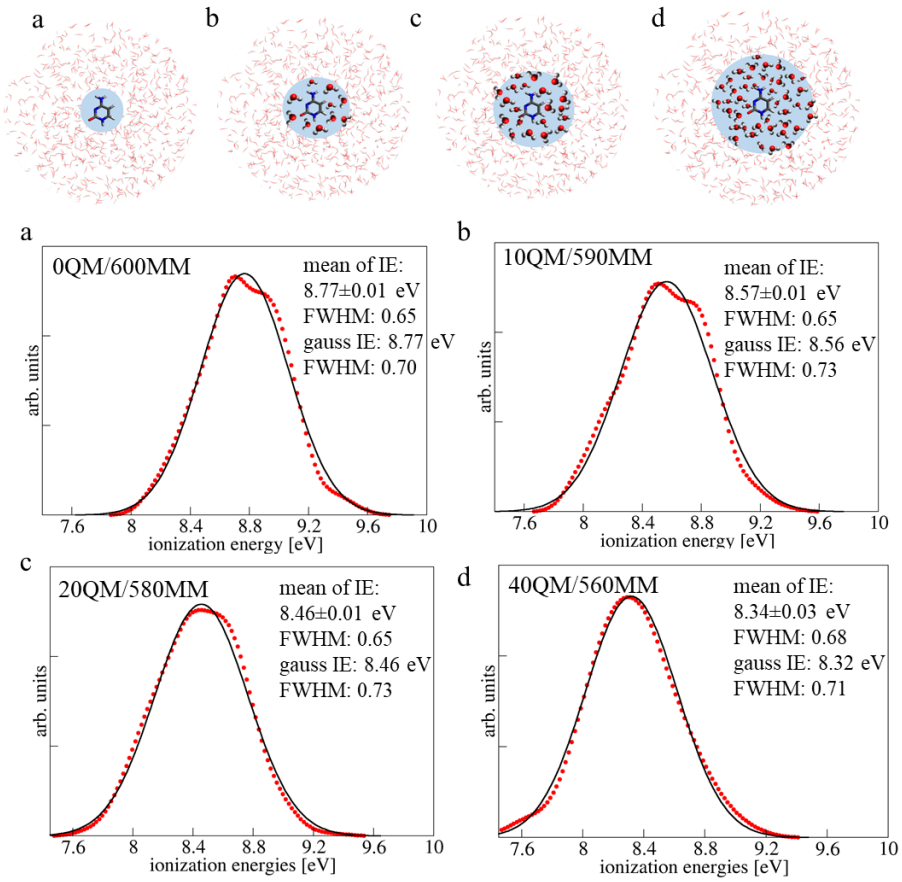

Figure 3. VIE of cytosine within the electrostatic embedding QM/MM model as a function of an increasing number of QM water molecules: a) 0, b) 10, c) 20, d) 40. Calculated mean values of the VIE and FWHM are shown in the inset of the picture. SPC/E charges were used for water molecules. Data in graphs a), b), c) were collected from 400 independent calculations of VIE, graph d) from 100. The individual geometries were selected randomly from the MD simulation of cytosine in a box. The red curve corresponds to kernel density estimation method, in which each single point was replaced with a narrow Gaussian, the black curve shows a Gaussian fit of the distribution.

	HF		BMK		PBE		LC-ωPBE	
# water in QM	VIE	Q^+(cyt)	VIE	Q^+(cyt)	VIE	Q^+(cyt)	VIE	Q^+(cyt)
0	7.430±0.020	-	8.718±0.016	-	8.442±0.008	-	8.892±0.012	-
5	7.660±0.042	0.960±0.025	8.773±0.029	0.957±0.009	8.017±0.018	0.598±0.009	8.984±0.030	0.996±0.007
10	7.630±0.042	1.021±0.009	8.689± 0.031	0.855±0.014	7.692±0.020	0.459±0.012	8.936±0.035	0.969±0.011
20	7.513±0.041	1.010±0.015	8.431±0.027	0.756±0.019	7.282±0.045	0.189±0.023	8.716±0.031	0.933±0.015
40	7.315±0.057	0.885±0.046	8.170±0.028	0.604±0.023	6.701±0.053	0.305±0.031	8.556±0.050	0.747±0.047
60	7.332±0.059	0.887±0.042	8.076±0.045	0.544±0.035	-	-	8.538±0.053	0.689±0.055

Table 1. Calculated mean value of the VIE for clusters of cytosine and water. Q^+(cyt) is the charge localized on the ionized cytosine molecule. The system contained from 5 to 60 water molecules explicitly calculated at the HF, BMK, PBE and LC-ωPBE/6-31+g level. Data presented in the table were obtained for a cluster cut from a box simulation.*

Next, we present the data for the fragmentation QM:QM scheme which, as we claim, is ideally suited for the description of the ionization in the condensed phase. Figure 4 shows the calculated VIE of cytosine in water clusters of increasing size, the corresponding data are shown in Table S3. We compare two types of clusters – a) clusters of given sizes (50, 100, 200, 400, 600 water molecules) simulated *in vacuo* or b) clusters cut from a large box simulated with periodic boundary conditions. In both types of clusters, each molecule represents a fragment and the fragments communicate only electrostatically. For both types of clusters, we can see in Figure 4 a gradual shift of ionization energy from 8.73 eV in the gas phase to 8.36 eV (for a) or to 8.27 eV (for b). In both cases, the solvent shift of VIE is close to the value observed for cytidine. At the same time, the value is significantly above the value calculated with the polarizable continuum model, which is 7.95 eV. The specific intermolecular interactions of the solvent clearly are important and should not be excluded. Unfortunately, the graphs in Figure 4 show that even for the biggest clusters the complete convergence has not been reached. Yet from the observed trends we can answer how many water molecules mimics the solution – when it comes to ionization of neutral organic species in solutions, at least several hundreds of water molecules must be included.

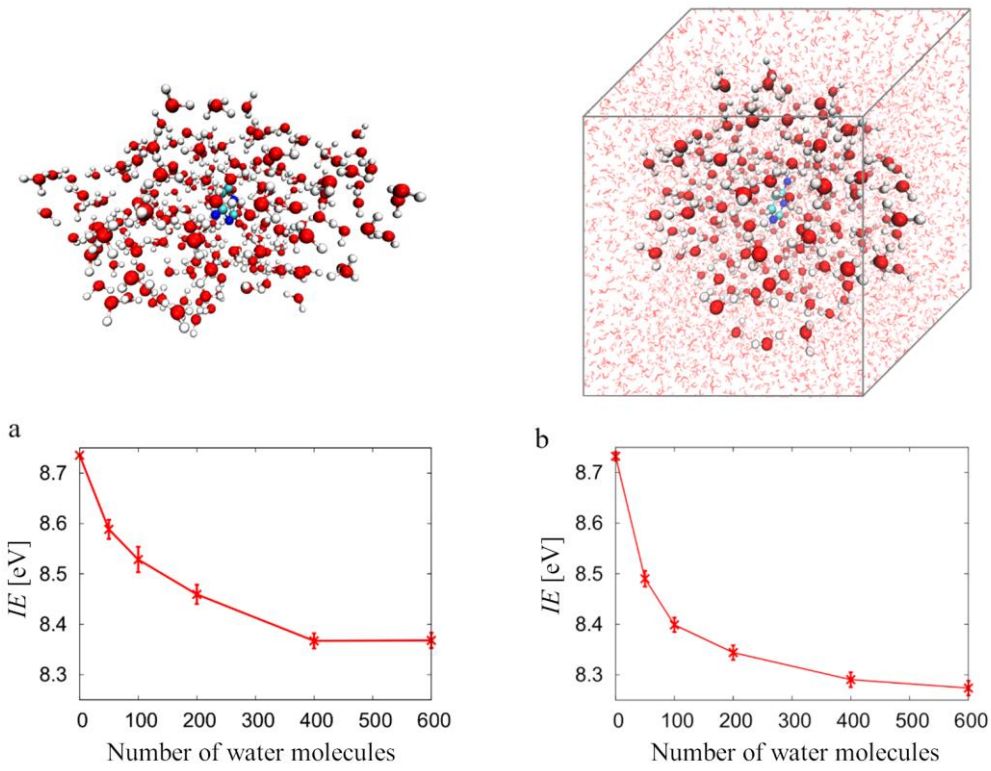

Figure 4. VIE of cytosine within the QM:QM model as a function of an increasing system size. In the QM:QM scheme, the BMK functional with 6-31+g basis set was used. Results are*

presented for clusters obtained either from classical simulations of a given cluster size in vacuo (a), or for clusters obtained by cutting from a large box simulation (b).

Concerning the widths of the spectra, the FWHM of the distribution changes much less with the cluster sizes and the convergence is essentially obtained for clusters containing 200 water molecules. The widths of the distribution converges to 0.68 eV (see Table S3). The widths are at the same time close to the one calculated with the SPC/E charges for water (compare Tables S1 and S3). The width of the distribution relates via the Marcus formula to the reorganization energy, which is not affected by the electronic part of the polarization.

The values obtained within the QM:QM approach can be corrected by $\Delta VIE_{\mathrm{solv}}$ calculated from dielectric continuum assumptions

$$\Delta\mathrm{VIE}_{\mathrm{solv}} = \frac{e^2}{8\pi\varepsilon_0 R}\left(1 - \frac{1}{\varepsilon_{\mathrm{opt}}}\right). \qquad (3)$$

Figure 5 shows the evolution of the VIE with the system size compared to the value obtained with dielectric continuum correction and with the polarizable continuum model. The radius R of the spherical clusters was estimated by the MSMS (Maximal Speed Molecular Surface) code.[74] We can see that the result obtained via polarizable continuum model (7.95 eV) is well below what appears to be the limit of the QM:QM calculation. We also see that the corrected cluster calculations are rather stable with the cluster size and that the two calculations slowly but gradually converge to the same limit. The calculated VIE in the limit is below the experimental value which can be attributed to the inaccuracy of the BMK functional; it provides the gas phase ionization energy of cytosine 0.1–0.2 eV below the experiment.[6] It follows from Figure 5 that the solvent shift is described with the QM:QM scheme within 0.1 eV; 8.1 eV being the lower bound and 8.27 eV being the upper bound.

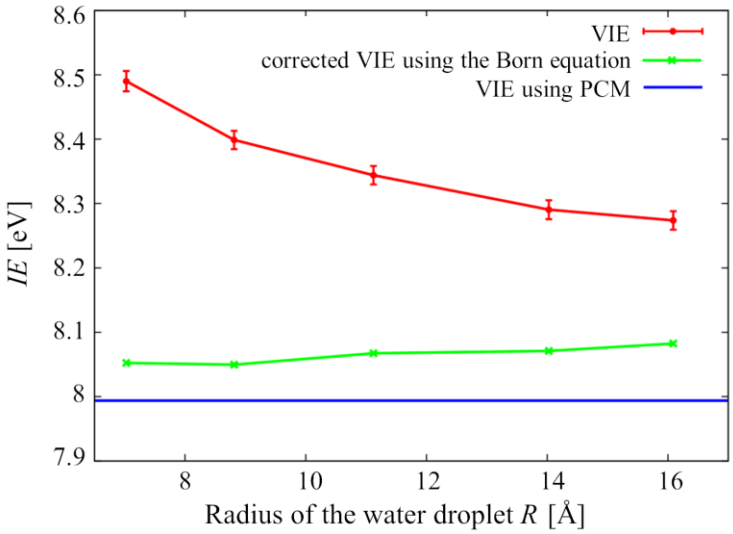

Figure 5. Corrected VIE via Born formula (green), result obtained via the polarizable continuum model (blue) and the QM:QM dependence of VIE on the size of the cluster (red).

We shall now briefly discuss several more technical aspects of the QM:QM simulations. We start with a discussion of the convergence of the fragments charge. The simulations always started with the charges calculated for isolated molecules. It then required 3, at most 4 cycles to converge both in the ground and ionized states within 0.001 eV.

To understand the polarization effect of a charged solute in the solvent, the dipole moments of water molecules were calculated and projected onto the vector pointing from the centre of mass of water to the centre of mass of cytosine. The ionized solute increases the solvent dipole moment components pointing to the direction of cytosine as shown in Figure 6. The induced dipole moment generated by an ion is $\mu = \alpha E = \alpha \frac{ze}{4\pi\varepsilon_0 \varepsilon R^2}$, where E is the electric field of the ion and α is the polarizability of water. While the dipole moment of water molecules will be affected also by modified dipoles of the other water molecules, the leading term for the induced dipole should be inversely proportional to R^2. In accordance with this expectation, the observed dipole moment change of water molecules in Figure 6 is inversely proportional to the distance squared.

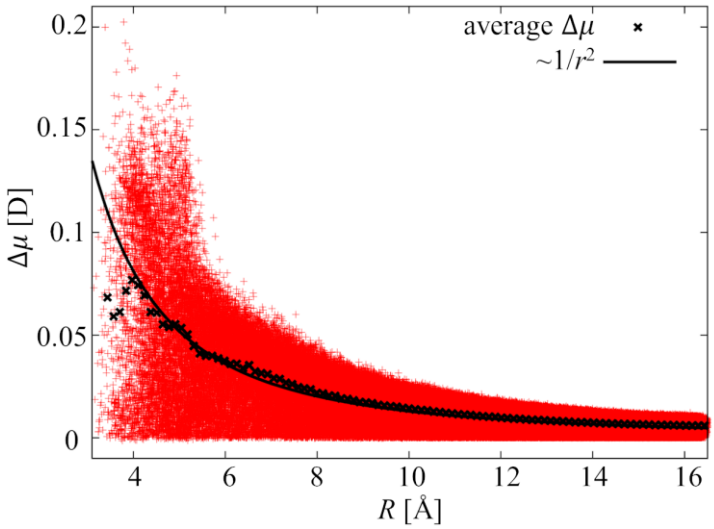

Figure 6. The increase of a dipole moment component of water pointing to the direction of the cytosine molecule caused by ionization of cytosine. Each point represents a difference between the dipole moment of a water molecule in the ionized and ground state of the system. The results are shown for 400 clusters containing 600 water molecules. R is the distance between the centres of mass of water and cytosine.

We further addressed a question of statistical convergence with the number of samples from a perspective of both the ionization energy and the width of the spectrum. It is clear that 100 samples are enough to get quantities with error bars within the systematic error of the DFT method itself. The results are presented in the Supporting Information Table S4 and S5.

So far we explored the QM:QM approach on an example of an ionized neutral organic molecule. The solvent effects in this case are relatively modest and the solvation can be reasonably well described within a linear regime. The long-term goal is, however, to use the QM:QM approach for highly charged particles formed within radiation chemistry processes, *e.g.* within the Auger decay.[75,76] Here, the solvent shift can easily exceed 10 eV as doubly (and generally multiply) charged ions are formed.

We thus tested the performance of the QM:QM approach for the vertical ionization of a sodium cation, *i.e.* for a process $Na^{+} \rightarrow Na^{++}$. The experimental gas phase energy for this reaction is 47.3 eV while the experimental value in water is 35.5 eV.[54] The observed solvent shift is 12.2 eV. The theoretical VIE for an isolated sodium atom in the gas phase is 49.88 eV (using the BMK/6-31+g* approach). We observe that the QM:QM approach provides us with a solvent shift of 10.7 eV already for 50 water molecules and then it gradually increases (see Fig. 7). The solvent shift for the largest cluster used with 1200 water molecules gives a solvent shift of 12.7

eV; the convergence with respect to the cluster size is still not fully reached. If QM/MM calculations with the SPC/E water model are used, we find much smaller values of the solvent shift (8.5 eV for the largest cluster). The polarizable continuum model provides a reasonable value of the solvent shift (10.5 eV).

Note that it is conceptually rather difficult to make an explicit calculation for the present system without fragmentation. For doubly charged sodium-water clusters, states with charge delocalized between more solvent units would be lower in energy. Yet these states are not formed under experimental conditions (a doubly charged sodium cation is formed).

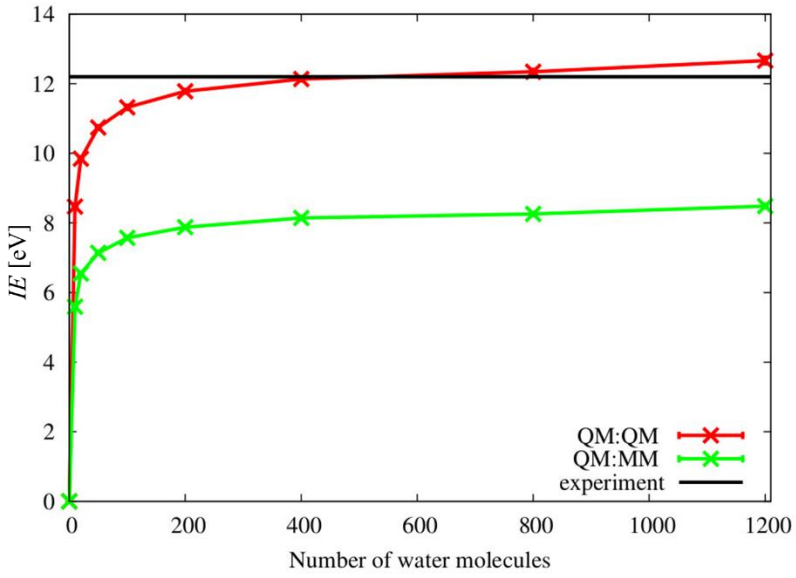

Figure 7 Solvent shift of VIE of sodium cation within the QM:QM model as a function of an increasing system size. In the QM:QM scheme, the BMK functional with 6-31+g basis set was used. Value for an isolated sodium atom surrounded by a polarizable continuum as well as the experimental values both in the gas phase and in solutions are marked.*

4. Conclusions

We investigated the performance of QM:QM scheme for vertical ionization of cytosine and sodium cation in water. We show that a direct simulation of the ionization of organic solutes in liquid water with DFT methods provides spurious results. When we investigated large solvated clusters, we observed an artificial charge delocalization from the originally ionized unit. We could observe the limits of DFT for simulations of extended systems, which have become

technically feasible only recently. We argue, that in particular cases, the results should be weighted very carefully.[28]

We furthermore argue, that simple QM/MM techniques based on electrostatic embedding scheme tend to overestimate the ionization energies of solvated molecules due to the lack of electronic polarization. Therefore QM/MM can be useful *e.g.* for stabilizing anions in the ground state, but its applicability for VIEs is limited. The combination of the QM/MM with an extended QM zone can compensate this problem but only as a matter of error cancellation. On the other hand, the ionization energies calculated within the polarizable continuum model are shifted to lower values and furthermore, the spectral width is not accessible.[77]

We show that the fragmentation QM:QM approach represents a reliable approach for ionization in the condensed phase. We can observe a gradual transition of VIE from an isolated molecule to the bulk with the solvent shift below 0.5 eV (for largest studied clusters). Yet even for a large number of water molecules, the solvent shift is not completely converged. The width of the spectra is estimated with QM:QM reasonably well.

The present approach is similar in its spirit to the widely used EFP method,[29,46,47] which has been successfully used for ionized systems in the liquid phase. We believe, that the QM:QM scheme in its present form has some advances and will be useful for highly charged systems where the linear polarizability is not sufficient (iply charged anions [78]). Another useful application is the description of energetics in autoionization processes. Here, multiply charged cations are formed and their energetics in the condensed phase is not well known. We think that fragment-based methods will help access the accurate energetics with the use of ground state electronic structure methods. Fragmentation is then a natural way to describe these systems within all-atom simulations. Our tests on sodium cation showed a reasonable quantitative performance of the model. However, a large number of solvating molecules is required for converged results. In practice, it would be meaningful to use extrapolation schemes to reach the experimental values.

Acknowledgement

The authors gratefully acknowledge the support of Czech Science Foundation (project number 18-23756S).

References

[1] Zeng X, Hu H, Hu X, Cohen A J and Yang W 2008 Ab Initio QM/MM Simulation of Electron Transfer Process: Fractional Electron Approach *J. Chem. Phys.* **128** 124510

[2] Blumberger J 2015 Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions *Chem. Rev.* **115** 11191–238

[3] Maroncelli M, Macinnis J and Fleming G R 1989 Polar Solvent Dynamics and Electron-Transfer Reactions *Science* **243** 1674–81

[4] Warshel A 1982 Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions *J. Phys. Chem.* **86** 2218–24

[5] Marcus R A 1956 Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization. I *J. Chem. Phys.* **24** 979–89

[6] Pluhařová E, Slavíček P and Jungwirth P 2015 Modeling Photoionization of Aqueous DNA and Its Components *Acc. Chem. Res.* **48** 1209–17

[7] Slavíček P, Winter B, Faubel M, Bradforth S E and Jungwirth P 2009 Ionization Energies of Aqueous Nucleic Acids: Photoelectron Spectroscopy of Pyrimidine Nucleosides and ab Initio Calculations *J. Am. Chem. Soc.* **131** 6460–7

[8] Jagoda-Cwiklik B, Slavíček P, Cwiklik L, Nolting D, Winter B and Jungwirth P 2008 Ionization of Imidazole in the Gas Phase, Microhydrated Environments, and in Aqueous Solution *J. Phys. Chem. A* **112** 3499–505

[9] Winter B and Faubel M 2006 Photoemission from Liquid Aqueous Solutions *Chem. Rev.* **106** 1176–211

[10] Seidel R, Thürmer S and Winter B 2011 Photoelectron Spectroscopy Meets Aqueous Solution: Studies from a Vacuum Liquid Microjet *J. Phys. Chem. Lett.* **2** 633–41

[11] Ayala R and Sprik M 2008 A Classical Point Charge Model Study of System Size Dependence of Oxidation and Reorganization Free Energies in Aqueous Solution *J. Phys. Chem. B* **112** 257–69

[12] Bayly C I, Merz K M, Ferguson D M, Cornell W D, Fox T, Caldwell J W, Kollman P A, Cieplak P, Gould I R and Spellmeyer D C 1995 A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **117** 5179–97

[13] MacKerell A D, Bashford D, Bellott M, Dunbrack R L, Evanseck J D, Field M J, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F T K, Mattos C, Michnick S, Ngo T, Nguyen D T, Prodhom B, Reiher W E, Roux B, Schlenkrich M, Smith J C, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D and Karplus M 1998 All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins *J. Phys. Chem. B* **102** 3586–616

[14] Schmid N, Eichenberger A P, Choutko A, Riniker S, Winger M, Mark A E and van Gunsteren W F 2011 Definition and testing of the GROMOS force-field versions 54A7 and 54B7 *Eur. Biophys. J.* **40** 843

[15] Jorgensen W L, Maxwell D S and Tirado-Rives J 1996 Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids *J. Am. Chem. Soc.* **118** 11225–36

[16] Aguilar M A, Olivares del Valle F J and Tomasi J 1993 Nonequilibrium solvation: An ab initio quantum-mechanical method in the continuum cavity model approximation *J. Chem. Phys.* **98** 7375–84

[17] Cammi R and Mennucci B 1999 Linear response theory for the polarizable continuum model *J. Chem. Phys.* **110** 9877–86

[18] Cossi M and Barone V 2001 Time-dependent density functional theory for molecules in liquid solutions *J. Chem. Phys.* **115** 4708–17

[19] Chipman D M 2009 Vertical electronic excitation with a dielectric continuum model of solvation including volume polarization. I. Theory *J. Chem. Phys.* **131** 014103

[20] Chipman D M 2009 Vertical electronic excitation with a dielectric continuum model of solvation including volume polarization. II. Implementation and applications *J. Chem. Phys.* **131** 014104

[21] Pluhařová E, Jungwirth P, Bradforth S E and Slavíček P 2011 Ionization of Purine Tautomers in Nucleobases, Nucleosides, and Nucleotides: From the Gas Phase to the Aqueous Environment *J. Phys. Chem. B* **115** 1294–305

[22] Simm G N, Türtscher P L and Reiher M 2019 Systematic Microsolvation Approach with a Cluster-Continuum Scheme and Conformational Sampling *ArXiv190906664 Phys.*

[23] Loco D, Polack É, Caprasecca S, Lagardère L, Lipparini F, Piquemal J-P and Mennucci B 2016 A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations *J. Chem. Theory Comput.* **12** 3654–61

[24] Ghosh D, Isayev O, Slipchenko L V and Krylov A I 2011 Effect of Solvation on the Vertical Ionization Energy of Thymine: From Microhydration to Bulk *J. Phys. Chem. A* **115** 6028–38

[25] Ghosh D, Kosenkov D, Vanovschi V, Williams C F, Herbert J M, Gordon M S, Schmidt M W, Slipchenko L V and Krylov A I 2010 Non-covalent interactions in extended systems described by the Effective Fragment Potential method: Theory and application to nucleobase oligomers *J. Phys. Chem. A* **114** 12739–54

[26] Mori-Sánchez P, Cohen A J and Yang W 2006 Many-electron self-interaction error in approximate density functionals *J. Chem. Phys.* **125** 201102

[27] Mori-Sánchez P, Cohen A J and Yang W 2008 Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction *Phys. Rev. Lett.* **100** 146401

[28] Isborn C M, Mar B D, Curchod B F E, Tavernelli I and Martínez T J 2013 The Charge Transfer Problem in Density Functional Theory Calculations of Aqueously Solvated Molecules *J. Phys. Chem. B* **117** 12189–201

[29] Gordon M S, Fedorov D G, Pruitt S R and Slipchenko L V 2012 Fragmentation Methods: A Route to Accurate Calculations on Large Systems *Chem. Rev.* **112** 632–72

[30] Herbert J M 2019 Fantasy versus reality in fragment-based quantum chemistry *J. Chem. Phys.* **151** 170901

[31] Ikegami T, Ishida T, Fedorov D G, Kitaura K, Inadomi Y, Umeda H, Yokokawa M and Sekiguchi S 2005 Full Electron Calculation Beyond 20,000 Atoms: Ground Electronic State of Photosynthetic Proteins *SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing* SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing pp 10–10

[32] Collins M A and Bettens R P A 2015 Energy-Based Molecular Fragmentation Methods *Chem. Rev.* **115** 5607–42

[33] Li W, Li S and Jiang Y 2007 Generalized Energy-Based Fragmentation Approach for Computing the Ground-State Energies and Properties of Large Molecules *J. Phys. Chem. A* **111** 2193–9

[34] Li S, Li W and Ma J 2014 Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates *Acc. Chem. Res.* **47** 2712–20

[35] Li Y, Yuan D, Wang Q, Li W and Li S 2018 Accurate prediction of the structure and vibrational spectra of ionic liquid clusters with the generalized energy-based fragmentation approach: critical role of ion-pair-based fragmentation *Phys. Chem. Chem. Phys.* **20** 13547–57

[36] Hua W, Fang T, Li W, Yu J-G and Li S 2008 Geometry Optimizations and Vibrational Spectra of Large Molecules from a Generalized Energy-Based Fragmentation Approach *J. Phys. Chem. A* **112** 10864–72

[37] Zhang D W, Xiang Y and Zhang J Z H 2003 New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin−Biotin Interaction Energy *J. Phys. Chem. B* **107** 12039–41

[38] Zhang D W and Zhang J Z H 2003 Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy *J. Chem. Phys.* **119** 3599–605

[39] Ganesh V, Dongare R K, Balanarayan P and Gadre S R 2006 Molecular tailoring approach for geometry optimization of large molecules: Energy evaluation and parallelization strategies *J. Chem. Phys.* **125** 104109

[40] Sahu N and Gadre S R 2014 Molecular Tailoring Approach: A Route for ab Initio Treatment of Large Clusters *Acc. Chem. Res.* **47** 2739–47

[41] Kitaura K, Ikeo E, Asada T, Nakano T and Uebayasi M 1999 Fragment molecular orbital method: an approximate computational method for large molecules *Chem. Phys. Lett.* **313** 701–6

[42] Fedorov D G and Kitaura K 2004 The importance of three-body terms in the fragment molecular orbital method *J. Chem. Phys.* **120** 6832–40

[43] Dahlke E E and Truhlar D G 2007 Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters *J. Chem. Theory Comput.* **3** 46–53

[44] Dahlke E E and Truhlar D G 2007 Electrostatically Embedded Many-Body Correlation Energy, with Applications to the Calculation of Accurate Second-Order Møller−Plesset Perturbation Theory Energies for Large Water Clusters *J. Chem. Theory Comput.* **3** 1342–8

[45] Sorkin A, Dahlke E E and Truhlar D G 2008 Application of the Electrostatically Embedded Many-Body Expansion to Microsolvation of Ammonia in Water Clusters *J. Chem. Theory Comput.* **4** 683–8

[46] Gordon M S, Freitag M A, Bandyopadhyay P, Jensen J H, Kairys V and Stevens W J 2001 The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry *J. Phys. Chem. A* **105** 293–307

[47] Gordon M S, Slipchenko L, Li H and Jensen J H 2007 Chapter 10 The Effective Fragment Potential: A General Method for Predicting Intermolecular Interactions *Annual Reports in Computational Chemistry* vol 3, ed D C Spellmeyer and R Wheeler (Elsevier) pp 177–93

[48] Stone A J 1981 Distributed multipole analysis, or how to describe a molecular charge distribution *Chem. Phys. Lett.* **83** 233–9

[49] DeFusco A, Minezawa N, Slipchenko L V, Zahariev F and Gordon M S 2011 Modeling Solvent Effects on Electronic Excited States *J. Phys. Chem. Lett.* **2** 2184–92

[50] Yoo S, Zahariev F, Sok S and Gordon M S 2008 Solvent effects on optical properties of molecules: A combined time-dependent density functional theory/effective fragment potential approach *J. Chem. Phys.* **129** 144112

[51] Arora P, Slipchenko L V, Webb S P, DeFusco A and Gordon M S 2010 Solvent-Induced Frequency Shifts: Configuration Interaction Singles Combined with the Effective Fragment Potential Method *J. Phys. Chem. A* **114** 6742–50

[52] Slipchenko L V 2010 Solvation of the Excited States of Chromophores in Polarizable Environment: Orbital Relaxation versus Polarization *J. Phys. Chem. A* **114** 8824–30

[53] Pluhařová E, Schroeder C, Seidel R, Bradforth S E, Winter B, Faubel M, Slavíček P and Jungwirth P 2013 Unexpectedly Small Effect of the DNA Environment on Vertical Ionization Energies of Aqueous Nucleobases *J. Phys. Chem. Lett.* **4** 3766–9

[54] Winter B, Weber R, Hertel I V, Faubel M, Jungwirth P, Brown E C and Bradforth S E 2005 Electron Binding Energies of Aqueous Alkali and Halide Ions: EUV Photoelectron Spectroscopy of Liquid Solutions and Combined Ab Initio and Molecular Dynamics Calculations *J. Am. Chem. Soc.* **127** 7203–14

[55] Genereux J C, Boal A K and Barton J K 2010 DNA-Mediated Charge Transport in Redox Sensing and Signaling *J. Am. Chem. Soc.* **132** 891–905

[56] Morgan W F 2003 Non-targeted and Delayed Effects of Exposure to Ionizing Radiation: II. Radiation-Induced Genomic Instability and Bystander Effects In Vivo, Clastogenic Factors and Transgenerational Effects *Radiat. Res.* **159** 581–96

[57] Giese B 2002 Long-Distance Electron Transfer Through DNA *Annu. Rev. Biochem.* **71** 51–70

[58] Singh U C and Kollman P A 1984 An approach to computing electrostatic charges for molecules *J. Comput. Chem.* **5** 129–45

[59] Boese A D and Martin J M L 2004 Development of density functionals for thermochemical kinetics *J. Chem. Phys.* **121** 3405–16

[60] Abraham M J, Murtola T, Schulz R, Páll S, Smith J C, Hess B and Lindahl E 2015 GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers *SoftwareX* **1–2** 19–25

[61] Berendsen H J C, Grigera J R and Straatsma T P 1987 The missing term in effective pair potentials *J. Phys. Chem.* **91** 6269–71

[62] Malde A K, Zuo L, Breeze M, Stroet M, Poger D, Nair P C, Oostenbrink C and Mark A E 2011 An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0 *J. Chem. Theory Comput.* **7** 4026–37

[63] Joung I S and Cheatham T E 2008 Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations *J. Phys. Chem. B* **112** 9020–41

[64] Bussi G, Donadio D and Parrinello M 2007 Canonical sampling through velocity rescaling *J. Chem. Phys.* **126** 014101

[65] Radhakrishna Rao C 1973 *Linear Statistical Inference and Applications* (New York: John Wiley & Sons)

[66] Rosenblatt M 1956 Remarks on Some Nonparametric Estimates of a Density Function *Ann. Math. Stat.* **27** 832–7

[67] Parzen E 1962 On Estimation of a Probability Density Function and Mode *Ann. Math. Stat.* **33** 1065–76

[68] Silverman B W 2018 *Density estimation: For statistics and data analysis*

[69] Trofimov A B, Schirmer J, Kobychev V B, Potts A W, Holland D M P and Karlsson L 2005 Photoelectron spectra of the nucleobases cytosine, thymine and adenine *J. Phys. B At. Mol. Opt. Phys.* **39** 305–329

[70] Yu C, O'Donnell T J and LeBreton P R 1981 Ultraviolet photoelectron studies of volatile nucleoside models. Vertical ionization potential measurements of methylated uridine, thymidine, cytidine, and adenosine *J. Phys. Chem.* **85** 3851–5

[71] Schroeder C A, Pluhařová E, Seidel R, Schroeder W P, Faubel M, Slavíček P, Winter B, Jungwirth P and Bradforth S E 2015 Oxidation Half-Reaction of Aqueous Nucleosides and Nucleotides via Photoelectron Spectroscopy Augmented by ab Initio Calculations *J. Am. Chem. Soc.* **137** 201–9

[72] Zhang Y, Wang J and Yang S 2019 Notable effect of water on excess electron attachment to aqueous DNA deoxyribonucleosides *Phys. Chem. Chem. Phys.* **21** 8925–32

[73] Milanese J M, Provorse M R, Alameda E and Isborn C M 2017 Convergence of Computed Aqueous Absorption Spectra with Explicit Quantum Mechanical Solvent *J. Chem. Theory Comput.* **13** 2159–71

[74] Sanner M F, Olson A J and Spehner J-C 1996 Reduced surface: An efficient way to compute molecular surfaces *Biopolymers* **38** 305–20

[75] Unger I, Seidel R, Thürmer S, Pohl M N, Aziz E F, Cederbaum L S, Muchová E, Slavíček P, Winter B and Kryzhevoi N V 2017 Observation of electron-transfer-mediated decay in aqueous solution *Nat. Chem.* **9** 708–14

[76] Slavíček P, Kryzhevoi N V, Aziz E F and Winter B 2016 Relaxation Processes in Aqueous Systems upon X-ray Ionization: Entanglement of Electronic and Nuclear Dynamics *J. Phys. Chem. Lett.* **7** 234–43

[77] Rubešová M, Jurásková V and Slavíček P 2017 Efficient modeling of liquid phase photoemission spectra and reorganization energies: Difficult case of multiply charged anions *J. Comput. Chem.* **38** 427–37

[78] Pluhařová E, Ončák M, Seidel R, Schroeder C, Schroeder W, Winter B, Bradforth S E, Jungwirth P and Slavíček P 2012 Transforming anion instability into stability: contrasting photoionization of three protonation forms of the phosphate ion upon moving into water *J. Phys. Chem. B* **116** 13254–64