Coupled-cluster based approach for core-level states in condensed phase: Theory and application to different protonated forms of aqueous glycine

Arman Sadybekov, and Anna I. Krylov

AIP | The Journal of Chemical Physics

PERSPECTIVES

Coupled-cluster based approach for core-level states in condensed phase: Theory and application to different protonated forms of aqueous glycine

Arman Sadybekov and Anna I. Krylov
Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA

A theoretical approach for calculating core-level states in condensed phase is presented. The approach is based on the equation-of-motion coupled-cluster (EOM-CC) theory and effective fragment potential (EFP) method. By introducing approximate treatment of double excitations in the EOM-CC with single and double substitutions ansatz, we address poor convergence issues that are encountered for the core-level states and significantly reduce computational costs. While the approximations introduce relatively large errors in the absolute values of transition energies, the errors are systematic. Consequently, chemical shifts, changes in ionization energies relative to reference systems, are reproduced reasonably well. By using different protonation forms of solvated glycine as a benchmark system, we show that our protocol is capable of reproducing the experimental chemical shifts with a quantitative accuracy. The results demonstrate that chemical shifts are very sensitive to the solvent interactions and that explicit treatment of a solvent, such as within EFP framework, is essential for achieving quantitative accuracy. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4990564]

I. INTRODUCTION

By providing tunable high-energy radiation, advanced light sources enable a variety of X-ray based spectroscopies. Among those, X-ray photoelectron spectroscopy (XPS) and near-edge X-ray absorption spectroscopy (NEXAS/XANES) can be described as higher-energy analogs of the VUV-based techniques.[1] XPS determines ionization energies (IEs) of core electrons, whereas NEXAS probes electronic transitions from the core orbitals to low-lying valence or Rydberg orbitals. Owing to the localized nature of the core orbitals, these transitions enable probing the local environment, thus providing complementary information to valence-based techniques.

As in the case of VUV-based techniques, theoretical modeling is required to unambiguously relate experimental measurements to molecular structures. The advances in XPS and NEXAS have stimulated the development of theoretical techniques aiming at describing core-level transitions.[2–19] These theoretical studies have revealed that accurate description of core states is more challenging than the description of valence states.

Superficially, ionization or electronic excitations of core electrons appear to be similar to valence transitions. That is, qualitatively, core-ionized states can be described by the Koopmans theorem, whereas near-edge transitions correspond to single excitations of the core electrons. Thus, one may expect that electronic-structure methods developed for valence transitions can be trivially extended to core levels. Yet, numerical experiments have shown that direct application of standard approaches to core-level transitions leads to unsatisfactory results. For example, application of time-dependent density functional theory (TDDFT) to core excitations results in errors of tens of electron volts (typical errors

are 10 eV for the second row atoms and are much larger for heavier elements). The errors are often systematic but functional-dependent.[5] Recently, Verma and Bartlett have shown[18] that some long-range corrected functionals[20,21] tuned[22–24] to yield Kohn-Sham IEs consistent with the ΔSCF values perform very well for core-ionized and core-excited states: they reported impressive mean absolute errors (MAEs) of less than 1 eV for core-ionized and core-excited states when using their IP-tuned QTP(0,0) functional.[24]

ΔSCF approaches show somewhat better performance,[8,25] but they require a separate calculation for each core state and are often formulated in a spin-incomplete way. These features complicate evaluation of the spectra and, especially, transition properties. (One possible way to circumvent this problem is via the static exchange approach.[26])

The experimentation with low-order wave function methods has been disappointing. The application of the CIS(D) method (configuration interaction singles with perturbative account of double excitations) gives unsystematic errors (MAE of 2.6 eV, with errors as large as 8 eV), especially for Rydberg states, whose energies are considerably underestimated.[7] The resulting spectra are, therefore, qualitatively incorrect. The situation is somewhat improved by using the scaled-opposite-spin version (MAE of 1-2 eV). Application of coupled-cluster methods to core states[11–17,27] has shown promise, but also revealed some challenges, as elaborated below.

Let us examine the essential features of core-ionized and core-excited states. First, these states have open-shell character.[28] Thus, similar to their valence counterparts, application of ground-state methods [such as HF, MP2, CCSD (coupled-cluster with single and double substitutions), or Kohn-Sham DFT] leads to spin-contaminated solutions. Spin-contamination is relatively moderate in the case of

core-ionized states, which can be well represented by a single-determinantal ansatz. More serious issues can arise in molecules with symmetry-equivalent atoms due to symmetry breaking. In the case of open-shell excited states, at least two determinants are needed for spin-completeness; consequently, a single-determinant ansatz leads to severe spin-contamination.[28]

Second, the core-level states lie very high in energy. For example, the IEs of $1s$ electrons in C, N, and O are around 300 eV, 400 eV, and 540 eV, respectively. Thus, in order to find the corresponding solutions, one needs to modify the SCF or iterative diagonalization algorithms. In the SCF procedure, constrained optimization is possible in which the algorithm locks the desired orbital occupation;[8,25] one robust algorithm of this type is called the maximum overlap method (MOM).[25] While at the SCF level these algorithms perform well, using these high-energy solutions in post-HF calculations (e.g., CCSD) often leads to numerically unstable behavior, i.e., CC ansatz would attempt to rotate the orbitals and find a lower-energy solution.[29] In the methods involving diagonalization of an effective Hamiltonian [TDDFT, CIS, equation-of-motion coupled-cluster (EOM-CC), and algebraic diagrammatic construction (ADC)], the Davidson procedure can be modified to solve for the eigenstates dominated by the desired transition (MOM-like) or lying within the desired energy range.[14,15] Even more drastic reduction of the target space can be achieved by employing the core-valence separation (CVS) approximation[30] in which pure valence excitations are filtered out.[9,10,16]

Third, these high-lying states are metastable with respect to electron ejection.[31–34] More specifically, they are Feshbach resonances that can autoionize via two-electron transition in which one valence electron fills the core hole and the second valence electron is ejected. Thus, they are embedded in ionization continuum and their description within Hermitian quantum mechanics is problematic. Owing to their metastable nature, an attempt to compute such states often produces pseudo-continuum states in which one electron occupies the most diffuse orbital.[35] In the Hermitian quantum mechanics, the resonances are not represented by a single state, but rather by an increased density of states in the continuum;[36] thus, the representation of resonances in a discretized continuum is inherently prone to numeric instabilities.

Fourth, orbital relaxation effects are much more important for core states than for valence states because the outer orbitals are more delocalized and better shielded from the nuclear charge than the tight and localized core orbitals. This means that removal of an electron from a valence orbital does not perturb other orbitals to the same extent as removing an electron from a core orbital.

Relativity has the effect of lowering the energy of core orbitals, but for the second row elements the effect is small (Evangelista[8] and Belsey[7] reported corrections of 0.1, 0.2, and 0.3 eV for C, N, and O, respectively). Here we focus on light atoms and, therefore, we are not concerned about relativistic effects.

In this paper, we report a reduced-cost extension of EOM-CC theory to core-level states. Our aim here is to develop an economical and a robust approach that can be applied for modeling core states in moderate-size molecules in condensed phase. Since many experiments are conducted in the condensed phase, it is important to include the effect of the environment on the electronic structure. Due to different electronic distributions, the solvent leads to preferential stabilization of some electronic states relative to others, which results in solvent-induced shifts of transition energies. Shifts in excitation energies can be as large as 1 eV.[37,38] The effect on ionization/electron attachment energies is even more pronounced—shifts of several electron volts in polar solvents and proteins are rather common.[39–41] For valence-ionized and excited states, the combination of EOM-CC with electrostatic embedding models,[42] such as QM/MM (quantum mechanics/molecular mechanics)[43] or QM/EFP (Effective Fragment Potential),[44] leads to accurate results.[39–41,45–47] Unfortunately, using these explicit solvent models requires an extensive sampling of equilibrium solvent configurations (e.g., at least 20-50 snapshots needed to be included in averaging for the converged results[48]), which makes the requirements for the cost and the robustness of the underlying QM method even more stringent. Interestingly, despite the localized nature of the core states, the effect of the solvent is critically important for interpreting the experimental spectra.[49]

Below we present a low-order approximation to EOM-CCSD that reduces the cost of the calculations and, more importantly, eliminates any convergence issues arising due to the metastable nature of the core states. The errors in absolute transition energies are relatively large but systematic. Our benchmarks illustrate that chemical shifts, i.e., changes in core IEs due to the local chemical environment, can be reliably reproduced. The resulting method is inexpensive and robust; thus, it can be easily combined with explicit solvent models. By considering different protonation states of aqueous glycine, we show that our approach can accurately account for solvent effects and reproduce experimental chemical shifts.

II. THEORETICAL APPROACH

The EOM-CC family of methods enables a robust and reliable description of various electronic states including ionized, electron-attached, and electronically excited states.[50–56] The target states are described by using the following ansatz:

$$\Psi = \hat{R}e^{\hat{T}}\Phi_0, \tag{1}$$

where Φ_0 is the reference determinant, operator $\hat{R}$ is a general excitation operator, and $\hat{T}$ is the coupled-cluster operator for the reference state. As a less computationally expensive alternative to full EOM-CCSD, T_2 amplitudes from an MP2 calculation can be used to construct $\bar{H}$, giving rise to the EOM-MP2 family of methods.[57–60] The amplitudes of the operator $\hat{R}$ are found by solving the following eigen-problem:

$$\bar{H}R^k = \Omega_k R^k, \tag{2}$$

where $\bar{H} = e^{-T}He^{T}$. Here we consider the variant of the theory that includes single and double excitations (EOM-CCSD). In this case, the operator $\hat{T}$ includes hole-particle

(1h1p) and two-hole-two-particle (2h2p) configurations and satisfies the CCSD equations for the reference state. The form of EOM operators $\hat{R}$ depends on the nature of the target state. To describe an open-shell ionized state, the operator $\hat{R}$ includes single (1h) and double (2h1p) ionizing configurations,

$$\hat{R}^{IP} = \sum_i r_i i + \frac{1}{2} \sum_{ija} r_{ij}^a a^\dagger j i, \tag{3}$$

giving rise to the EOM-IP-CCSD method. Electronically excited states can be described by using electron-conserving operator $\hat{R}$,

$$\hat{R}^{EE} = r_0 + \sum_{ia} r_i^a a^\dagger i + \frac{1}{4} \sum_{ijab} r_{ij}^{ab} a^\dagger b^\dagger j i, \tag{4}$$

giving rise to the EOM-EE-CCSD method. Here $a^\dagger, b^\dagger, \ldots$ denote creation operators and $i, j, \ldots$ denote annihilation operators. The reference state in EOM-EE and EOM-IP calculations of the ionized and excited states corresponds to the closed-shell ground state of the system (see Fig. 1). The reference determines the separation of orbital space into the occupied and virtual subspaces. Here we use indices $i, j, k, \ldots$ and $a, b, c, \ldots$ to denote the orbitals from the two subspaces.

Because of linear parameterization of the target states, EOM-CC methods are capable of describing multiple electronic states in a balanced way. For valence states dominated by one-electron transitions, such as Koopmans-like ionized states or singly excited states, the errors of EOM-CCSD methods are less than 0.3 eV. The errors in the relative energies of the target states are usually smaller. The EOM-CC ansatz is capable of describing states of different nature (e.g., valence and Rydberg) as well as of mixed character on the same footing. Importantly, EOM-CC provides a consistent description of ionized and excited states, i.e., the onsets of ionization continua in EOM-EE calculations correspond to the EOM-IP IEs. Thus, Rydberg series in EOM-EE calculations converge to the correct correlated ionization thresholds. The single excitation part of the operator $\hat{R}$ (1h or 1h1p) describes leading configurations of the target state, whereas the doubly excited part (2h1p and 2h2p) describes correlation and orbital relaxation effects of the target sates. The similarity transformation of the Hamiltonian wraps in electronic correlation and insures size extensivity, which leads to improved accuracy of EOM-CC (and EOM-MP2) relative to the respective CI counterparts (in which the target states are described by the same linear ansatz, but the operator $\hat{T}$ is zero).

Can one extend EOM-CCSD methodology to describe core-ionized and core-excited states? The algorithm for solving eigen-problem (2) can be trivially extended[14,15] to look for the eigenstates of desired character [e.g., 1s(N) hole or 1s(N)-π^* excitation] or states lying within a specified energy range. In some cases, the equations converge smoothly to the desired states yielding reasonable energies.[14] However, applications to a larger set of molecules revealed that, more often than not, the EOM-CC equations fail to converge. Problematic convergence (100 Davidson iterations) can be seen in the reported data in Ref. 15; it was also mentioned in Ref. 16.

The analysis of the EOM-CC wave functions in the cases of problematic convergence allows us to attribute this behavior to the resonance nature of these states.[31–34] That is, in EOM-IP calculations of a core-ionized state, the amplitudes of 2h1p operator keep increasing, in an attempt to converge to a detached state (see Fig. 2). This behavior is akin to the well-documented case of shape resonances, metastable states that can decay via one-electron process (see, for example, Fig. 2 in Ref. 61). Upon increasing one-electron basis set, the resonances dissolve in the discretized continuum because in the Hermitian quantum mechanics, the resonances correspond to an increased density of states in the continuum.[36] Thus, attempts to compute such states as discrete states using L^2-integrable methods are inherently unreliable because the representation of the resonance depends on the discretization of the continuum in the relevant energy region. That is, in one basis set, which happens to have a relatively sparse spectrum around the resonance, the resonance may appear as an isolated state, whereas in a slightly different basis, which happens to have a slightly denser spectrum in that energy range, the resonance dissolves in the (pseudo)continuum. Only by invoking special techniques, such as complex-variable extensions of standard approaches, these states can be reliably computed as eigenstates of a modified non-Hermitian Hamiltonian.[31,32,34] EOM-CC methods have recently been extended to describe resonances by using complex-scaled and CAP-augmented
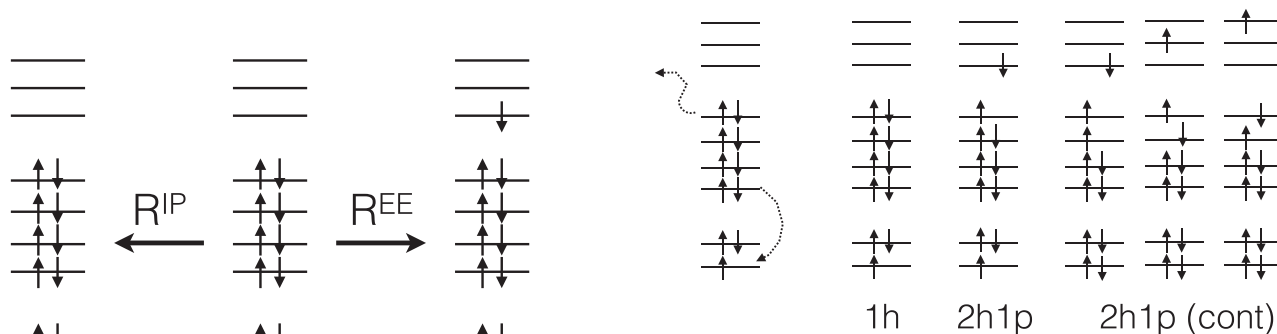
FIG. 1. Target core-ionized (left) and core-excited (right) states are generated by the EOM-IP and EOM-EE R_1-operators from a closed-shell reference of the neutral.
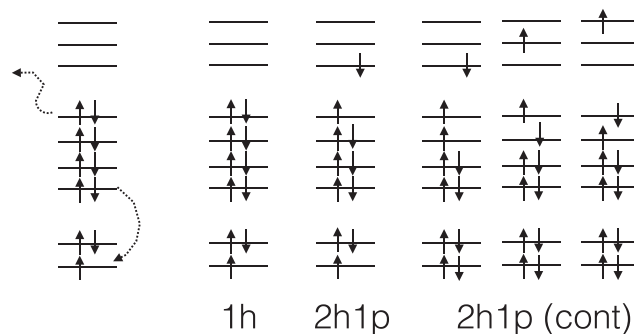
FIG. 2. Core-ionized states are Feshbach resonances. They can autoionize via a two-electron process in which one electron fills the core hole providing enough energy to eject the second electron. To describe this process, the wavefunction ansatz needs to include 2h1p excitations, such as the three right-most configurations.

approaches.[61–64] Such calculations of resonances are rigorous and provide not only energies of the resonances but also their lifetimes. However, such calculations are much more expensive than regular bound-state calculations. Thus, our aim here is to introduce a less expensive approximate method for finding energies of core-level states. From the methodological point of view, there is an important difference between the two types of resonances. Since the shape resonances can decay via one-electron detachment, their metastable nature manifest itself already at the CIS level (as in the example from Ref. [61]). Thus, the attempt to compute these states in a reasonably large basis set will always be complicated by their mixing with discretized pseudo-continuum states. In contrast, the Feshbach resonances can only couple to the continuum by two-electron excitations. Thus, at the CIS level, they behave as bound states. Here we exploit this difference as follows.

Our approach is based on an approximation to the EOM-CCSD eigen-problem, which is similar to the CC2 or CIS(D) approach. We consider EOM-CCSD (or EOM-MP2) $\bar{H}$ as our full Hamiltonian. We then introduce zero-order wave functions, which contain only $1h$ and $1h1p$ configurations. The amplitudes of the respective R_1 operators are found by diagonalizing the singles block of $\bar{H}$. We then employ the 2nd order Rayleigh-Schrödinger perturbation theory to evaluate energy correction due to double excitations. We follow the same formalism as in Refs. [65] and [66]. The programmable expressions for zero-order states and energy correction can be easily derived from the EOM-CCSD expressions for σ-vectors;[67,68] they are given in the supplementary material. The resulting methods can be called EOM-IP/EE-CCSD-S(D). A similar approach can be applied to the EOM-MP2 ansatz giving rise to the EOM-IP/EE-MP2-S(D) methods. We note that since double excitations are described perturbatively, the most expensive EOM step is non-iterative. Thus, the scaling of the EOM-IP-MP2-S(D) calculation is N^5 and the N^5 step is non-iterative, which greatly reduces storage requirements. We note that the CVS approximation[9,10,16] effectively mitigates the problem of coupling with the continuum states by removing pure valence excitations (such as the three rightmost configurations in Fig. [2]) and, therefore, reducing the density of states in the continuum. However, straightforward implementation[16] of CVS using a projector operator that zeros out parts of the EOM amplitudes that do not involve core electrons has the same computational scaling and cost as regular EOM-CCSD.

Below we illustrate the numeric performance of our approximate models for valence and core states by considering a set of small molecules. We then show that when combined with explicit solvent models, these models can accurately reproduce experimental chemical shifts for core-ionized states. We note that our approach for core states is similar, in some aspects, to the methodology developed by Ghosh[60] for valence states.

III. MODELING SOLVENT EFFECTS FOR CORE-LEVEL STATES

Including solvent effects is critically important for interpreting experimental spectra. For example, one exciting application of core-level spectra is for distinguishing between different protonation states of amino-acids in realistic environments (protein, solution).[49,69] As was illustrated by Winter and co-workers,[49] although the protonation can change gas-phase IEs by several eV (e.g., 1-9 eV for glycine), the solvent screening results in much smaller chemical shifts (0.2-2 eV). Calculations on model systems[49] have shown that the implicit solvent models capture the effect qualitatively, but they are not sufficient for quantitative predictions of chemical shifts in solutions. Here we systematically investigate the effect of the solvent on core IEs of glycine using the following approaches:

1. Non-equilibrium polarizable continuum model[70] (PCM). We use the zero-order approach in which no state-specific solvent response is included (see Refs. [71] and [72] for detailed benchmarks of different versions of the PCM using ADC wave functions). Following Ref. [49], we consider two sets of models: bare glycine with PCM solvent and an optimized glycine molecule clustered with six explicit water molecules embedded in the PCM solvent (note that the structure of the cluster is static and no equilibrium averaging is performed in this calculation).

2. Explicit solvent represented by either point charges (QM/MM) or by EFP (QM/EFP). In these approaches, the spectra are computed by averaging IEs over snapshots obtained from molecular dynamics (MD) simulations. For valence states, such protocols have been shown to yield very accurate spectroscopic and thermochemical results.[39,41,48,73,74] We compare two approaches: (i) only glycine is included in the QM part and (ii) the QM part comprises glycine and several explicit water molecules. The structures of the QM and MM parts are taken from the MD snapshots.

    In contrast to QM/MM, EFP is a polarizable embedding method in which the electrostatic field of the EFP solute changes in response to changes in charge distribution of the QM solute. EFP[45,47,75–78] is similar to the PE (polarizable embedding) approach.[79–82]

3. To better understand different contributions to solvent shifts, we also considered model structures taken from representative snapshots from the MD trajectory. For these snapshots, we compare the magnitude of shifts computed by various solvent models without performing equilibrium averaging.

IV. COMPUTATIONAL DETAILS

All electronic structure and QM/MM calculations were performed using the Q-Chem[83,84] electronic structure package. QM/EFP calculations were performed with libefp.[85,86] MD simulations were performed using the NAMD package.[87]

A. Benchmark calculations

As a benchmark set, we considered several small molecules for which gas-phase core states are available. The structures for N_2, CO, and C_2H_2 are taken from Ref. [66]; all other structures were optimized with RI-MP2/cc-pVTZ. All

Cartesian geometries are given in the supplementary material. Several Dunning basis sets were considered: cc-pVTZ, aug-cc-pVTZ, and cc-pCVTZ. We found that the cc-pVTZ basis is sufficient for core-ionized states. Using cc-pCVTZ did not result in better performance.

For comparison purposes, we also report ΔSCF calculations with DFT using the ωB97X-D functional[88,89] and the cc-pVTZ basis set. In these calculations, core IEs are computed as total energy differences between the neutral and core-ionized systems. The SCF solutions for core-ionized states are found using the MOM algorithm,[25] as implemented in Q-Chem.[83]

B. Calculations of different forms of aqueous glycine

To carry out equilibrium sampling of solvated glycine, we performed MD simulations using the NAMD software[87] with CHARMM forcefield,[90] using an NPT ensemble (T = 300 K, P = 1 atm, density equals 1.0 g/ml^3) with periodic boundary conditions, rigid TIP3P water molecules, and a simulation time step of 2 fs. Prior to running production trajectories, the water box was allowed to equilibrate for 2 ns. We first performed nanosecond relaxation of the solvent box with solute frozen and then allowed the full system to equilibrate. After that, we performed production run of 500 ps, taking the snapshots every 5 ps.

In the simulation of the zwitterionic form of glycine, the box contained one glycine molecule dissolved in 1486 water molecules. In the case of the deprotonated form, the box contained single deprotonated glycine dissolved in 1828 water molecules and a sodium atom added to neutralize the charge of the system. The parameters for the deprotonated glycine N-terminus atom were taken from the deprotonated form of lysine residue. The N(NH$_2$-terminal)–C(CH$_2$)–C(COO-terminal) angle and dihedral C(COO-terminal)–C(CH$_2$)–N(NH$_2$-terminal)–H(CH$_2$) angle were treated using the parameters from the corresponding angles of the neutral (canonical) form of glycine. In the simulation of the protonated form of glycine, the simulation box contained one glycine molecule dissolved in 1877 water molecules and a single chloride ion added to neutralize the charge of the system. In the MD simulation, the charge of C in the CH$_2$ group was adjusted to balance the total charge of the molecule. The size of the water box was 35-40 Å, which was shown[39,91] to be sufficient for obtaining converged valence IEs.

For the definition of the first solvation shell and for the hydrogen-bond analysis, we employed the following constraints. We defined the cutoff for the first solvation shell at 3.0 Å: if any atom of a water molecule was within this distance from either oxygen or nitrogen, this water was assigned to the solvation shell. The average number of water molecules assigned to the first solvation shell using 3 Å cutoff was 8.2 for the zwitterionic form, 7.2 for the anionic form, and 5.1 for the cationic form. The respective distributions are shown in Fig. S3 of the supplementary material. This definition of the solvation shell is used in selecting the QM subsystem in the QM/MM calculations. We also tested a shorter cutoff radius (2.7 Å) and found that using a shorter cutoff radius significantly affects the number of molecules assigned to the first solvation shell (see

Fig. S3 in the supplementary material) and leads to large errors in IEs. We defined hydrogen bonds between the solute and the water molecules in the first solvation shell using a bond length of 3.0 Å or less and a hydrogen bond angle of less than 30°. The resulting distribution of the hydrogen bonds is discussed below.

Running averages of the total energy and RMSD of glycine, which are shown in the supplementary material, confirm that the simulation length was sufficient to fully relax both the solvent box and the solute. The radial distribution function of the distance between the solute and the counterion for simulations of protonated and deprotonated states is shown in Fig. S4 of the supplementary material. The solute and counterion are separated by several solvation shells during roughly 99% of the simulation time. Due to the screening effect of water, the resulting effect of the counterion on the electronic structure of the solute is relatively small.

1. Model structures and QM/MM and QM/EFP calculations

As described in Sec. III, we compared several approaches to model solvent effects. The structure of the canonical form of glycine (Gly$_{can}$), which is representative of the gas-phase glycine, was optimized by RI-MP2/cc-pVTZ. All chemical shifts were computed with respect to this structure. The structures representing other forms of glycine (Gly$_{ZI}$, Gly$_+$, and Gly$_-$) were optimized using PCM/ωB97x-D/cc-pVTZ. The respective Cartesian geometries are given in the supplementary material.

To compare with the approach of Winter and co-workers,[49] we computed solvent shifts using the PCM for the above-optimized structures of different forms of glycine (these results are denoted by Gly$_{XX}^{PCM}$) as well as model structures[49] of glycine with six water molecules embedded in the PCM (these calculations are denoted by Gly$_{6w,XX}^{PCM}$).

To compute IEs with explicit solvent models, we followed protocols similar to those developed earlier for valence states.[39,41,48,73,74] Specifically, we performed QM/MM and QM/EFP calculations using two choices of the QM system. In small QM calculations, only glycine was included, similar to our previous calculations of valence IEs;[39,41] these calculations are denoted by Gly$_{XX}$/MM or Gly$_{XX}$/EFP. In large QM calculations, glycine and water molecules from the first solvation shell (defined using a cutoff of 3 Å, as explained above) were included in the QM part. The number of water molecules in these calculations (denoted as Gly$_{XX,w}$/MM or Gly$_{XX,w}$/EFP) varied from 2 to 8; the distributions for different protonation forms are shown in Fig. S3 of the supplementary material. For the EFP calculations, we report IEs computed with and without a solvent polarization response correction;[45] these are denoted by EFP and EFP$_{fz}$, respectively. The QM/MM and QM/EFP IEs were computed by averaging over 100 snapshots taken every 5 ps from 500 ps equilibrium trajectories. We note that prior to the QM/MM and QM/EFP calculations, each snapshot was translated such that glycine is always in the center of the box. For the detailed analysis of different solvent models, we also considered model clusters from selected snapshots from our equilibrium MD simulations (the structures of these snapshots are described below; the

respective Cartesian geometries are given in the supplementary material).

V. RESULTS AND DISCUSSION

A. Benchmark calculations

Before proceeding to calculations of core-level states, we consider the performance of these approximations by looking at selected valence ionized and excited states. Table S1 in the supplementary material shows the results for the valence ionized and lowest excited states of formaldehyde. As one can see, the errors in IEs introduced by the perturbative account of double excitations range between 0.36 and 0.72 eV. The behavior appears to be systematic: the EOM-CCSD-S(D) values are blue-shifted with respect to full EOM-IP-CCSD, which is expected since the reference state is described by full EOM-CCSD and the target states by an approximate S(D) ansatz. We note that the EOM-MP2 approximation to EOM-IP-CCSD performs really well, both in the case of full EOM-IP-MP2 (errors of 0.01-0.04 eV) and in the case of EOM-IP-S(D) (errors of 0.42-0.83 eV). We observe similar errors in ethylene and ethane (Tables S2 and S3 in the supplementary material). Thus, on the basis of these benchmark calculations, the MAEs for EOM-IP-CCSD-S(D) and EOM-IP-MP2-S(D) are 0.23 and 0.51 eV, respectively, with standard deviations 0.23 and 0.18 eV. The errors for excited states are also systematic (energies are blue-shifted), but they are larger, e.g., 1.00-1.31 eV for EOM-EE-CCSD-S(D), suggesting that correlation and orbital relaxation effects are more significant for excited states.

We now proceed to the core-ionized states. The modified Davidson procedure[14] converges very fast for EOM-IP-CCSD-S(D) and EOM-IP-MP2-S(D). The core states have clear Koopmans character. The IEs obtained from diagonalizing the singles-only block of $\bar{H}$ are very close to the respective orbital energies (Koopmans IEs). The magnitude of the (D) correction is large: it varies between 10 and 20 eV. For the full EOM-IP-CCSD/MP2 methods, we were able to obtain results only for a subset of molecules. We begin by analyzing the results for selected small molecules.

Table I shows the results for model ammonia clusters in different protonation states. For this system, we can also compare our results with previously reported ADC(4) calculations.[92] For the cases when converged EOM-IP-CCSD results are available, the errors of IP-CCSD-S(D) range between 0.4 and 0.7 eV (again, the errors are systematic). As one can see, the protonation state and clustering have a noticeable effect on core IEs. These chemical shifts relative to the reference NH_3 system are shown in parentheses; they range from -0.7 to $+11.6$ eV. The errors of approximate S(D) methods in chemical shifts relative to full EOM-IP-CCSD are much smaller: the largest error of IP-CCSD-S(D) is 0.4 eV for NH_4^+, which has a chemical shift of 11.61 eV. Thus, one can expect that due to the systematic nature of the errors, IP-CCSD-S(D) might be able to reliably reproduce chemical shifts. We note that EOM-IP-MP2 versions perform similar to EOM-IP-CCSD; the differences in the computed chemical shifts between the two models are less than 0.2 eV.

Tables S4-S9 in the supplementary material show the results for core-ionized states of selected molecules. Tables S4, S6, and S8 in the supplementary material show absolute values of IE for carbon, oxygen, and nitrogen 1s levels, respectively. The chemical shifts computed against reference molecules are collected in Tables S5, S7, and S9 of the supplementary material. The experimental values were taken from the synchrotron studies[93,94] in the case of methane, ethene, ethyne, and carbon monoxide and from the compilation[95] by Jolly *et al.* for other species. The results in Tables S4-S9 of the supplementary material were obtained with the cc-pVTZ basis. For selected molecules, we also performed cc-pCVTZ and aug-pCVTZ calculations and found that the effect of the basis set on core ionized states is rather small: chemical shifts were affected by less than 0.1 eV and vertical IEs by $\sim+0.5$ eV.

Let us first consider errors in absolute values against the experimental IEs. For a small subset of molecules for which we were able to obtain full EOM-IP-CCSD results, the errors against the experimental values are larger than for valence IEs and for most cases are systematic (IEs are overestimated). The standard deviation (non-systematic errors) is about 0.2 eV. For carbon and oxygen edges in CH_4, C_2H_6, H_2O, and CH_3OH,

TABLE I. Core-ionized states of model ammonia systems.[a] The chemical shifts relative to the reference NH_3 system are shown in parentheses.

Molecule	ADC(4)[b]	IP-CCSD	IP-CCSD-S(D)	IP-MP2-SD	IP-MP2-S(D)
NH_3	406	406.44 (0.00)	406.97 (0.00)	406.87 (0.00)	407.59 (0.00)
NH_4^+	418	417.64 (+11.2)	418.58 (+11.61)	418.04 (+11.17)	419.09 (+11.5)
NH_2^-	395	395.52 (−10.92)	395.89 (−11.08)	396.18 (−10.69)	396.78 (−10.81)
$NH_4^+NH_3$	413.97	414.89 (+8.45)	415.38 (+8.41)	415.33 (+8.46)	415.97 (+8.38)
	411.12	411.97 (+5.53)	412.61 (+5.64)	DNC	413.20 (+5.61)
$(NH_3)_2$	405.72	DNC	406.31 (−0.66)	DNC	406.94 (−0.65)
	404.65	DNC	406.05 (−0.92)	DNC	406.69 (−0.90)
$NH_2^-NH_3$	399.16	DNC	400.15 (−6.82)	DNC	400.80 (−6.79)
	395.23	DNC	396.80 (−10.17)	DNC	397.69 (−9.90)

[a] ri-MP2/cc-pVTZ geometry. All EOM calculations were performed with the cc-pVTZ basis. Calculations marked with DNC did not converge in 60 cycles.
[b] ADC(4) results from Ref. 92. ADC(4) calculations were performed with the DZP basis; the structures were not reported.

oxygen edges in CO, HCOOH, CH$_3$COOH, and CH$_2$O, and nitrogen edges in N$_2$ and CH$_3$NH$_2$, the errors of the full EOM-IP-CCSD method are less than 1 eV. The inspection of the results obtained with EOM-IP-CCSD-S(D) and EOM-IP-MP2-S(D) reveals more substantial errors in the absolute values of IEs. Relative to full EOM-IP-CCSD, the average errors increase by about 1 eV; however, the increase in the standard deviation is less. When comparing with the experimental values, we note that except for CH$_3$OH, the sign of error is always positive, which is similar to the valence states. The errors increase up to 1.5 eV when using the S(D) approximation. We observe larger errors for carbon edges in CO, C$_2$H$_4$, C$_2$H$_2$, CH$_3$OH, CH$_2$O, and CH$_3$COO and nitrogen edge in NH$_3$: errors are less than 1.5 and 3 eV for full IP-CCSD and IP-CCSD-S(D), respectively. MAEs against the experimental values are 1.96 eV and 0.55 eV for C and O edges, respectively. The respective standard deviations are 0.38 and 0.3 eV. For localized core states (i.e., in non-symmetric molecules), ΔSCF methodology reproduces experiment remarkably well because the relaxation effects are described explicitly in these calculations. However, the ΔSCF errors are much larger in C$_2$H$_6$, C$_2$H$_4$, C$_2$H$_2$, and N$_2$, where this single-determinantal approach fails to represent the delocalized hole (this well documented problem can be remedied by, for example, performing small configuration interaction calculation in the basis of the two symmetry-broken configurations[96]).

The chemical shifts relative to the reference systems are given in Tables S5, S7, and S9 of the supplementary material. Here, we computed chemical shifts for carbon 1s, nitrogen 1s, and oxygen 1s edges against the following reference systems: CH$_4$, NH$_3$, and H$_2$O. The errors in chemical shifts against the experimental values are shown in Figs. 3 and 4. An alternative representation, the computed shifts versus the experimental values, is given in Figs. S1 and S2 of the supplementary material. Despite relatively large errors in absolute IEs, chemical shifts are reproduced by EOM-CC rather well. For the C edge, the full EOM-IP-CCSD and the S(D) approximations show

errors of less than 0.5 eV and 1 eV, respectively. The errors are larger for the O edge. The performance of EOM-MP2 variants is very similar to the respective CCSD variants. The ΔSCF approach fails in the same cases as for the absolute energies.

Overall, the results confirm the importance of orbital relaxation and show that perturbative treatment of double excitations in the S(D) approximation leads to larger errors relative to full CCSD. On the basis of excellent agreement between full EOM-CCSD and CVS approximations reported by Coriani and Koch,[16,17] it appears that the relaxation effect is captured extremely well within CVS. For comparison, we also computed core IEs for selected molecules using the frozen valence approximation (in such calculations, all valence occupied orbitals are frozen). These calculations, which describe the relaxation of the core orbitals but neglect the effect of core electrons on the valence states, lead to much larger (10-20 eV) and non-systematic errors in IEs. Importantly, despite relatively large errors for absolute values of IEs, the performance of IP-CCSD/MP2-S(D) for chemical shifts is satisfactory, which suggest that our approximate methods might be useful for studying core-ionization spectra in large molecules and complex environments. We anticipate even smaller errors in applications where chemical shifts are computed within a similar class of compounds, such as in different protonation states of amino-acids. Our results for glycine presented below confirm this.

Preliminary investigations of the core-excited states have shown larger errors for core-excited states than for core-ionized states, which is consistent with the observations for valence states. These results will be discussed elsewhere.

B. Core-ionized states of different protonation forms of glycine

Figure 5 shows structures of glycine in different protonation states. In the gas phase, glycine exists in the canoni-
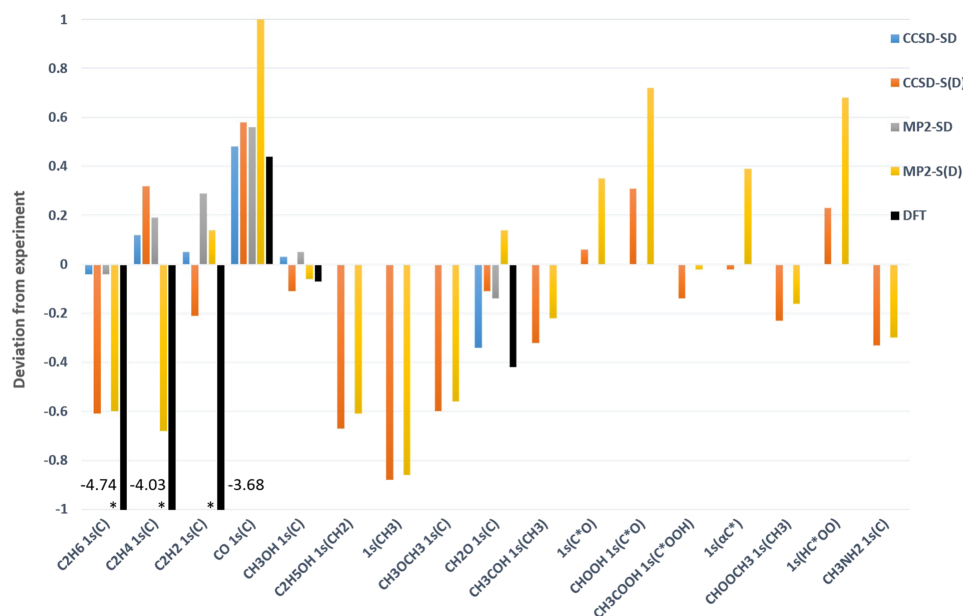

FIG. 3. Errors in chemical shifts (eV) of carbon (1s) ionized states for selected molecules. Shifts are computed relative to methane. Stars mark errors that are too large and do not fit on the plot (the corresponding values are shown next to the bars).
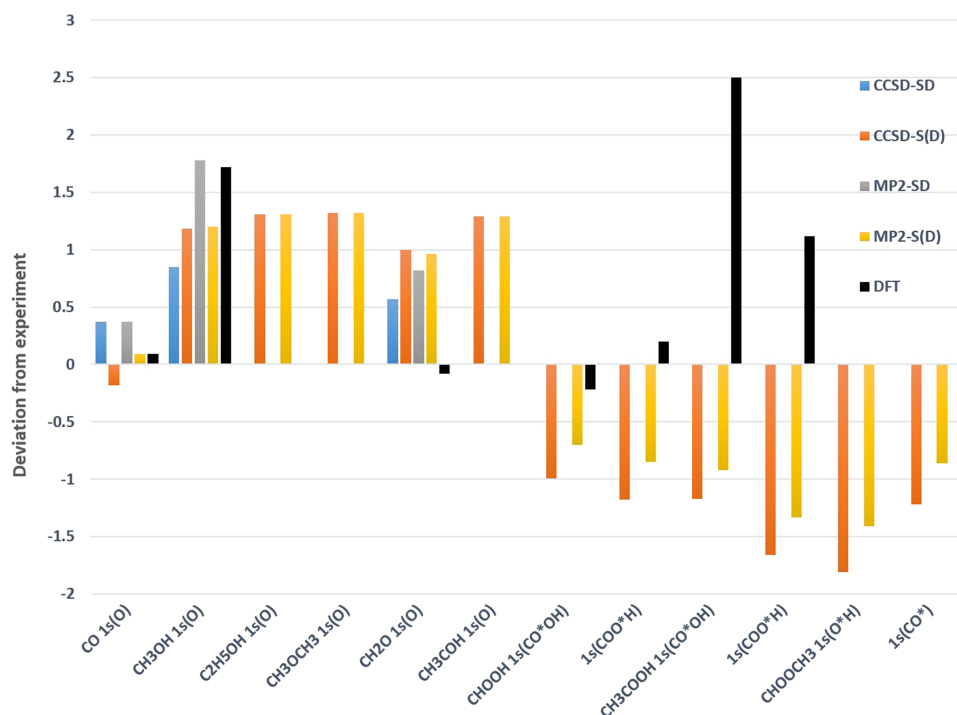
FIG. 4. Errors in chemical shifts (eV) of oxygen (1s) ionized states for selected molecules. Shifts are computed relative to water.

cal form (Gly$_{can}$), but in solution it assumes the zwitterionic form (Gly$_{ZI}$) in which a proton is transferred from the carboxyl group to the amino group. At low pH, both the carboxyl and the amino groups are protonated, giving rise to the cationic form (Gly$_+$), and at high pH, it exists in the deprotonated, anionic form (Gly$_-$). Table II presents experimental core IEs of glycine measured in the gas phase[97] and in liquid-jet synchrotron experiments.[49] Table II also shows chemical shifts of core IEs computed as the difference between condensed-phase IEs and gas-phase IEs of the canonical form: positive chemical shifts correspond to
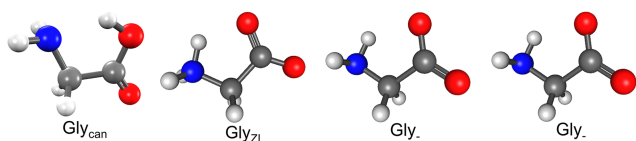

FIG. 5. Structures of different protonation forms of glycine: canonical (Gly$_{can}$), zwitterionic (Gly$_{ZI}$), deprotonated (Gly$_-$), and protonated (Gly$_+$).

TABLE II. Experimental core IEs and chemical shifts against gas-phase values of different forms of glycine (in eV).

Core IEs	Gly$_{can}$[a]	Gly$_{ZI}$[b]	Gly$_-$[b]	Gly$_+$[b]
1s (N)	405.40	406.81	404.30	406.91
1s (COO)	295.20	293.55	293.22	294.53
1s (CH$_2$)	292.30	291.43	290.67	291.88
		Chemical shifts		
1s (N)		1.18	−1.10	1.51
1s (COO)		−1.65	−1.98	−0.67
1s (CH$_2$)		−0.87	−1.63	−0.42

[a]From Ref. 97.
[b]From Ref. 49.

blue-shifted IEs and negative chemical shifts correspond to red-shifted IEs. As one can see, changes in protonation states have distinct spectroscopic signatures. The main trends can be readily explained by electrostatics: the protonation of the amino group in Gly$_{ZI}$ and Gly$_+$ results in the higher IE of N(1s) (positive chemical shift), whereas the deprotonation of the carboxyl group in Gly$_{ZI}$ and Gly$_+$ results in the red-shifted carbon edge.

Table III shows IEs of the canonical form and chemical shifts against Gly$_{can}$ of other protonation forms of glycine computed with IP-CCSD-S(D)/cc-pVTZ. The results with IP-MP2-S(D) and cc-pVDZ are given in the supplementary material (Table S10). Consistent with the benchmark calculations, we observe errors of 1-2 eV in the IEs of the canonical form. The computed chemical shifts for other protonation forms follow general trends expected from the electrostatic considerations, but their magnitudes are grossly exaggerated (by up to 7 eV). Moreover, for some IEs even the sign of the shift is reversed relative to the experiment. The computed shifts and large differences from the experimental values are in line with the results of DFT calculations from Ref. 49. On the basis of PCM calculations and calculations for micro-solvated model clusters, the authors have attributed the large discrepancy between the gas-phase calculations and the measured aqueous IEs to solvent effects. The solvent effectively screens the charges thus resulting

TABLE III. Computed gas-phase IEs and chemical shifts against Gly$_{can}$ of different forms of glycine (in eV). EOM-IP-CCSD-S(D)/cc-pVTZ.

	Gly$_{can}$	Gly$_{ZI}$	Gly$_-$	Gly$_+$
1s (N)	406.58	4.06	−6.32	9.09
1s (COO)	296.91	−1.43	−6.84	6.25
1s (CH$_2$)	294.70	0.58	−6.18	6.30

in much smaller shifts due to protonation/deprotonation. Table S11 in the supplementary material shows calculated IEs and chemical shifts computed with the IP-CCSD-S(D) and PCM. Following Ref. 49, we compare the results of bare glycine embedded in the PCM versus model clusters (glycine with six water molecules) surrounded by the PCM solvent (we use the same structures as in Ref. 49). Similar to the DFT-based calculations,[49] we observe that the PCM greatly reduces the magnitude of the shifts, bringing them to a better (but not perfect) agreement with the experiment. The authors of Ref. 49 reported better agreement when six explicit water molecules were included in the PCM calculations. Our results from Table S11 in the supplementary material support this finding. A significant effect due to the inclusion of the explicit water molecules can be explained by the inability of PCMs to describe specific solvent-solute interactions such as hydrogen bonding. While this effect is captured by considering small optimized clusters (such as various Gly_{6w} structures from Ref. 49), these calculations do not account for the fluctuation of structures due to equilibrium thermal motions. More appropriate calculations should include averaging over snapshots from equilibrium MD simulations. Our results below illustrate that this approach results in much better agreement with the experiment.

1. Analysis of hydrogen-bonding interactions of different protonation forms of glycine

The distributions of hydrogen bonds between the solvated glycine and nearby water molecules along the equilibrium MD trajectory are shown in Fig. 6. In the zwitterionic form, in which both protic groups are charged, we observe an average of about 2 hydrogen bonds formed by the amino group. The number of hydrogen bonds formed by the carboxyl group, which has two O^- centers, is four. In the deprotonated form, the amino group is no longer charged resulting in the smaller average number of hydrogen bonds (one), but the number of hydrogen bonds formed by the carboxyl group increases from four to five. In the protonated form, the carboxyl group is neutral and forms only two hydrogen bonds (the number of hydrogen bonds formed by the amino group is the same as in the zwitterionic form). As illustrated by the calculations below, these different hydrogen bond patterns are partially responsible for the observed trends in the core IEs. If a group acts as a hydrogen-bond donor (which is the case for the carboxyl group in both forms and the deprotonated amino group), one should expect a blue shift in IEs, as the electron density on the donating group is depleted. Conversely, if a group acts as a hydrogen-bond acceptor (which is the case for the protonated form of the carboxyl group and both forms of the neutral and protonated amino groups), one should expect red-shifted IEs.

2. Comparison of different solvent treatment by using a single snapshot from the equilibrium MD trajectory

Before proceeding to computing averages using equilibrium MD trajectories, we analyze the effects of different solvent treatment using a single representative snapshot for each structure. We selected snapshots that have the number of hydrogen bonds equal to the average number of hydrogen
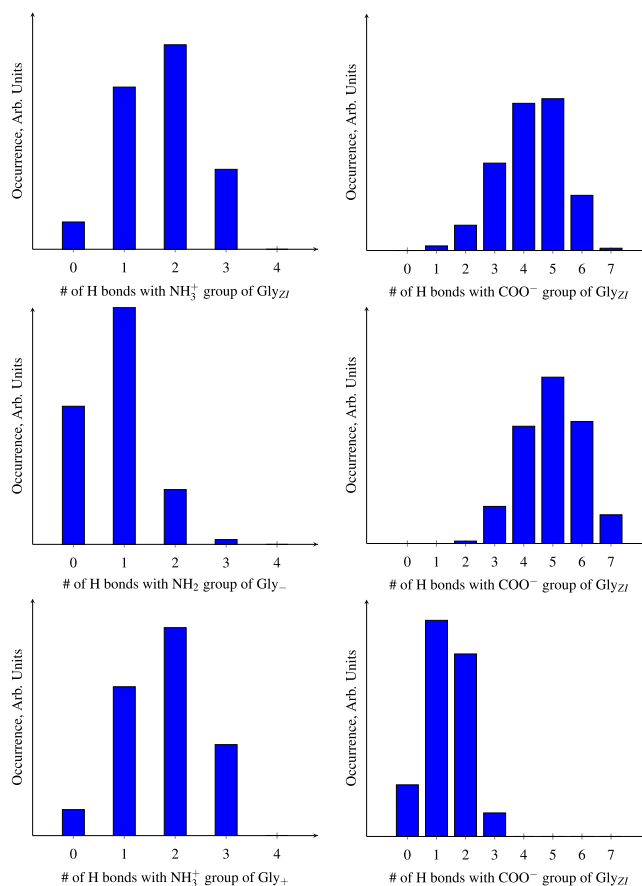

FIG. 6. Distribution of hydrogen bonds between the solute and water molecules from the first solvation shell for different forms of glycine.

bonds for each type of structure (see Fig. 6). The structures of glycine and the nearest solvent molecules from each snapshot are shown in Fig. 7.

We compare QM/MM and QM/EFP treatment using two different setups for the QM part: bare glycine and glycine with the nearest hydrogen-bonding water molecules. The results of IP-CCSD-S(D) calculations are collected in Table IV. In this set of calculations, $QM_{XX,w}$/EFP represents the highest level of theory (in this calculation, the nearest hydrogen-bonded water molecules are included into the QM part and the rest of water molecules are described by EFP including state-specific solvent polarization correction). We observe that even when the nearest water molecules are included into the QM part, polarization correction is important and can contribute up to 0.7 eV to the chemical shift of core IEs. The differences between the $QM_{XX,w}$/MM and $QM_{XX,w}$/EFP shifts can be as large as 0.5 eV, which is substantial, given the magnitude of the shifts. Contrary to our calculations of valence states,[39,41] we observe noticeable differences between EFP
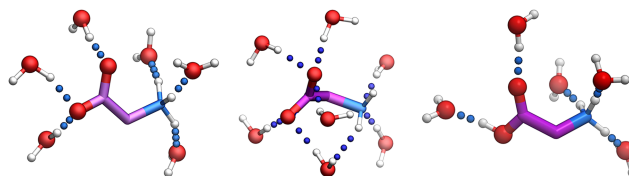

FIG. 7. Representative snapshots for Gly_{ZI}, Gly_-, and Gly_+ from equilibrium MD simulation.

TABLE IV. Core IEs and chemical shifts (eV) relative to Gly$_{can}$ of different protonation forms of glycine computed using representative single snapshots from the equilibrium trajectory. IP-CCSD-S(D)/cc-pVTZ. EFP$_{fz}$ denotes EFP calculations without solvent polarization correction.

	QM$_{ZI}$/MM	QM$_{ZI}$/EFP$_{fz}$	QM$_{ZI}$/EFP	QM$_{ZI,w}$/MM	QM$_{ZI,w}$/EFP$_{fz}$	QM$_{ZI,w}$/EFP
1s N	2.39	1.90	1.14	1.26	1.11	1.33
1s C(COO)	−1.02	0.12	−0.55	−1.69	−1.92	−1.22
1s C(CH$_2$)	0.01	0.77	0.21	−0.65	−0.21	−0.93
	QM$_-$/MM	QM$_-$/EFP$_{fz}$	QM$_-$/EFP	QM$_{-,w}$/MM	QM$_{-,w}$/EFP$_{fz}$	QM$_{-,w}$/EFP
1s N	−0.29	−0.79	−0.44	−0.54	−0.21	−0.76
1s C(COO)	−0.63	−0.48	−0.99	−1.25	−1.25	−1.72
1s C(CH$_2$)	−0.36	−0.22	−0.76	−0.68	−0.33	−0.98
	QM$_+$/MM	QM$_+$/EFP$_{fz}$	QM$_+$/EFP	QM$_{+,w}$/MM	QM$_{+,w}$/EFP$_{fz}$	QM$_{+,w}$/EFP
1s N	3.61	0.64	0.89	1.69	0.95	1.16
1s C(COO)	1.26	−0.02	−0.26	−0.23	−0.47	−0.47
1s C(CH$_2$)	1.21	−0.66	−1.13	−0.41	−0.75	−0.21

calculations with small and extended QM, i.e., the largest observed difference between QM$_{XX}$/EFP and QM$_{XX,w}$/EFP is 0.7 eV. Even larger differences (up to 1.1 eV) are observed between QM$_{XX}$/MM and QM$_{XX,w}$/MM. Higher sensitivity of QM/MM compared to QM/EFP can be explained by the lack of polarization in the former, where the electron distribution in the nearby water molecules cannot adjust to the charges in QM. Along the same lines, we observe larger differences between small and extended QM parts for EFP calculations without polarization correction. In summary, the results in Table IV indicate high sensitivity of the core IEs to the solvent-specific interactions and highlight the importance of using polarizable solvent models, such as EFP. But even with EFP, the inclusion of hydrogen-bonded water molecules into the QM part is necessary for accurate results.

In order to reduce computational costs when performing equilibrium averaging, we compared the magnitude of chemical shifts computed with EOM-IP-CCSD-S(D) and EOM-IP-MP2-S(D). We also quantified the effect of using a smaller basis set, cc-pVDZ, instead of cc-pVTZ. The results are given in Table S12 of the supplementary material. Contrary to the strong dependence of the shifts on the solvent model, we observe that chemical shifts are rather insensitive to the basis set and to using MP2 rather than CCSD amplitudes. Among all solvent models, the largest difference between EOM-IP-CCSD/cc-pVTZ and EOM-IP-MP2/cc-pVDZ calculations is ∼0.1 eV. For our best model, QM$_{XX,6w}$/EFP, the differences for most states are 0.01-0.05 eV, and the largest difference of 0.1 eV is observed for the state with the largest chemical shift, C(1s) from the carboxyl group. Thus, in equilibrium sampling calculations, one can safely employ EOM-IP-MP2-S(D)/cc-pVDZ. On our machine, the EOM-IP-MP2-S(D)/cc-pVDZ calculation with large QM is an order of magnitude faster than EOM-IP-CCSD-S(D)/cc-pVTZ.

C. Core IEs of different protonation forms of glycine computed using equilibrium averaging with explicit solvent models

Table V and Fig. 8 show chemical shifts of different protonation states of glycine computed by averaging over

TABLE V. Chemical shifts (eV) relative to Gly$_{can}$ of different protonation forms of glycine.[a]

	QM$_{ZI}$/MM	QM$_{ZI}$/EFP$_{fz}$	QM$_{ZI}$/EFP	QM$_{ZI,w}$/MM	QM$_{ZI,w}$/EFP$_{fz}$	QM$_{ZI,w}$/EFP	Exp$_{ZI,aq}$
1s N	2.96	2.74	2.02	1.73	1.78	1.52	1.41
1s C(COO)	−0.19	1.11	−0.53	−0.99	−0.93	−1.43	−1.65
1s C(CH$_2$)	0.48	1.00	0.50	−0.39	−0.22	−0.75	−0.87
	QM$_-$/MM	QM$_-$/EFP$_{fz}$	QM$_-$/EFP	QM$_{-,w}$/MM	QM$_{-,w}$/EFP$_{fz}$	QM$_{-,w}$/EFP	Exp$_{-,aq}$
1s N	−0.09	0.08	−0.20	−1.05	−0.83	−1.29	−1.10
1s C(COO)	−0.45	−0.16	−0.82	−1.48	−1.38	−1.89	−1.98
1s C(CH$_2$)	−0.52	0.34	−0.29	−1.72	−1.53	−1.69	−1.63
	QM$_+$/MM	QM$_+$/EFP$_{fz}$	QM$_+$/EFP	QM$_{+,w}$/MM	QM$_{+,w}$/EFP$_{fz}$	QM$_{+,w}$/EFP	Exp$_{+,aq}$
1s N	3.85	1.89	1.10	1.47	1.30	1.60	1.51
1s C(COO)	1.59	0.34	−0.09	−1.06	−1.12	−0.82	−0.67
1s C(CH$_2$)	1.58	0.01	−0.44	−0.30	−0.05	−0.47	−0.42

[a]The results with small QM were computed with IP-CCSD-S(D)/cc-pVTZ. The calculations with large QM part, QM$_{XX,w}$/MM and QM$_{XX,w}$/EFP, were computed with IP-MP2-S(D)/cc-pVDZ.
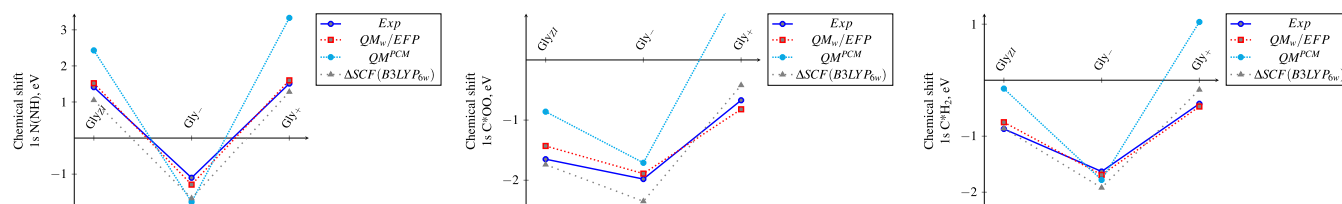
FIG. 8. Chemical shifts of glycine computed with different levels of theory: nitrogen 1s edge, carbon (1s) from carboxyl, and carbon (1s) from methyl group.

snapshots from the equilibrium MD trajectory. We observe that $QM_{ZI,w}$/EFP results are in excellent agreement with the experiment, i.e., the signs and the relative values of the shifts are correctly reproduced and the errors in shifts range from 0.05 to 0.22 eV (again, the largest error is observed for the largest-magnitude shift). By inspecting the data in Table V, one can clearly see that this agreement is not coincidental, as the differences between other lower-level treatments are larger. This observation is in agreement with the behavior observed for the representative snapshots. For example, the errors of EFP calculations with small QM, which does not include the nearest water molecules, can be as large as 1 eV and the largest error of $QM_{XX,w}$/MM is ~0.6 eV.

VI. CONCLUSION

We presented a new computational approach based on the EOM-CC theory for modeling core-ionized states. By using an approximate EOM-CC ansatz in which doubly excited amplitudes are treated perturbatively, we circumvent the convergence problems arising due to metastable (resonance) character of the core-level states. The benchmark calculations for small gas-phase molecules show that this approximation leads to errors of about 1-2 eV in core-ionized states; however, chemical shifts computed relative to a reference compound can be computed with higher accuracy. We applied this method, in combination with different solvent models, to model chemical shifts in core IEs of different protonation forms of glycine. We showed that explicit solvent treatment by the EFP method (with solvent polarization correction) leads to the best results and that, even with EFP, the nearest water molecules should be included in the QM part to achieve quantitative agreement with the experimental shifts. These results once again highlight the high sensitivity of core states to the local environment and specific interactions with the surrounding solvent. Our best protocol, QM/EFP calculation with the extended QM part including glycine and the nearest water molecules that are hydrogen-bonded to glycine, results in excellent agreement with the experiment, i.e., the signs and the relative values of the shifts are correctly reproduced and the errors do not exceed 0.22 eV.

SUPPLEMENTARY MATERIAL

See supplementary material for programmable equations, relevant Cartesian geometries, benchmark results, graphical analysis of the errors, analysis of equilibrium MD simulations, validation of methodology for glycine, and additional results for glycine snapshots.

[1]O. Kostko, B. Bandyopadhyay, and M. Ahmed, Annu. Rev. Phys. Chem. **67**, 19 (2016).
[2]U. Ekström, P. Norman, V. Carravetta, and H. Ågren, Phys. Rev. Lett. **97**, 143001 (2006).
[3]U. Ekström and P. Norman, Phys. Rev. A **74**, 042722 (2006).
[4]N. A. Besley, M. J. G. Peach, and D. J. Tozer, Phys. Chem. Chem. Phys. **11**, 10350 (2009).
[5]N. A. Besley and F. A. Asmuruf, Phys. Chem. Chem. Phys. **12**, 12024 (2010).
[6]P.-F. Loos and X. Assfeld, Int. J. Quantum Chem. **107**, 2243 (2007).
[7]F. A. Asmuruf and N. A. Besley, Chem. Phys. Lett. **463**, 267 (2008).
[8]W. D. Derricotte and F. A. Evangelista, Phys. Chem. Chem. Phys. **17**, 14360 (2015).
[9]J. Wenzel, M. Wormit, and A. Dreuw, J. Comput. Chem. **35**, 1900 (2014).
[10]J. Wenzel, M. Wormit, and A. Dreuw, J. Chem. Theory Comput. **10**, 4583 (2014).
[11]S. Coriani, O. Christiansen, T. Fransson, and P. Norman, Phys. Rev. A **85**, 022507 (2012).
[12]T. J. Watson and R. J. Bartlett, Chem. Phys. Lett. **555**, 235 (2013).
[13]J. Kauczor, P. Norman, O. Christiansen, and S. Coriani, J. Chem. Phys. **139**, 211102 (2013).
[14]D. Zuev, E. Vecharynski, C. Yang, N. Orms, and A. I. Krylov, J. Comput. Chem. **36**, 273 (2015).
[15]B. Peng, P. J. Lestrange, J. J. Goings, M. Caricato, and X. Li, J. Chem. Theory Comput. **11**, 4146 (2015).
[16]S. Coriani and H. Koch, J. Chem. Phys. **143**, 181103 (2015).
[17]S. Coriani and H. Koch, J. Chem. Phys. **145**, 149901 (2016).
[18]P. Verma and R. J. Bartlett, J. Chem. Phys. **145**, 034108 (2016).
[19]M. Nisoli, P. Decleva, F. Calegari, A. Palacios, and F. Martín, "Attosecond electron dynamics in molecules," Chem. Rev. (published online).
[20]H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).
[21]R. Baer and D. Neuhauser, Phys. Rev. Lett. **94**, 043002 (2005).
[22]U. Salzner and R. Baer, J. Chem. Phys. **131**, 231101 (2009).
[23]R. Baer, E. Livshits, and U. Salzner, Annu. Rev. Phys. Chem. **61**, 85 (2010).
[24]P. Verma and R. J. Bartlett, J. Chem. Phys. **140**, 18A534 (2014).
[25]N. A. Besley, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **130**, 124308 (2009).
[26]H. Ågren, V. Carravetta, O. Vahtras, and L. G. M. Pettersson, Chem. Phys. Lett. **222**, 75 (1994).
[27]M. Nooijen and R. J. Bartlett, J. Chem. Phys. **102**, 6735 (1995).
[28]A. I. Krylov, "The quantum chemistry of open-shell species," in *Reviews in Computational Chemistry*, edited by A. L. Parrill and K. B. Lipkowitz (John Wiley & Sons, 2017), Vol. 30, Chap. 4, pp. 151–224.
[29]Special algorithms exist for addressing this type of issues, e.g., quadratically convergent MCSCF module in Dalton developed by Jensen and Ågren.
[30]L. S. Cederbaum, W. Domcke, and J. Schirmer, Phys. Rev. A **22**, 206 (1980).
[31]W. P. Reinhardt, Annu. Rev. Phys. Chem. **33**, 223 (1982).
[32]N. Moiseyev, Phys. Rep. **302**, 212 (1998).
[33]S. Klaiman and I. Gilary, Adv. Quantum Chem. **63**, 1 (2012).
[34]T.-C. Jagau, K. B. Bravaya, and A. I. Krylov, Annu. Rev. Phys. Chem. **68**, 525 (2017).

[35] D. Zuev, K. B. Bravaya, T. D. Crawford, R. Lindh, and A. I. Krylov, J. Chem. Phys. **134**, 034310 (2011).

[36] N. Moiseyev, *Non-Hermitian Quantum Mechanics* (Cambridge University Press, 2011).

[37] C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed. (VCH, Weinheim, 1990).

[38] A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, J. Phys. Chem. Lett. **2**, 2184 (2011).

[39] D. Ghosh, O. Isayev, L. V. Slipchenko, and A. I. Krylov, J. Phys. Chem. A **115**, 6028 (2011).

[40] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, J. Phys. Chem. B **115**, 8296 (2011).

[41] D. Ghosh, A. Roy, R. Seidel, B. Winter, S. Bradforth, and A. I. Krylov, J. Phys. Chem. B **116**, 7269 (2012).

[42] M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Annu. Rev. Phys. Chem. **64**, 553 (2013).

[43] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[44] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, J. Phys. Chem. A **105**, 293 (2001).

[45] L. V. Slipchenko, J. Phys. Chem. A **114**, 8824 (2010).

[46] D. Kosenkov and L. V. Slipchenko, J. Phys. Chem. A **115**, 392 (2011).

[47] P. Kumar, A. Acharya, D. Ghosh, D. Kosenkov, I. Kaliman, Y. Shao, A. I. Krylov, and L. V. Slipchenko, J. Phys. Chem. B **120**, 6562 (2016).

[48] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, and A. I. Krylov, J. Am. Chem. Soc. **138**, 4807 (2016).

[49] N. Ottosson, K. J. Børve, D. Spångberg, H. Bergersen, L. J. Saethre, M. Faubel, W. Pokapanich, G. Öhrwall, O. Björneholm, and B. Winter, J. Am. Chem. Soc. **133**, 3120 (2011).

[50] R. J. Bartlett, Annu. Rev. Phys. Chem. **32**, 359 (1981).

[51] R. J. Bartlett and J. F. Stanton, Rev. Comput. Chem. **5**, 65 (1994).

[52] R. J. Bartlett, Int. J. Mol. Sci. **3**, 579 (2002).

[53] A. I. Krylov, Annu. Rev. Phys. Chem. **59**, 433 (2008).

[54] R. J. Bartlett, Mol. Phys. **108**, 2905 (2010).

[55] K. Sneskov and O. Christiansen, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 566 (2012).

[56] R. J. Bartlett, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 126 (2012).

[57] J. F. Stanton and J. Gauss, J. Chem. Phys. **103**, 1064 (1995).

[58] M. Nooijen and J. G. Snijders, J. Chem. Phys. **102**, 1681 (1995).

[59] D. Ghosh, J. Chem. Phys. **139**, 124116 (2013).

[60] D. Ghosh, J. Chem. Phys. **140**, 094101 (2014).

[61] K. B. Bravaya, D. Zuev, E. Epifanovsky, and A. I. Krylov, J. Chem. Phys. **138**, 124106 (2013).

[62] D. Zuev, T.-C. Jagau, K. B. Bravaya, E. Epifanovsky, Y. Shao, E. Sundstrom, M. Head-Gordon, and A. I. Krylov, J. Chem. Phys. **141**, 024102 (2014).

[63] T.-C. Jagau, D. Zuev, K. B. Bravaya, E. Epifanovsky, and A. I. Krylov, J. Phys. Chem. Lett. **5**, 310 (2014).

[64] T.-C. Jagau and A. I. Krylov, J. Phys. Chem. Lett. **5**, 3078 (2014).

[65] P. U. Manohar and A. I. Krylov, J. Chem. Phys. **129**, 194105 (2008).

[66] P. U. Manohar, J. F. Stanton, and A. I. Krylov, J. Chem. Phys. **131**, 114112 (2009).

[67] S. V. Levchenko and A. I. Krylov, J. Chem. Phys. **120**, 175 (2004).

[68] P. A. Pieniazek, S. E. Bradforth, and A. I. Krylov, J. Chem. Phys. **129**, 074104 (2008).

[69] B. Xu, M. I. Jacobs, O. Kostko, and M. Ahmed, Chem. Phys. Chem. **18**, 1503 (2017).

[70] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).

[71] J.-M. Mewes, J. M. Herbert, and A. Dreuw, Phys. Chem. Chem. Phys. **19**, 1644 (2017).

[72] J. M. Mewes, Z. Q. You, M. Wormit, T. Kriesche, J. M. Herbert, and A. Dreuw, J. Phys. Chem. A **119**, 5446 (2015).

[73] S. Bose, S. Chakrabarty, and D. Ghosh, J. Phys. Chem. B **120**, 4410 (2016).

[74] D. Ghosh, J. Phys. Chem. A **121**, 741 (2017).

[75] M. S. Gordon, L. Slipchenko, H. Li, and J. H. Jensen, "The effective fragment potential: A general method for predicting intermolecular interactions," in *Annual Reports in Computational Chemistry*, edited by D. C. Spellmeyer and R. Wheeler (Elsevier, 2007), Vol. 3, Chap. 10, pp 177–193.

[76] D. Ghosh, D. Kosenkov, V. Vanovschi, C. Williams, J. Herbert, M. S. Gordon, M. Schmidt, L. V. Slipchenko, and A. I. Krylov, J. Phys. Chem. A **114**, 12739 (2010).

[77] D. Ghosh, D. Kosenkov, V. Vanovschi, J. Flick, I. Kaliman, Y. Shao, A. T. B. Gilbert, L. V. Slipchenko, and A. I. Krylov, J. Comput. Chem. **34**, 1060 (2013).

[78] P. Arora, L. V. Slipchenko, S. P. Webb, A. Defusco, and M. S. Gordon, J. Phys. Chem. A **114**, 6742 (2010).

[79] J. M. Olsen, K. Aidas, and J. Kongsted, J. Chem. Theory Comput. **6**, 3721 (2010).

[80] K. Sneskov, T. Schwabe, O. Christiansen, and J. Kongsted, Phys. Chem. Chem. Phys. **13**, 18551 (2011).

[81] J. M. H. Olsen, C. Steinmann, K. Ruud, and J. Kongsted, J. Phys. Chem. A **119**, 5344 (2015).

[82] M. N. Pedersen, E. D. Hedegård, J. M. H. Olsen, J. Kauczor, P. Norman, and J. Kongsted, J. Chem. Theory Comput. **10**, 1164 (2014).

[83] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kus, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguires, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Laog, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stuck, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Z. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).

[84] A. I. Krylov and P. M. W. Gill, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 317 (2013).

[85] I. A. Kaliman and L. V. Slipchenko, J. Comput. Chem. **34**, 2284 (2013).

[86] I. A. Kaliman and L. V. Slipchenko, J. Comput. Chem. **36**, 129 (2015).

[87] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).

[88] J.-D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106 (2008).

[89] J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[90] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, Jr., J. Comput. Chem. **31**, 671 (2009).

[91] S. E. Bradforth and P. Jungwirth, J. Phys. Chem. A **106**, 1286 (2002).

[92] N. V. Kryzhevoi and L. S. Cederbaum, J. Phys. Chem. Lett. **3**, 2733 (2012).

[93] M. Tronc, G. C. King, and F. H. Read, J. Phys. Chem. B **12**, 137 (2001).

[94] M. Coreno, M. de Simone, K. C. Prince, R. Richter, M. Vondráček, L. Avaldi, and R. Camilloni, Chem. Phys. Lett. **306**, 269 (1999).

[95] W. L. Jolly, K. D. Bomben, and C. J. Eyermann, At. Data Nucl. Data Tables **31**, 433 (1984).

[96] W. R. Daasch, E. R. Davidson, and A. V. Hazi, J. Chem. Phys. **76**, 6031 (1982).

[97] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, and V. Carravetta, J. Phys. Chem. A **111**, 10998 (2007).