A DFT/MRCI Hamiltonian Parameterized Using Only *Ab Initio* Data: II. Core-Excited States

Teagan Shane Costain,[1] Jibrael B. Rolston,[1] Simon P. Neville,[2] and Michael S. Schuurman[1,2]
[1)]*Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Canada*
[2)]*National Research Council Canada, 100 Sussex Dr., Ottawa, Canada, K1A 0R6*

(*Electronic mail: Michael.Schuurman@uottawa.ca)

(*Electronic mail: Simon.Neville@nrc-cnrc.gc.ca)

(Dated: 8 July 2024)

A newly parameterized combined density functional theory and multi-reference configuration interaction (DFT/MRCI) Hamiltonian, termed CVS-QE12, is defined for the computation of K-shell core-excitation and core-ionization energies. This CVS counterpart to the recently reported QE8 Hamiltonian [J. Chem. Phys, **160**, 224106 (2024)] is parameterized by fitting to benchmark quality *ab initio* data. The definition of the CVS-QE12 and QE8 Hamiltonians differ from previous CVS-DFT/MRCI parameterizations in three primary ways: (i) the replacement of the BHLYP exchange-correlation functional with QTP17 to yield a balanced description of both core and valence excitation energies, (ii) the adoption of a new, three-parameter damping function, and (iii) the introduction of separate scaling of the core-valence and valence-valence Coulombic interactions. Crucially, the parameters of the CVS-QE12 Hamiltonian are obtained via fitting exclusively to highly accurate *ab initio* vertical core-excitation and ionization energies computed at the CVS-EOM-CCSDT level of theory. The CVS-QE12 Hamiltonian is validated against further benchmark computations and is found to furnish K-edge core vertical excitation and ionization energies exhibiting absolute errors $\leq$ 0.5 eV at low computational cost.

I. INTRODUCTION

The recent push to develop electronic structure methods for X-ray spectroscopy, including those suitable for the simulation of time-resolved experiments, can be attributed at least in part to the expansion of X-ray free-electron laser facilities (XFELs) around the world. XFELs have the ability to generate X-ray radiation at tunable wavelengths that are also orders of magnitude brighter than previous technologies allowed.[1,2] Furthermore, the high-order harmonic generation (HHG) process, which can be generated in laboratory settings, can extend up to the soft X-ray range and can emit ultrashort pulses in the attosecond domain.[2] The principal advantage of ultrafast X-ray spectroscopy is the ability to probe structural dynamics at the fundamental timescales of atomic and electronic motion in an element-specific manner.[2]

These experimental advancements have spurred a corresponding development of *ab initio* electronic structure methods capable of describing the core-excited states accessed by X-ray absorption processes, as well as the simulation of the spectroscopy these light sources enable. As with the application of quantum chemistry to describe valence electronic structure, obtaining a high level of (controllable) accuracy more often than not implies the adoption of a wave function theory (WFT) based approach, with an accompanying trade-off in computational expediency. The consequence of this trade-off, then, is that these methods are generally not applicable beyond the study of small molecular species. Likewise, methods based upon density functional theory (DFT) are well-known for their computational efficiency, yet exhibit their own short-comings, including difficulty in the simultaneous description of states of different electronic character (i.e., charge-transfer, Rydberg, multi-reference, etc.), although the extent of these problems is exchange-correlation (XC) functional dependant.[3,4]

A standard approximation, employed in both WFT and DFT-based methods, for the computation of core-excited states is the core-valence separation (CVS) approximation,[5] in which the coupling between valence and core-excited configurations is assumed to be zero, thereby leading to a block diagonalization of the Hamiltonian between the two subspaces and allowing direct access to the core-excited states of interest.

Linear-response time-dependant density functional theory (LR-TDDFT) is currently the most common choice for computing excitation spectra, being far more computationally efficient than WFT-based methods.[6] Early applications of the CVS approximation to TDDFT[7–9] determined it to be a reliable method of reproducing and interpreting experimental spectra, and efforts have been made to develop XC functionals for the description of both core and valence states.[10] Nevertheless, TDDFT suffers from systemic limitations; it struggles to describe charge-transfer states, double-excitations, and Rydberg states.[11,12] Moreover, as a single-reference method, TDDFT cannot be used to describe states of multireference character, inhibiting its general use for the calculation of excited state core-level spectra, for which the initial states are often dominated by more than one electronic configuration. Another TDDFT variant used, albeit less often, in the calculation of core-level spectra is real-time TDDFT (RT-TDDFT).[11,13–15] While the excitation energies yielded by LR-TDDFT and RT-TDDFT are formally identical, the RT formulation can be more effective at yielding complete spectra of large systems featuring a high density of states.[11]

A time-independent alternative to TDDFT is orthagonality-constrained DFT (OCDFT).[16,17] Derricote and Evangelista

compared the performance of OCDFT with that of standard TDDFT, demonstrating the latter method as being able to describe orbital relaxation and CT states, while predicting core-excitation energies with an accuracy of around 1 eV. In common with TDDFT however, OCDFT cannot describe strongly correlated systems. Orbital optimized (OO) DFT approaches have also been used to overcome of number of the shortcomings of TDDFT. Hait *et al.* describe OO-DFT, based upon a Restricted Open-Shell Kohn–Sham (ROKS) framework, for computing valence and core-excited state spectra.[18–22] Employing an orbital optimized approach, one may target and directly solve for individual states, including doubly-excited states,[18] with sub-eV error.[18–22] The challenge with this general approach instead lies in solving for the target states; variational collapse and slow convergence issues when using the maximum overlap method (MOM) and initial MOM (IMOM) approaches, and both can exhibit spin-contamination, although recent developments[16,18,19,23,24] have sought to address this these issues. However, in general, employing OO methods requires some prior knowledge about the states of interest and individual computations are required for individual excited states, complicating the determination of entire spectra.

WFT-based methods offer a diverse and extensively developed range of approaches amenable to the description of core-excited states. The Algebraic Diagrammatic Construction (ADC) approach is one such family of methods and is the first approach in which the CVS approximation was applied[5] with the introduction of a CVS-ADC(2) implementation by Barth and Schirmer in 1985.[25] The CVS-ADC(2) method is found to accurately reproduce experimental transition energies and spectral intensities, with a maximum error of around 1 eV with respect to experiment. However, despite proving to be reliable, the CVS-ADC(2) method is limited by an $\mathcal{O}(N^5)$ asymptotic scaling, imposing a limit on the size of systems that can be feasibly studied. The CVS-ADC scheme has since been expanded to an extended second-order scheme, CVS-ADC(2)-x, and a third-order scheme, CVS-ADC(3).[26–28] In particular, the ADC(2)-x method has been shown to highly accurate for K-edge core-excited states; which is attributed to a fortuitous cancellation of errors.[27,28] However, the ADC(2)-x, and ADC(3) methods formally scale as $\mathcal{O}(N^6)$, again limiting the scope of their applicability to small and medium-sized molecules.

Recent years have seen the introduction of many new coupled-cluster (CC) methods for the computation of core-excited states and the simulation of X-ray spectroscopy. In particular, the CVS approximation has been applied to the linear response(LR)/equation-of-motion (EOM) coupled cluster singles and doubles (CCSD) approach.[29,30] This method formally scales as $\mathcal{O}(N^6)$ and has been demonstrated capable of producing highly accurate vertical excitation energies,[31–35] as well as absorption,[29–31,35] and photoelectron spectra.[36–38] The CVS-EOM-CCSD level of theory, as well as the approximate CC2 variant, exhibits errors in core-excitation energies on the order of 1 eV. The CVS-EOM-CCSDT method, benefiting from the full inclusion of triples, is found to converge within 0.1 eV of some experimental vertical core-excitation energies at the complete basis set limit.[31] The ionization potential variant, CVS-EOMIP-CCSDT, is found to be similarly accurate, furnishing ionization energies within 0.1 eV of experimentally determined values.[39] However, CVS-EOM-CCSDT has a steep asymptotic scaling of $\mathcal{O}(N^8)$,[40] making it prohibitively expensive beyond application to small molecules. Finally, we note that the multilevel CCSD approach, MLCCSD, allows for accurate results comparable to standard CCSD at a greatly reduced computational cost, but at the expense of a non-blackbox nature.[41]

We note that the description of static correlation in multi-reference systems presents a separate but equally relevant issue for the simulation of X-ray spectroscopy, particularly for electronically excited initial states, as encountered, e.g., in ultrafast X-ray spectroscopy studies. The reliable computation of electronic energies and properties of multi-reference states typically falls to the complete and restricted active-space self-consistent field (CASSCF and RASSCF) family of methods.[42,43] Dynamical correlation energy is generally accounted for using second-order perturbation theory, most commonly via the use of the CASPT2 and RASPT2 methods.[44–46] In contrast to the single-reference methods, active-space methods are infamously non-black-box and significant experience-informed intuition is often required to achieve stable and reliable results. That said, when performed correctly, these methods can achieve extremely high accuracy in the computation of core-excited states. For example, Odelius *et al.* used the RASSCF/RASPT2 calculations to simulate the L-edge spectra of a number of different transition metal complexes[47,48] with sub-eV deviations in peak position with respect to experiment not uncommon.

Lastly, The ADC and CC family of methods are both formally single-reference (SR) methods, but have also been formulated employing CAS and RAS and based references. For example, de Moura and Sokolov describe CVS-MR-ADC(2)-x for simulating X-ray photoelectron spectra (XPS).[49,50] Within this scheme, the ground electronic state is described by a CASSCF wave function. The results of this application appear promising, yielding a qualitatively and quantitatively correct description of the carbon K-edge XPS for *ortho*−, *meta*−, and *para*−benzyne.[49] Likewise, MR-EOM-CC theory is similarly implemented.[51]

This brief survey of quantum chemical methods for core-excited states demonstrates that, while the number of available approaches continues to expand, there is still a need for electronic structure methods that are simultaneously accurate enough to be predictive, sufficiently general to describe a wide range of molecular systems and electronic states, but also computationally efficient so as to be applied to large molecules. Additionally, a black box nature, requiring only the specification of the basis set and number of roots of interest, is highly desirable. A method that, somewhat uniquely, has the potential to simultaneously satisfy all of these requirements is the combined density functional theory and multi-reference configuration interaction (DFT/MRCI) approach.[52] Here, short, automatically tailored CI expansions aimed at the capturing of static correlation are paired with DFT-specific corrections to efficiently recover the preponderance of the re-

maining dynamic correlation. Hence, DFT/MRCI simultaneously incorporates the desirable characteristics from both WFT and DFT-based methods, while mitigating their respective drawbacks. Previous DFT/MRCI Hamiltonians have been parameterized to accurately reproduce valence excitation energies.[53–57] In particular, in a previous paper,[57] we presented a new Hamiltonian, termed QE8, that is fitted to highly accurate theoretical best estimates of valence excitation energies. The QE8 Hamiltonian is capable of furnishing valence excitation energies with a sub-0.2 eV accuracy.

The application of the CVS approximation to the DFT/MRCI method (CVS-DFT/MRCI) was previously explored[58] and found to furnish relatively accurate *relative* core-excitation energies. However, an unbalanced description of core and valence excitation energies was observed, with the CVS-DFT/MRCI core-excitation energies for K-edge transitions carrying *absolute* errors of up to 5 eV. This has since been attributed to the use of an exchange-correlation (XC) functional (BHLYP) that yields KS orbital energies whose differences give an unbalanced zeroth-order approximation of valence and core-excitation energies.[57] In Reference 57, with an eye to the present work, a new XC functional was identified that gives a significantly more balanced description of valence and core-excited states: QTP17. In this work, we explore the construction of a new DFT/MRCI Hamiltonian tailored to the description of core-excited states that employs this XC functional. Additionally, new DFT-specific Hamiltonian corrections are introduced that treat core-valence and valence-valence Coulomb interactions on a different footing, leading to yet more accurate results. The resulting Hamiltonian, termed CVS-QE12, is parameterized via fitting to reference CVS-EOM-CCSDT core-excitation and ionization energies and is demonstrated to have sub-eV accuracy for K-edge transitions.

In Sec. II, we give a brief description of the DFT/MRCI method, deferring much of this discussion to a companion work[57] that describes an analogous new parameterization tailored to valence excited states. Secs. III and IV describe the technical details of the parameterization optimization and an analysis of the results, respectively. Sec. V presents the results of some representative applications of the CVS-QE12 Hamiltonian, and Sec VI provides some summary results a prospectus for future work.

II. THE DFT/MRCI METHOD

In the following section we will present an abbreviated description of the DFT/MRCI method, focusing on those features that are the subject of the present work. For a more detailed account of the approach, we refer the reader to the Refs. 52–57,59.

A. Standard Formulation

As in an *ab initio* MRCI calculation, the DFT/MRCI wave function *ansatz* may, in general, be written in the following form:

$$|\Psi_I\rangle = \sum_{\mathbf{w}\omega\in\mathscr{R}} C_{\mathbf{w}\omega,I}|\mathbf{w}\omega\rangle + \sum_{\mathbf{w}\omega\in\mathscr{F}} C_{\mathbf{w}\omega,I}|\mathbf{w}\omega\rangle. \qquad (1)$$

Here, the wave functions $|\Psi_I\rangle$ are expanded in terms of spin-adapted configuration state functions $|\mathbf{w}\omega\rangle$ specified by a spatial occupation vector $\mathbf{w}$, giving the occupations of the spatial orbitals, and a spin coupling pattern ω, specifying how the open shells in $\mathbf{w}$ are spin-coupled. The total CSF basis is partitioned into two subsets: (i) the reference space $\mathscr{R}$, chosen to recover the static correlation of the system and to give a good zeroth-order description of the states of interest, and (ii) the first-order interacting space (FOIS) $\mathscr{F}$, obtained by the application of single and double excitation operators to $\mathscr{R}$, which, to first-order in perturbation theory, corrects the reference space description of the wave functions of interest.

The bottleneck in an *ab initio* MRCI calculation is the size of the FOIS, which, even for small molecules, can quickly become unmanageable with increasing basis and reference space dimensions. The main role of the FOIS is to recover the missing dynamic electron correlation via its coupling to the reference space. The DFT/MRCI method attempts to circumvent this by employing instead DFT-specific corrections to the on-diagonal Hamiltonian matrix elements to effectively account for the dynamic correlation, thus rendering almost all of the FOIS superfluous. This is achieved via the explicit replacement of Hartree-Fock orbital energy differences, which naturally appear in the working equations of the on-diagonal Hamiltonian matrix elements,[60,61] with their Kohn–Sham (KS) DFT equivalents. The latter gives an improved zeroth-order description of excitation energies, thereby incorporating a significant portion of the dynamic correlation energy in a manner removed from the coupling of the reference and FOIS spaces. As a result of this improved zeroth-order description, the accompanying Coulomb and exchange integral contributions to the on-diagonal Hamiltonian matrix elements must be empirically down-scaled. The on-diagonal DFT/MRCI Hamiltonian matrix elements may thus be conveniently written as a correction to the *ab initio* matrix elements in the following form:

$$\langle \mathbf{w}\omega|\hat{H}^{DFT} - E_{DFT}|\mathbf{w}\omega\rangle = \langle \mathbf{w}\omega|\hat{H} - E_{SCF}|\mathbf{w}\omega\rangle + \sum_i (\varepsilon_{ii}^{KS} - F_{ii})\Delta \mathrm{w}_i + \Delta E_x + \Delta E_c. \qquad (2)$$

Here, E_{SCF} and $\boldsymbol{F}$ are the SCF energy and Fock matrix evaluated in the given one-electron (orbital) basis, and ε_i^{KS} are the canonical KS DFT orbital energies. The terms ΔE_c and ΔE_x represent the aforementioned Coulomb and exchange integral corrections, the exact form of which depend on the DFT/MRCI parameterization.[52] In this work, we will focus on the R2017 parameterization that was the subject of a prior CVS implementation.[58] Here, these corrections terms take the following form:

$$\Delta E_c = -p_C \left[\sum_{i<j} V_{iijj} \Delta \mathrm{w}_i \Delta \mathrm{w}_j + \frac{1}{2} \sum_{i \in S_{\overline{\mathbf{w}}}} V_{iiii} |\Delta \mathrm{w}_i| + \sum_i V_{iiii} \delta_{\Delta \mathrm{w}_i, 2} \right] + \frac{1}{4} \sum_{i \in S_{\overline{\mathbf{w}}}} V_{iiii}, \tag{3}$$

$$\Delta E_x = p_X \left[- \sum_{\substack{i,j \\ j \in S_{\mathbf{w}}}} V_{ijij} |\Delta \mathrm{w}_i| + \frac{1}{2} \sum_{\substack{i \in C_{\mathbf{w}} \\ j \in A_{\mathbf{w}}}} V_{ijij} \Delta \mathrm{w}_i \Delta \mathrm{w}_j + \sum_{\substack{i<j \\ i,j \in S_{\mathbf{w}}}} V_{ijij} \eta_{ij}^{ji} \right]. \tag{4}$$

Here, $C_{\mathbf{w}}$ and $A_{\mathbf{w}}$ are the sets of indices or orbitals created and annihilated, respectively, in $\mathbf{w}$ relative to the base configuration $\overline{\mathbf{w}}$, and $S_{\mathbf{w}}$ and $S_{\overline{\mathbf{w}}}$ are the sets of indices of the singly-occupied orbitals in $\mathbf{w}$, and $\overline{\mathbf{w}}$, respectively. V_{ijkl} are standard two-electron integrals in chemists' notation, η_{ij}^{ji} are spin coupling coefficients (see Reference 57 for their definition), and $\Delta \mathrm{w}_i = \mathrm{w}_i - \overline{\mathrm{w}}_i$ denotes the difference of the occupation of the ith spatial orbital relative to the base occupation.

As it stands, the introduction of the DFT-specific corrections introduces a double-counting of dynamic correlation. To ameliorate this, the off-diagonal elements of the DFT/MRCI Hamiltonian are damped by an amount that is dependent on the energetic separation of the bra and ket CSFs. The key idea is that as the energy of a given CSF increases relative to the reference configuration, the contribution to static correlation is presumed to decrease, while its role in capturing dynamic correlation will increase. For bra and ket CSFs linked by a single excitation between orbitals i and a, the corresponding damped DFT/MRCI off-diagonal Hamiltonian matrix element reads

$$\left\langle \mathbf{w}\omega \middle| \hat{H}^{DFT} - E_{DFT} \middle| \mathbf{w}'\omega' \right\rangle = D(\Delta E_{\mathbf{w}\mathbf{w}'}) \left(\left\langle \mathbf{w}\omega \middle| \hat{H} - E_{SCF} \middle| \mathbf{w}'\omega' \right\rangle - F_{ia} \right), \tag{5}$$

where the corresponding off-diagonal Fock-matrix element is subtracted before damping. For bra and ket CSFs linked by a double excitation, the damped Hamiltonian matrix element is given by

$$\left\langle \mathbf{w}\omega \middle| \hat{H}^{DFT} - E_{DFT} \middle| \mathbf{w}'\omega' \right\rangle = D(\Delta E_{\mathbf{w}\mathbf{w}'}) \left\langle \mathbf{w}\omega \middle| \hat{H} - E_{SCF} \middle| \mathbf{w}'\omega' \right\rangle, \tag{6}$$

Here, $\Delta E_{\mathbf{w}\mathbf{w}'}$ denotes the spin-coupling averaged difference in on-diagonal Hamiltonian matrix elements for the bra and ket CSFs $|\mathbf{w}\omega\rangle$ and $|\mathbf{w}'\omega'\rangle$ (see Reference 57 for details).

The damping of these matrix elements decouples a large part of the FOIS from the reference space, thus making it effectively redundant, and allowing for its removal from the calculation. The subset of FOIS configurations that will contribute significantly to the damped DFT/MRCI Hamiltonian matrix is estimated *a priori* using a simple orbital energy-based selection criterion.[59] For each FOIS configuration $\mathbf{w}$, the quantity

$$d_{\mathbf{w}} = \sum_i \Delta \mathrm{w}_i \varepsilon_i^{KS} - \delta E_{sel} \tag{7}$$

is computed, where δE_{sel} is a free parameter. If $d_{\mathbf{w}}$ is less than the highest reference space eigenvalue of interest, then all the CSFs generated from the configuration $\mathbf{w}$ are selected for inclusion, else they are discarded. This value has previously been chosen to be either 1.0 or 0.8 E_h, with the latter correspondingly to a more aggressive truncation of the space of allowed excitations thereby leading to increased computational savings.

B. Improved description of core-excited states

The original CVS-DFT/MRCI Hamiltonian[58] systematically underestimates core-excitation energies, an effect that is observed to be particularly pronounced for states for which the (core) hole and (valence) particle orbitals of the dominant electronic configuration have a large degree of spatial overlap. This can be attributed to an overly large stabilization of the core-valence Coulombic interaction and suggests that a more severe down-scaling of the Coulomb corrections to the on-diagonal Hamiltonian matrix elements may be necessary.

Furthermore, this assessment is consistent with the observation that TDDFT core-excitation energies are generally more accurate for XC functionals with larger fractions of exact (HF) exchange, x_{HF}.[62–64] This manifests itself as the subtraction of $x_{HF} V_{iiaa}$ from the on-diagonal matrix elements, where V_{iiaa} is a Coulomb integral between the (core) hole and (valence) particle orbitals. Guided by these observations, one could expect that a simple increase in the value of the Coulomb damping parameter p_C might ameliorate the underestimation of core-excitation energies by the original CVS-DFT/MRCI implementation. However, as the CVS-DFT/MRCI configuration space is not limited to single excitations relative to the base configuration, valence-valence Coulomb integrals will also contribute to the on-diagonal Hamiltonian matrix elements. As the original R2017 Hamiltonian describes valence-valence interactions well, these should not be modified. Correspondingly, we are led to the introduction of separate core-valence and valence-valence Coulomb scaling parameters, denoted by $p_C^{(c,v)}$ and $p_C^{(v,v)}$, respectively. The on-diagonal Hamiltonian matrix elements thus take the following modified form:

$$\Delta E_c = -p_C^{(v,v)}\Bigg[\sum_{\substack{i<j\\ i,j\in\mathscr{V}}} V_{iijj}\Delta \mathrm{w}_i\Delta \mathrm{w}_j + \frac{1}{2}\sum_{i\in S_{\mathrm{w}}} V_{iiii}|\Delta \mathrm{w}_i| + \sum_i V_{iiii}\delta_{\Delta \mathrm{w}_i,2}\Bigg] - p_C^{(c,v)}\sum_{\substack{i<j\\ i\in\mathscr{C}}} V_{iijj}\Delta \mathrm{w}_i\Delta \mathrm{w}_j + \frac{1}{4}\sum_{i\in S_{\mathrm{w}}}^{n_{mo}} V_{iiii} \tag{8}$$

where $\mathscr{V}$ and $\mathscr{C}$ are the set of indices of the valence and core orbitals, respectively.

III. COMPUTATIONAL DETAILS

The optimization of the DFT/MRCI Hamiltonian scaling and damping parameters requires both the calculation of vertical excitation energies for a given parameter set, as well as the iterative variation and optimization of those parameters to minimize the error between the computed core-excitation energies and those of the reference values within the training set, where the latter are determined using a benchmark level *ab initio* quantum chemistry method. Iterative optimization of the empirical Hamiltonian parameters is performed by minimizing the root-mean-square deviation (RMSD) of the computed vertical excitation energies relative to the reference values. As the state ordering may vary as a function of the parameters, overlaps between the DFT/MRCI wave functions and those of a pre-computed set of the correct target character are computed at each iteration, allowing an unambiguous assignment of the states of interest. We refer the reader to a companion work[57] for a detailed discussion of the optimization procedure. The optimization step is followed by a validation step, in which the errors of the DFT/MRCI vertical excitation energies are computed relative to a separate subset of the computed reference data, the validation set, that is not used in the optimization step.

The following section details the construction of the reference dataset and the optimization and validation steps. Following this, computational details of selected representative applications are provided.

A. Training and Validation Set Data

The training and validation data is comprised of vertical core-excitation and ionization energies computed at the EOM-CCSDT and EOMIP-CCSDT[65,66] levels of theory using the aug-cc-pCVTZ basis set.[67,68] The MRCC package[69–71] was used to perform these calculations. The training set used is similar to that reported previously by Matthews *et al.*[33]; however, to facilitate state assignments, it was deemed expedient to re-compute these excitations in order to unambiguously identify the symmetry labels of each state.

The training set is composed of small organic molecules of photochemical relevance: C_2H_2, C_2H_4, CH_4, H_2O, NH_3, HF, NH_2F, CH_2NH, and HOF. The external validation dataset comprises the molecules CO, N_2, H_2CO, HCN, HNC, HNO, CH_3F, and C_2H_2O. For each molecule, singlet and triplet excited state energies and core-ionized doublet state energies were computed. Core-excited states corresponding to excitation into π^*, σ^*, and $3s$ and $3p$ Rydberg type orbitals were included. We note that the restriction to $n = 3$ Rydberg states was enforced to ensure that the basis sets used had good support for the states in the training set, noting that the correct description of higher-lying Rydberg states requires the inclusion of increasingly diffuse basis functions. Furthermore, only the core-electron transitions of second-row elements were considered. This includes carbon, nitrogen, oxygen, and fluorine. Organic photochemistry depends heavily on the former three elements; the strong polarizing effect of fluorine serves to diversify the chemical environments modeled in the training set. Thus, the present work should be considered first and foremost a parameterization for the description of K-edge transitions of organic molecules. We note that limiting the training set to small molecules increases the efficiency of the optimization procedure, and it allows for higher quality reference data to be considered; computing excitation energies at the selected level of theory, aug-cc-pCVTZ/CCSDT, becomes infeasible beyond very small molecules. Larger molecules are instead reserved for the computation of validation spectra for the re-optimized Hamiltonian. Finally, as CCSDT is preferentially accurate for one-electron transitions, doubly-excited core-excited states were excluded from the training set. This ensures that the reference excitation energy errors are on the order of 0.1 eV.[31,39] A few molecules in the training set, such as hypofluorous acid, nevertheless feature a high density of low-lying two-electron transitions, which are thus excluded from this analysis.

A summary of the training and validation datasets and their contents is provided in Table S1 of the Supplementary Information. The optimized ground state geometries of the molecules in the datasets are taken from literature sources (generally, from the QUEST electronic structure databases)[33,72,73] and are additionally provided as supplementary information (SI). In general, these correspond to aug-cc-pVTZ/CC3[74] optimized structures.

The DFT/MRCI computations were performed using the GRaCI software package,[75] which employs PySCF for KS-DFT calculations and integral transformations.[76–80] Density-fitting was employed for the evaluation of the two-electron integrals and employed an optimized auxiliary basis set.[81] For each molecule in either dataset, the number of roots computed is determined to include all excitations within the near-edge X-ray absorption fine structure (NEXAFS) region of the spectrum. This, in general, involves the computation of a large number of roots per molecule. As a consequence, the determination of an appropriate initial reference space is non-trivial, and the autoRAS procedure introduced in Reference 82 was employed to automate this process.

B. Representative Applications

Three sets of X-ray spectra are used to demonstrate the ability of the re-parameterized Hamiltonian to reproduce selected experimental data: (i) the XPS of the ethyl trifluoroacetate molecule is computed compared to the experimental spectrum, and (ii) The C, N, and O K-edge NEXAFS spectra of cytosine, and (iii) the ground and excited state O K-edge NEXAFS spectra of thymine. Each application is compared to experimental data.[83–86] The cytosine spectra are additionally compared to high-level *ab initio* levels of theory, namely, using fc-CVS-EOM-CCSD and CVS-ADC(2)-x. In the ethyl trifluoroacetate and thymine calculations, the aug-cc-pCVTZ basis was used, while the aug-cc-pCVDZ basis set was used in the cytosine calculations. The EOM-CCSD calculations were performed using the QChem program,[87] and the ADC-Connect code[88] was used for the ADC(2)-x computations. The ground and excited state spectra of thymine were computed using the relevant geometries (e.g. ground state minimum, $\pi\pi^*$, $n\pi^*$ excited state minimum) provided in Reference 86.

For the computation of XPS of ethyl trifluoroacetate, the electronic ionization cross-sections were approximated by the squared norms of the corresponding Dyson orbitals,[89,90] ϕ^D_{if}:

$$\phi^D_{if} = \langle \Psi^C_f | \Psi^N_i \rangle \tag{9}$$

where Ψ^N_i is the N-electron neutral state wave function, which in this case is the simply the ground electronic state, and Ψ^C_f is the wave function for cationic state f. The electronic integral in Eq. 9 runs over $N-1$ electrons. The ionization probability integrated over all outgoing photoelectron kinetic energies is then given by the squared norm of the Dyson orbital, G^D_{if}:

$$G^D_{if} = \langle \phi^D_{if} | \phi^D_{if} \rangle \tag{10}$$

Previous work has shown this quantity to be good a approximation, in the weak field limit, to the relative ionization probabilities to different cationic final states in the case of one-photon ionization processes.[91] In the present application, the neutral and cationic wave functions are expanded in a non-orthognoal orbital bases corresponding to the ground-state singlet (RKS) for the neutral and doublet (ROKS) orbitals for the cation, and the Dyson orbitals are computed using an extension of the wave function overlap algorithm of Plasser *et al.*[92]

IV. RESULTS

A. Selection of XC Functional

All previous DFT/MRCI Hamiltonians have been parameterized for use with the BHLYP XC functional. As noted in a companion work,[57] BHLYP-derived KS-orbital energy differences give a somewhat unbalanced zeroth-order description of core-excitation energies; quantities that enter directly into the diagonal matrix elements of the DFT/MRCI Hamiltonian (Equation 2). Specifically, the previously implemented CVS-DFT/MRCI method systematically underestimated core-excitation energies due to the underlying BHLYP KS orbital energies being too high in energy relative to the valence orbital energies. This will be characteristic of many global XC functionals, as it has been previously noted that the importance of the exchange contribution differs significantly between the core and valence orbitals.[93] This observation prompted a reconsideration of the underlying XC functional employed to compute the KS orbital basis. However, given the desire to develop a consistent family of methods that can be used for both valence and core-excitation, it is not sufficient to identify XC functionals that are appropriate solely for core-excitation. Rather, XC functionals capable of yielding a *balanced*, as well as accurate, description of core and valence excitation processes are required.

The ideal approach to identifying a new XC functional would be to perform parameter optimizations for each candidate. However, in terms of computational expense, this is computationally prohibitive. Instead, we used Tamm–Dancoff approximation time-dependent DFT (TDA-TDDFT) excitation energies as a suitable, computationally expedient proxy for the DFT/MRCI excitation energies of interest. The main idea underlying this choice is that the main contribution to both the DFT/MRCI and TDA-TDDFT on-diagonal matrix element is the differences between KS orbital energies.

TDA-TDDFT core and valence excitation energy calculations were performed for a series of small molecules, the full results of which are presented as supplementary material in a previous work.[57] These energies were compared to benchmark values, computed using aug-cc-pCVTZ/CVS-EOM-CCSDT for core-excitations, and taken as the QUEST database aug-cc-pVTZ/EOM-CCSDT value for valence excitations (see Table S6 of the supplementary material in Reference 57 for details). The TDA-TDDFT calculations were performed for the full set of XC functionals implemented in QChem version 5.4, and additionally for the QTP17 functional.

A comparison of the TDA-TDDFT and benchmark vertical excitation energies is given in Figure 1. Panel (a) gives a broad overview of the general performance, in terms of mean absolute error (MAE), over a large swath of the XC functionals considered, while panel (b) focuses on those functionals in the bottom left of panel (a) (i.e., those will small errors for both core and valence excitation) and shows the mean error of the most interesting candidates. The TDA-TDHF result (i.e., CIS) is shown in panel (a) for reference.

The functional used in the previous CVS-R2017 Hamiltonian, BHLYP, performs impressively well when evaluated by these metrics. At first glance, this would seem to imply that perhaps the ability to improve on the pre-existing KS orbital basis is minimal. However, the functional that exhibits the smallest MAE for core-excitation also furnishes mean errors for both valence and core-excitation that are closest to zero for all the XC functionals interrogated: QTP17.[94] The design

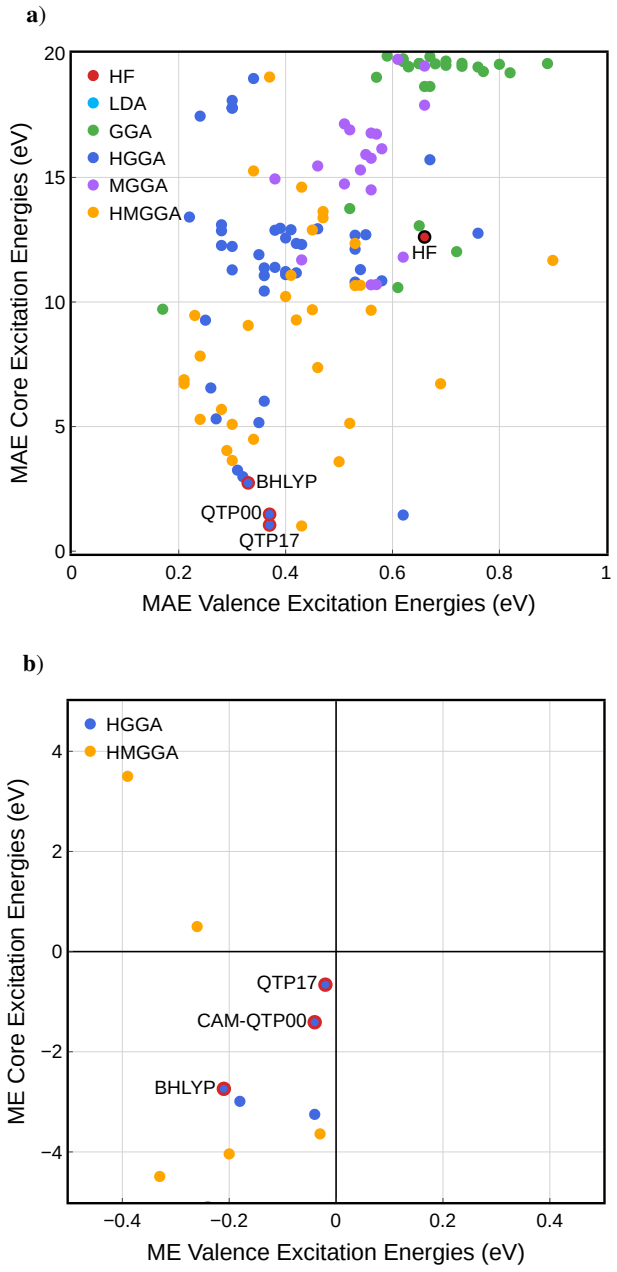

FIG. 1. Mean absolute errors (MAEs) for core and valence excitation energies for a variety of functionals computed using TDA-TDDFT are given in panel (**a**). Panel (**b**) presents a subset of functionals exhibiting the smallest mean errors (MEs) for valence and core-excitation energies. This figure is reproduced here from Reference 57. The numerical values for the data in this figure are given in the source publication.

principle of this functional was to construct a global hybrid functional such that KS orbital energies, i.e., Koopmans' ionization potentials (IPs), would reproduce each of the five vertical IPs of the water molecule. This design strategy results in QTP17 employing 62% exact HF exchange. The ability of this XC functional to furnish highly accurate core-excitation energies has been previously noted.[94]

The basis of this design principle may be explained as follows. Koopmans' theorem relates IP to the negative of the energy of the highest occupied molecular orbital (HOMO) within the Hartree-Fock scheme,[95] which is known as the Koopmans' IP. The extended Koopmans' theorem (EKT) correlates each orbital energy to a corresponding IP.[96–99] Equivalently, Janak's theorem is commonly cited as a proof that Koopmans' theorem extends equally to DFT.[100] However, it is generally stipulated that only the KS eigenvalue of the HOMO, which is approximately equal to the negative of the highest-lying IP, has any rigorous physical meaning. This interpretation is the IP theorem of Kohn–Sham DFT.[101] The Quantum Theory Project (QTP) family of functionals employ as a design principle rigorous adherence to a complete IP theorem.[94,102–104] Using the QTP17 functional, Jin and Bartlett report an MAE of approximately 0.9 eV with respect to experimental core-excitation energies, remarking that this method reduces the many-electron self-interaction error inherent to DFT-based methods.[94]

As discussed in Sec. II A, the KS orbital energies form the leading order description of the excitation energies in the DFT/MRCI formalism. It is notable that the QTP17 functional yields *absolute* excitation errors for core-excitation that are of the same magnitude as the valence vertical excitation energies, despite the former corresponding to energy differences that are 100× larger. On the basis of these results, in the following, we will employ the QTP17 functional to generate the underlying KS orbital basis for present parameterization.

B. New Parameterization: CVS-QE12

1. *Form of the Damping Function*

We employ the same general form of the damping function presented in Reference 57, namely, a three-parameter exponential function given by

$$D(\Delta E_{\mathbf{ww}'}) = d_1 \exp\left(-d_2 \Delta E_{\mathbf{ww}'}^{d_3}\right) \tag{11}$$

The damping function parameters d_1 and d_2 were optimized simultaneously with the scaling parameters, while d_3 was scanned over integer values between 6 and 16. Optimized damping functions for a series of CVS-QEn Hamiltonians (where $n = d_3$), as well as the damping function used in the CVS-R2017 Hamiltonian, are shown plotted in Figure 2.

Consistent with the recent QE8 parameterization,[57] we have found it prudent to enforce a constraint on the parameter d_2. Since the reference space is generated via a selected CI algorithm that ensures all reference states provide good support for the states of interest, the size of the reference space will be dependent on the number of roots, N_{root}, computed. With increasing N_{root} the span of the eigenspectrum of the reference space increases, leading to different effective scalings for different roots of the DFT/MRCI Hamiltonian. This, in turn, can result in a pronounced dependence of the computed excitation energies on the number of roots included.

In the present case, we observed that an unconstrained optimization resulted in damping functions that have significant magnitude for $\Delta E_{\mathbf{ww}'} > 1.0\ E_h$. Practically, this meant that the lowest-lying core-excitation could shift by up to 1.0 eV when varying N_{root} from 1 to 100. Given that the number of states in an XAS or XPS for moderately sized molecules can be large, this was deemed a highly undesirable characteristic of the method. Thus, we have chosen to impose the constraint $d_3 \geq 2.303$ which corresponds to a damping function value of 0.1 at $\Delta E_{\mathbf{ww}'} = 1$. Parameter sets optimized using this constraint were found to reduce to the errors in the lowest core transition to ≤ 0.5 eV, irrespective of value of N_{root}, while maintaining accurate error metrics in the excitation energies.

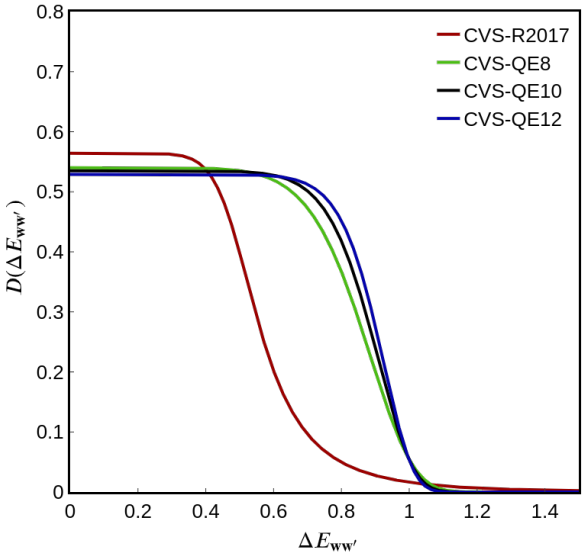

FIG. 2. The damping function applied to the off-diagonal matrix elements of the DFT/MRCI Hamiltonian for current the (CVS-QE*n*) and previous (CVS-R2017) parameterizations.

The damping functions resulting from this constrained optimization procedure are shown in Figure 2 for parameter values $d_3 = 8, 10,$ and 12. It is found that, while the $\Delta E_{\mathbf{ww}'} = 0$ limiting value of the damping decreases only slightly ($d_1 = 0.529$) compared to the CVS-R2017 value (0.564), the re-optimized damping functions exhibit much slower decays, thus capturing significantly more dynamics electron correlation. The MAE for the optimized CVS-QE*n* Hamiltonians decreases steadily with increasing value of the parameter d_3 up to $d_3 = 12$, upon which the improvement becomes marginal (see section S2 of the supplementary material). Thus, a value of $d_3 = 12$ was adopted, corresponding to the CVS-QE12 Hamiltonian.

2. *Analysis of Optimized Parameters*

Table I presents the optimized parameter values of the CVS-QE12 Hamiltonian, the corresponding error metrics, as well as those of the previous CVS-R2017 Hamiltonian. Separate parameter sets for energy-based configuration selection thresholds δE_{esel} of 0.8 and 1.0 E_h were determined. In the subsequent discussion, we will refer exclusively to the $\delta E_{sel} = 1.0\ E_h$ results, but the slightly less accurate δE_{sel}=1.0 E_h results are shown for consistency with previous parameterizations.

TABLE I. Hamiltonian Parameters and Error Summary (Mean Absolute Errors in eV)

Hamiltonian	E_{sel}	$p_C^{(v,v)}$	$p_C^{(c,v)}$	p_X	d_1	d_2	d_3
CVS-QE12	1.0	0.426	0.564	0.255	0.529	2.303	12
	0.8	0.419	0.560	0.258	0.546	2.303	12
	E_{sel}	p_J	p_X		p_1[a]	p_2	
CVS-R2017	1.0	0.503	0.359		0.564	1.86	
	0.8	0.501	0.357		0.574	1.93	

Training Data

Hamiltonian	E_{sel}	Singlet	Triplet	Doublet	Total
CVS-QE12	1.0	0.61	0.55	0.24	0.53
	0.8	0.54	0.51	0.25	0.49
CVS-R2017	1.0	3.66	3.58	5.28	3.87
	0.8	3.63	3.55	5.33	3.85

Validation Data

Hamiltonian	E_{sel}	Singlet	Triplet	Doublet	Total
CVS-QE12	1.0	0.63	0.51	0.33	0.54
	0.8	0.62	0.52	0.37	0.54
CVS-R2017	1.0	3.89	3.90	5.10	4.05
	0.8	3.81	3.82	5.07	3.98

[b] A different form of the damping function is employed in R2017 such that the parameters (d_1,d_2,d_3) and (p_1,p_2) are directly comparable.

We start with a comment on the optimized value of the Coulomb integral scaling terms, $p_C^{(v,v)}$ and $p_C^{(c,v)}$. This coefficient scales the integrals corresponding to the hole-particle Coulomb interaction and it has been previously noted[54,59] that equivalent terms in the TDDFT formalism are explicitly scaled by $1 - x_{HF}$, where x_{HF} denotes the proportion of exact exchange present in the XC functional. This relation was observed to hold for the QE8 valence Hamiltonian parameterization, in which an optimized value of $p_C = 0.426$ was obtained for the QTP17 functional (for which $1 - x_{HF} = 0.38$). Here, the valence-valence scaling parameter, $p_C^{(v,v)}$, remains unchanged from the QE8 valence parameterization, with an identical value of $p_C^{(v,v)} = 0.426$ being attained. In contrast, the optimized value of the core-valence Coulomb scaling parameter is $p_C^{(c,v)} = 0.564$, corresponding to an *increased* damping of this interaction. Selective down-scaling of the core-valence Coulomb interactions is similar in effect to the increased admixture of exact exchange in XC functionals, which is known to yield better accuracy for core transition energies. K-shell electrons, confined to a small spatial region, constitute a point-charge-like distribution relative to outer-shell electrons. Such electrons are described as approaching the high-density (weak-interaction, or weakly-correlated) limit in DFT,[105] wherein electron exchange is predominant. Either

by increasing the exact HF exchange in the XC functional or by down-scaling the two-electron Coulomb integrals, the self-interaction error (SIE) is corrected.

In TDDFT, a combination of global hybrid[10,62,63,94,106–108] and short-range corrected (SRC)[10,62,93,108] functionals have been designed to correct for the SIE of core electrons by employing increased fractions of exact HF exchange. The balance of exact and semi-local exchange in these approaches can nevertheless have drawbacks,[6,105,109] which has been characterized as a "zero-sum trade-off" between the over-localization and over-delocalization of electrons,[110] manifesting as either under-binding of covalent bonds or pervasive self-interaction error. In the present work, the separated Coulomb scaling for valence-valence and core-valence interactions is made necessary due to the *global* nature of the exact exchange admixture in the QTP17 functional.

3. *Performance on Training and Validation Set Data*

Figure 3 shows the MAEs in core-excitation and ionization energies yielded by the CVS-QE12 and CVS-R2017 Hamiltonians with respect to the training and validation datasets. For singlet vertical core-excitation energies, the CVS-QE12 Hamiltonian yields and MAE of 0.61 eV with respect to the transitions present in the training set. This is to be compared to an MAE of 3.7 eV for the old CVS-R2017 Hamiltonian. That is, an order of magnitude improvement is attained. For vertical core-ionization potentials, the CVS-QE12 Hamiltonian exhibits an MAE of 0.24 eV, compared to an MAE of 5.3 eV for the CVS-R2017 Hamiltonian. In addition to dramatically reducing the errors in these quantities, the CVS-QE12 Hamiltonian is also considerably more consistent in the description of both core-excitation and ionization energies relative to the old CVS-R2017 Hamiltonian. For a complete list of MAEs for the CVS-QE12 and CVS-R2017 Hamiltonians with respect to the training set, see Table I. The errors in the core-excitation and ionization energies furnished by the CVS-QE12 Hamiltonian for the validation set agree to within 0.1 eV of those for the training set: the CVS-QE12 Hamiltonian yields an MAE of 0.63 eV for the singlet excitation energies and an MAE of 0.54 eV over all states. The complete list of MAEs for both datasets is tabulated in Table I.

In Figure 4, we show the distribution of errors for the CVS-QE8 and CVS-R2017 Hamiltonians. For both the training set (top panel) and validation set (middle panel), the CVS-QE12 Hamiltonian yields distributions of errors almost centered around zero, with a slight positive bias for the training set and a negative bias for the validation set. In comparison, the CVS-R2017 Hamiltonian yields a much broader spread of errors approximately centered at -4 eV. This figure unambiguously shows the improvement, both in terms of accuracy and precision, of the CVS-QE8 Hamiltonian relative to the old CVS-R2017 Hamiltonian.

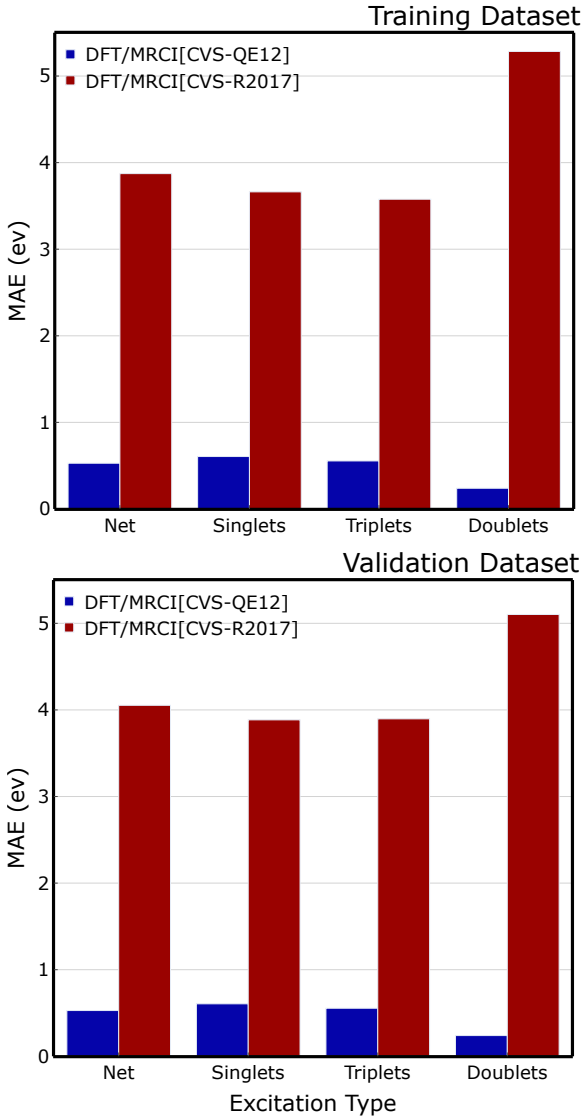

FIG. 3. MAE of DFT/MRCI with respect to the Reference Data using the CVS-QE12 and CVS-R2017 Hamiltonian. The performance of each Hamiltonian is compared for the data in the training set, and the external validation dataset.

4. *Performance of Approximate DFT/MRCI Methods*

Finally, we assess the performance of the pruned and perturbative DFT/MRCI variants, *p*-DFT/MRCI and DFT/MRCI(2), respectively, in conjunction with the CVS-QE12 Hamiltonian. These variants are known to greatly reduce the compute expense of the DFT/MRCI method, with a negligible trade-off in accuracy.[82,111] The distributions of errors for *p*-DFT/MRCI and DFT/MRCI(2) for the validation set are shown in the bottom set of panels of Figure 4, with the DFT/MRCI distribution shown alongside for reference. Overall, both *p*-DFT/MRCI and DFT/MRCI(2) perform very similar to the parent DFT/MRCI method, both in terms of distri-

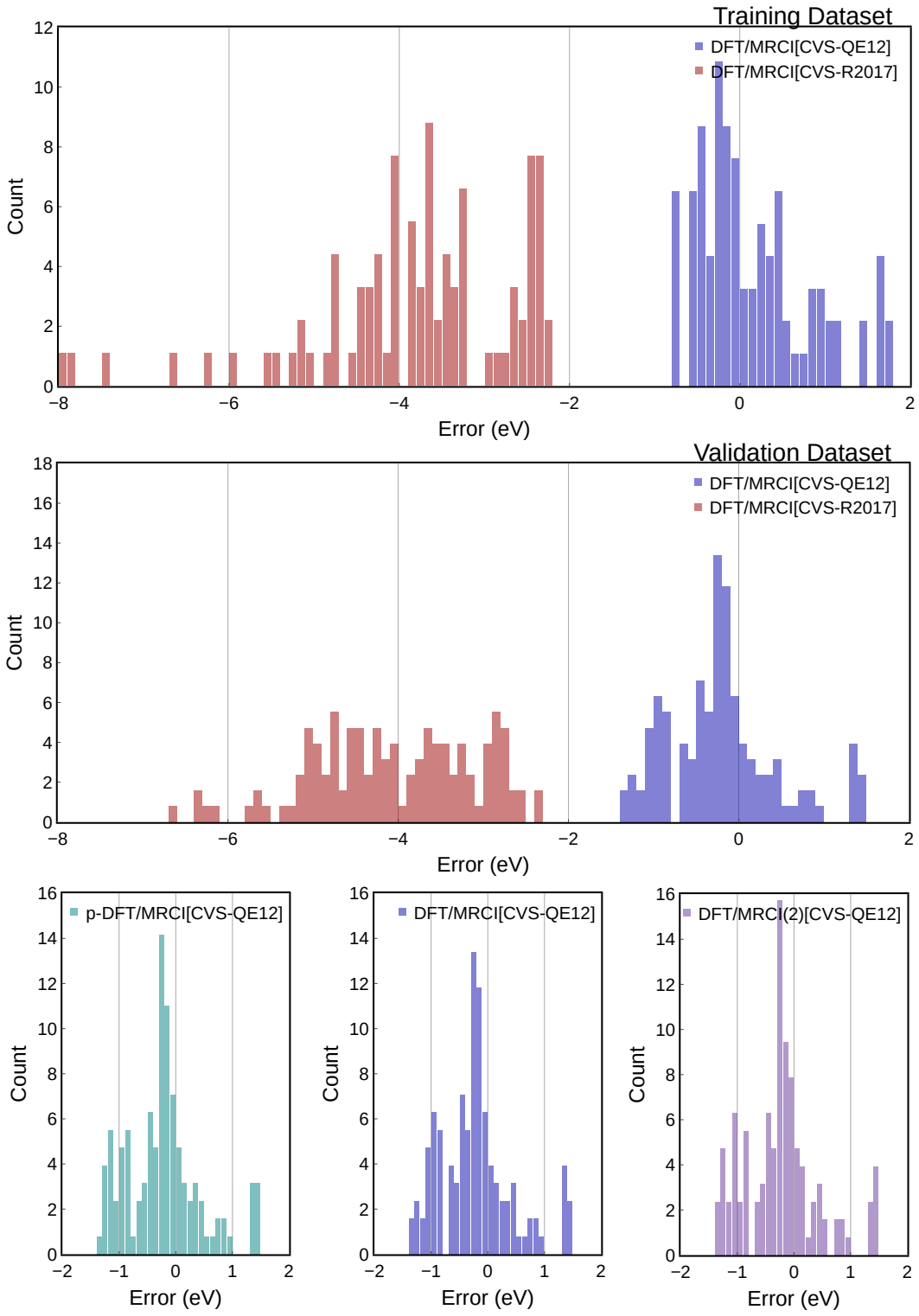

FIG. 4. Error distribution histograms for the CVS-R2017 (red) and the CVS-QE12 Hamiltonians (blue) with respect to the training (top), and the validation (middle) datasets. Error distribution histogram of the CVS-QE12 Hamiltonian with respect to the validation dataset (bottom) for the *p*-DFT/MRCI (teal), DFT/MRCI (blue), and DFT/MRCI(2) (purple) variants.

bution width and center.

The MAEs for the *p*-DFT/MRCI and DFT/MRCI(2) variants are given in Table II for both the training and validation datasets. To aid comparison, the DFT/MRCI results are shown alongside. Overall, both *p*-DFT/MRCI and DFT/MRCI(2) yield MAEs to within around 0.01 eV of the parent DFT/MRCI values. This corresponds to an impressive error of around just 0.05% error across transitions. Furthermore, the error histograms in Figure 4 show that not only do the aggregate error metrics for *p*-DFT/MRCI and DFT/MRCI(2) match those of the parent DFT/MRCI method, the error distributions are very similar, covering the same energy range and exhibiting the similar structure. In fact, across all excitation types the mean *absolute* error in the *p*-DFT/MRCI and DFT/MRCI(2) vertical excitation energies relative to the DFT/MRCI results are an astonishing −0.01 and −0.02 eV, respectively.

TABLE II. Comparison of Errors for DFT/MRCI and Approximate Methods [*p*-DFT/MRCI and DFT/MRCI(2)] (Mean Absolute Errors in eV)

Training Data				
Method[CVS-QE12]	Singlet	Triplet	Doublet	Total
DFT/MRCI	0.61	0.55	0.24	0.53
p-DFT/MRCI	0.62	0.57	0.23	0.54
DFT/MRCI(2)	0.62	0.58	0.23	0.54
Validation Data				
Method[CVS-QE12]	Singlet	Triplet	Doublet	Total
DFT/MRCI	0.63	0.51	0.33	0.54
p-DFT/MRCI	0.64	0.52	0.32	0.55
DFT/MRCI(2)	0.64	0.52	0.33	0.55

The main conclusions to be drawn from this analysis are two-fold. First, the previously noted high fidelity between the approximate *p*-DFT/MRCI and DFT/MRCI(2) methods and the parent DFT/MRCI method for valence excitation is equally valid for core-excitation energies. Second, the new CVS-QE8 parameterization does not adversely impact the ability of these approximate methods to quantitatively reproduce the DFT/MRCI results. The consistency in the computed excitation energies is quantitative, with *p*-DFT/MRCI and DFT/MRCI(2) displaying MAEs that differ from those of DFT/MRCI by 0.02 eV and 0.06 eV, respectively, for all excitation classes in the validation set.

V. REPRESENTATIVE APPLICATIONS

A. X-ray Photoelectron Spectra: The ESCA Molecule

To demonstrate the improved description of core-ionized states by the CVS-QE12 Hamiltonian, we consider the simulation of the X-ray photoelectron spectrum (XPS) of ethyl trifluoroacetate, also known as the ESCA molecule.[112] This is a benchmark spectrum, with the four carbon 1*s* orbitals being energetically well-separated from one another due to the presence of the trifluoromethyl and acetyl functional groups, leading to four distinct, non-overlapping peaks in the XPS. The large separation between the four carbon 1*s* binding energies helps to highlight any disagreement between the theoretical and experimental core-ionization energies, which can otherwise be ambiguous in systems with congested core-level spectra.

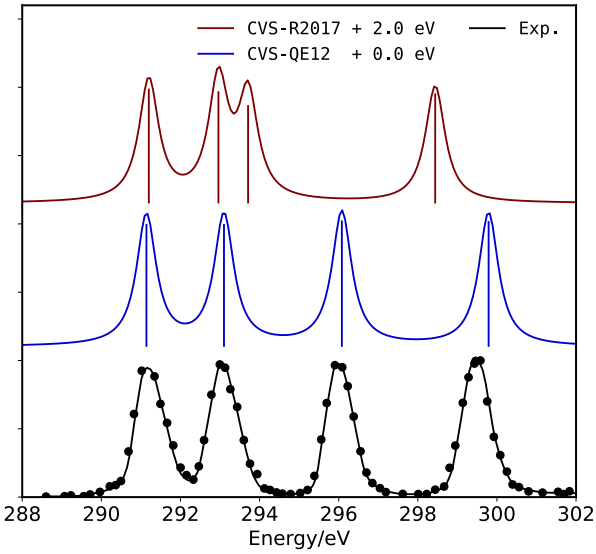

FIG. 5. Carbon XPS of Ethyl Trifluoroacetate Simulated with CVS-DFT/MRCI, using the CVS-R2017 Hamiltonian (red) and the CVS-QE12 Hamiltonian (blue), with Comparison to Experimental Gas-Phase X-Ray Photo-electron Spectrum (black). The simulated spectra consider only the anti-anti-conformer. Experimental data, shown with line of best fit, are digitized from source.[83]

Ethyl trifluoroacetate exists in two major conformations: *anti-anti* and *anti-gauche*. Both conformations have a CCOC dihedral angle of about 180°. In the *anti-anti* conformation, the COCC dihedral angle is about 180°; in the *anti-gauche* conformation, it is oriented *gauche*, such that the ethyl group appears bent into (or out of) the plane. Using gas electron diffraction, the conformers are determined to exist at a ratio of 44:56 in the gas phase at 293 Kelvin.[113] The energy difference between the conformers is experimentally estimated at approximately 1.10 kJ/mol, *anti-anti* being lower in energy. Accordingly, the differences in the binding C 1s energies of either conformer are small (around 0.1 eV) and do not significantly change the resultant spectrum.[114] Thus, only the *anti-anti* geometry is considered in the present work.

The carbon 1*s* XPS was simulated using both the CVS-QE12 and CVS-R2017 Hamiltonians and aug-cc-pCVTZ basis. The simulated spectra are shown in Figure 5 alongside the experimental gas-phase spectrum adapted from Reference 83. The stick spectra, determined from vertical ionization energies, are convoluted with a Lorentzian function with a full-width half-maximum (FWHM) of 0.6 eV to obtain the spectral envelope. This convolution function has been used for all spectral simulations in this section. The CVS-R2017 results are shown shifted by 2 eV to bring the simulated and experimental spectra into maximum agreement.

Although the relative positions of the first two peaks are well reproduced, the position of the third peak, corresponding to the carbonyl carbon atom, is drastically underestimated. The position of the fourth peak, corresponding to the carbon atom of nearest the fluorine substituent, is also somewhat underestimated. In contrast, the CVS-QE12 XPS is in excellent agreement with the experimental spectrum both qualitatively and quantitatively, with zero shift needed to align it with the experimental spectrum, and the relative positions of all four peaks correctly described.

B. X-ray Absorption Spectra: Cytosine

In Figure 6 we show the NEXAFS spectra for the carbon, nitrogen, and oxygen K-edges of cytosine, simulated using the CVS-QE12 Hamiltonian. For comparison, shown alongside are the results of CVS-ADC(2)-x and CVS-EOM-CCSD simulations as well as the experimental spectra. For each edge, four structures of cytosine are considered. The experimental spectra were recorded using an evaporation temperature of 450K. As a result, multiple thermally-accessible geometries must be considered. The same four structures as used in References 84 and 85 were used here. The individual spectra computed for the four tautomer structures were weighted and combined using the appropriate Boltzmann Population Ratios (BPRs) at 450K, as computed by Trygubenko *et al.*[84,115] Each spectrum was individually shifted to bring it into maximal alignment with the experimental spectrum, and these shifts are indicated in the legend of each plot in Figure 6.

We first note that the spectra computed with CVS-QE12 Hamiltonian require smaller shifts in order to align them with the experimental data than those simulated with ADC(2)-x and EOM-CCSD; the shifts applied to the CVS-DFT/MRCI spectra range from -0.7 eV to -1.1 eV, whilst the ADC(2)-x spectra are blue-shifted on average by about 1.9 eV, and the EOM-CCSD spectra by about 2.5 eV.

Moreover, the CVS-QE12 Hamiltonian yields excellent relative peak positions across all edges. Whereas EOM-CCSD yields poor separation of the first and second major peaks in the carbon XAS, both ADC(2)-x and DFT/MRCI describe the relative separation of these peaks well. A remarkable feature of the oxygen XAS is the presence of two peaks in the simulated CVS-DFT/MRCI spectrum which are not apparent in the ADC(2)-x and EOM-CCSD spectra. These features are the shoulder transition at approximately 531.8 eV, and the low-intensity central peak at 533.7 eV, which appears at around 534 eV in the experimental spectrum. Analysis of the natural transition orbitals for these transitions reveals that the former peak, absent from the EOM-CCSD and ADC(2)-x spectra, corresponds to an O $1s \rightarrow \pi*$ transition with significant doubly-excited character for the ground-state structure of cytosine. The peak at 533.7 eV corresponds to an analogous transition from the carbonyl oxygen in the tautomer, labeled structure 3 in Reference 84 (see Fig. S7 in the supplementary material).

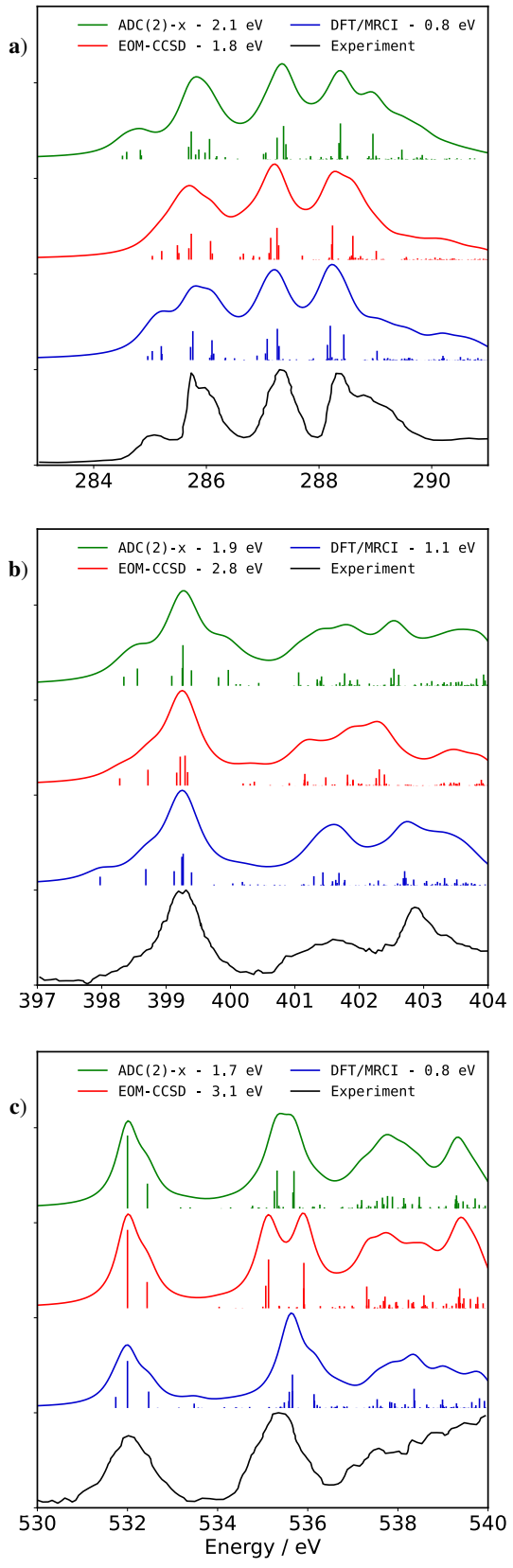

FIG. 6. NEXAFS Spectrum of Carbon (**a**), Nitrogen (**b**), and Oxygen (**c**) K-shells of Cytosine Simulated with CVS-ADC(2)-x (green), CVS-EOM-CCSD (orange), and CVS-DFT/MRCI (CVS-QE12) (blue). Experimental spectra (black) digitized from source.[84,85]

C. Excited-State Spectra: Thymine

One application that motivates the present work is the simulation of time-resolved pump-probe spectroscopies. Here, we consider the case of a UV-Vis pump, X-ray probe experiment. Here, the pump pulse creates an excited state wave packet in a manifold of low-lying valence-excited electronic states that is subsequently probed via either excited-state X-ray transient absorption or X-ray photoelectron spectroscopy. The quantum chemistry method employed to simulate such processes must be able to readily compute optical properties from valence-excited initial states.

To demonstrate the efficiency and accuracy of the DFT/MRCI family of methods applied to such problems, we consider the previously studied[116] photoprotective mechanism of the nucleobase thymine. Similar to other organic chromophores, thymine exhibits a low-lying bright transition to a $\pi\pi^*$ state, while the ground-state transition to the $n\pi^*$ state is optically dark. A rapid relaxation of the chemically reactive $\pi\pi^*$ state into the $n\pi^*$ state has been termed a photoprotective mechanism; UV-induced dimerization of nucleobases in DNA is mitigated by the rapid depopulation of the $\pi\pi^*$ state.[116] Due to the site-specificity of X-ray spectroscopy, the chemical shift of oxygen 1s K-edge transitions effectively probe the change in the non-bonding electron density at the oxygen atomic sites following transition to the $n\pi^*$ state. Wolf *et al.* have studied this process in detail.[86,117,118] Using Auger-Meitner spectroscopy (or resonant Auger spectroscopy), the population of the $n\pi^*$ state in thymine is observed to occur within 60 ± 30 fs of UV excitation.[86]

Shown in Figure 7(a) are the oxygen K-edge NEXAFS spectra computed using DFT/MRCI, p-DFT/MRCI, and DFT/MRCI(2). The experimental K-edge Auger spectra from Reference 86 are shown alongside in Figure 7(b). We note that the Auger spectrum is analogous to NEXAFS spectra as the integrated Auger electron yield is proportional to the absorption cross-sections of a typical photoabsorption spectrum. The ground state experimental spectrum, in black, is shown alongside the spectrum measured 2 ps after UV photoexcitation to the $\pi\pi^*$ state, in green. It was estimated that the initial pump pulse excites 13% of the sample,[86] and this scaling factor was applied to excited state oscillator strength to obtain the DFT/MRCI results in Figure 7(a).

We first note that the agreement between the experimental result and the NEXAFS simulations strongly supports the original assignment made in Reference 86. Specifically, that the time-delayed feature at 526.5 eV in Figure 7(b) is attributable to population of the $n\pi^*$ state. This is clearly seen in Figure 7(a) from the NEXAFS spectrum computed at the minimum energy geometry of the $n\pi^*$ state, which aligns very well with this feature. The fact that the $\pi\pi^*$ state (i.e., the orange line in Figure 7(a)) is predicted to yield very little signal at the O K-edge is also confirmed by the present computations. This can be readily rationalized by the poor overlap between the O 1s orbital and the π and π^* orbitals, which yields vanishingly small transition densities. Second, we note that no empirical shifts have been applied to any of the simulated spectra, highlighting the excellent (< 1 eV) accuracy of the CVS-QE12 Hamiltonian. That said, a uniform shift of both the ground and excited state spectra (the black and green curves, respectively), would bring the simulated results into near-perfect agreement with the experimental spectrum. Finally, the ability of the approximate DFT/MRCI methods, p-DFT/MRCI and DFT/MRCI(2), to yield near-indistinguishable results is again confirmed here.

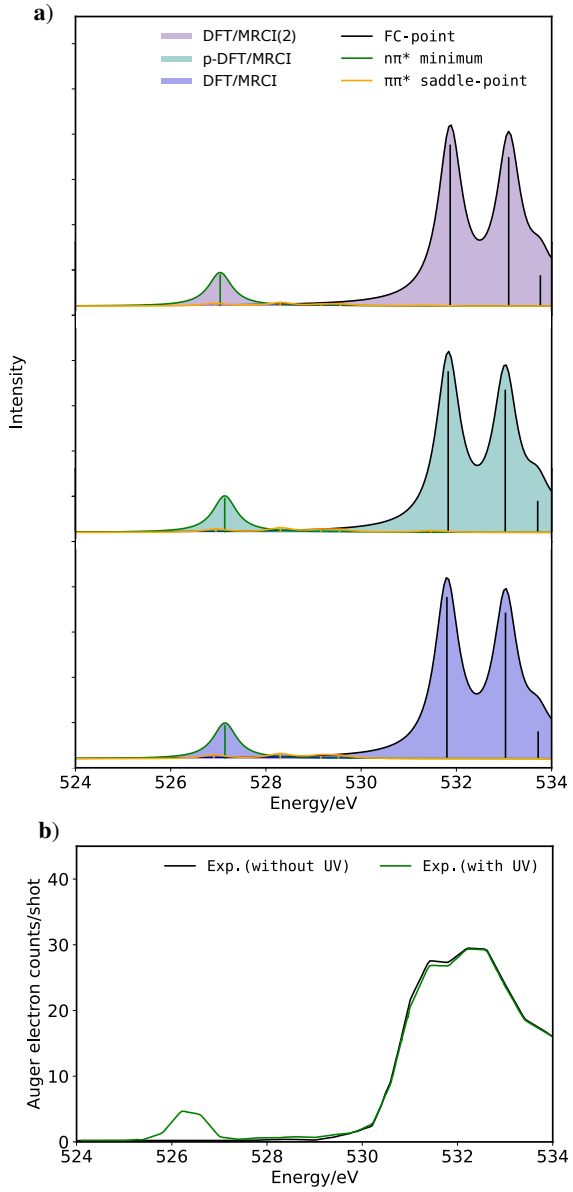

FIG. 7. Simulated Oxygen 1s NEXAFS Spectrum of Thymine (**a**) and Experimental Oxygen 1s Auger Spectrum (**b**). The simulated spectra (unshifted) are computed with DFT/MRCI(2) (purple), p-DFT/MRCI (teal), and DFT/MRCI (blue) using CVS-QE12 at the Franck-Condon point geometry (ground-state spectrum in black), the $n\pi^*$ minimum geometry ($n\pi^*$ excited-state spectrum in green), and the $\pi\pi^*/n\pi^*$ saddle-point geometry ($\pi\pi^*$ excited-state spectrum in orange); experimental spectra in (**b**) are digitized from source.[86]

D. Computational Efficiency: Timings

We conclude by presenting computational timings for the representative applications of this section. Table III gives the CPU times for all post-DFT aspects of the calculations, including the determination of the transition density matrices or Dyson orbitals for the spectral simulations. The results shown here were generated using the GRaCI package,[75] which uses the PySCF[76–80] library to solve the KS DFT equations. However, other packages may be employed in this respect, the computational efficiency of which will vary.

Figure 7, along with Figures S7 and S9 of the supplementary material, shows that DFT/MRCI, *p*-DFT/MRCI and DFT/MRCI(2) yield nearly identical spectral simulations. However, the latter two methods lead to order of magnitude speedups relative to the parent DFT/MRCI method. Indeed, every spectral simulation employing the *p*-DFT/MRCI and DFT/MRCI(2) methods took just over one minute to just under five minutes of CPU time.

TABLE III. CPU times (in seconds) of the DFT/MRCI, *p*-DFT/MRCI, and DFT/MRCI(2) calculations for the Representative Applications.[b]

Ethyl Trifluoroacetate				
DFT/MRCI	*p*-DFT/MRCI	DFT/MRCI(2)	Atom	*N* Roots
899	123	122	C	20
Cytosine				
DFT/MRCI	*p*-DFT/MRCI	DFT/MRCI(2)	Atom	*N* Roots
2690	309	231	C	50
530	69	65	N	20
7062	297	292	O	50
Thymine				
DFT/MRCI	*p*-DFT/MRCI	DFT/MRCI(2)	Structure	*N* Roots
602	87	98	*gs*	4
565	94	128	$n\pi^*$	4
597	87	99	$\pi\pi^*$	4

[c] Values in the table are CPU times (excluding the KS-DFT computation) and correspond to calculations run using a single thread of an Intel Xeon Gold 6130 processor. The aug-cc-pCVTZ basis set was used for ethyl trifluoroacetate, and the aug-cc-pCVDZ basis was used for Cytosine and Thymine.

This computational efficiency is particularly advantageous in the determination of X-ray spectra given the large number of core-excited or core-ionized states typically involved. For the ethyl trifluoroacetate XPS simulation, 20 carbon 1*s* core-ionized states were computed. For cytosine, 50, 20, and 50 roots are computed for the nitrogen, oxygen, and carbon K-edges, respectively. The number of roots computed in each case roughly scales as the density of states within NEXAFS region of the XAS and relates to the extent of the NEXAFS region. In the cytosine carbon K-edge NEXAFS, for example, there exist 50 core-excited states within the 284 eV to 291 eV energy range shown in Figure 6. We note that, for the calculation of large numbers of states, the DFT/MRCI approach, being a selected CI method, has a certain advantage. Namely, the reference space is tailored to the states under consideration and is kept optimally small. This, in combination with the energy-based configuration selection criterion, limits the growth of the FOIS CSF basis with increasing numbers of states. This makes the DFT/MRCI approach particularly well suited to the calculation X-ray spectra, for which dense manifolds of core-excited states spanning many eV typically have to be considered.

VI. CONCLUSION

In this work we report a new DFT/MRCI Hamiltonian, termed CVS-QE12, tailored for the calculation of accurate K-edge core-excitation and ionization energies for organic molecules. In line with a companion publication,[57] the QTP17 XC DFT functional has been employed to generate the underlying KS orbital basis; so adopted for its ability to generate orbital energies that enable a balanced description of both core and valence excitation energies. Furthermore, the current work shares with Reference 57 a core design philosophy in which the fitting of the Hamiltonian parameters is performed so as to reproduce benchmark *ab initio* excitation and ionization energies, here computed at the CVS-EOM-CCSDT and CVS-EOMIP-CCSDT levels of theory, respectively.

The present work also introduces an additional parameter to the Hamiltonian such that the on-diagonal Coulomb integrals involving hole-particle interactions between core and valence orbitals are now scaled separately from the valence-valence integrals. That these interactions should be differentially damped is understood conceptually via the long-recognized dominance of exchange interactions in the highly localized regions around the atomic cores, and is borne out by the dramatic improvement in the accuracy of the core-excitation energies furnished by the approach. The CVS-QE12 Hamiltonian yields vertical core-excitation and ionization energies that exhibit MAEs of ~ 0.5 eV relative to the benchmark values, while maintaining the favorable computational efficiency associated with the DFT/MRCI method. Furthermore, the previously observed ability of the fast *p*-DFT/MRCI and DFT/MRCI(2) variants to be highly accurate approximations to the parent DFT/MRCI method is maintained in the CVS-QE12 Hamiltonian, with both methods furnishing MAEs over the training and validation sets that are within 0.01 eV of the parent DFT/MRCI values.

While the present work focuses exclusively on K-edge core-excited states of second-row atoms in organic molecules, the range of applicability of the method is significantly broader than this specific application. Specifically, the performance of the current parameterization for L-edge spectra, as well as inorganic molecules, such as transition metal complexes, will be investigated in future studies. Given that the definition of "core" electrons becomes more nuanced in these applications, flexible approaches for generating KS orbital basis may become necessary. To this end, recent work demonstrating the ability of projected hybrid density functionals[119,120] to yield accurate core-excitation energies may perhaps provide a route to additional flexibility in the

choice of the XC functional employed to generate the KS-orbital basis, as well as a route to general methods capable of simultaneously furnishing accurate excitation energies of multiple core-levels.

ACKNOWLEDGMENTS

The authors would like to thank Sonia Coriani and Devin Matthews for helpful discussions during the preparation of this work. M.S.S. also acknowledges the Natural Sciences and Engineering Research Council (NSERC) Discovery grant program for financial support.

VII. SUPPORTING INFORMATION

Full training and validation sets, as well as the nuclear geometries of the molecules employed in representative applications are provided as Supplementary Material.

DATA AVAILABILITY STATEMENT

Any data not included in the published Supporting Information that was generated or employed by this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] C. Pellegrini, "The development of xfels," Nature Reviews Physics **2**, 330–331 (2020).

[2] R. Schoenlein, T. Elsaesser, K. Holldack, Z. Huang, H. Kapteyn, M. Murnane, and M. Woerner, "Recent advances in ultrafast x-ray sources," Philosophical Transactions of the Royal Society A **377**, 20180384 (2019).

[3] P. Norman and A. Dreuw, "Simulating x-ray spectroscopies and calculating core-excited states of molecules," Chemical reviews **118**, 7208–7248 (2018).

[4] C. D. Rankine and T. J. Penfold, "Progress in the theory of x-ray spectroscopy: From quantum chemistry to machine learning and ultrafast dynamics," The Journal of Physical Chemistry A **125**, 4276–4293 (2021).

[5] L. S. Cederbaum, W. Domcke, and J. Schirmer, "Many-body theory of core holes," Physical Review A **22**, 206 (1980).

[6] N. A. Besley, "Density functional theory based methods for the calculation of x-ray spectroscopy," Accounts of Chemical Research **53**, 1306–1315 (2020).

[7] M. Stener, G. Fronzoni, and M. d. de Simone, "Time dependent density functional theory of core electrons excitations," Chemical physics letters **373**, 115–123 (2003).

[8] J. Rehr and A. Ankudinov, "Time-dependent density functional theory calculations of x-ray absorption," International journal of quantum chemistry **95**, 487–492 (2003).

[9] Y. Imamura and H. Nakai, "Time-dependent density functional theory (tddft) calculations for core-excited states: Assessment of an exchange functional combining the becke88 and van leeuwen–baerends-type functionals," Chemical Physics Letters **419**, 297–303 (2006).

[10] N. A. Besley and F. A. Asmuruf, "Time-dependent density functional theory calculations of the spectroscopy of core electrons," Physical Chemistry Chemical Physics **12**, 12024–12039 (2010).

[11] K. Lopata, B. E. Van Kuiken, M. Khalil, and N. Govind, "Linear-response and real-time time-dependent density functional theory studies of core-level near-edge x-ray absorption," Journal of chemical theory and computation **8**, 3284–3292 (2012).

[12] X. Li, N. Govind, C. Isborn, A. E. DePrince III, and K. Lopata, "Real-time time-dependent electronic structure theory," Chemical Reviews **120**, 9951–9993 (2020).

[13] R. G. Fernando, M. C. Balhoff, and K. Lopata, "X-ray absorption in insulators with non-hermitian real-time time-dependent density functional theory," Journal of chemical theory and computation **11**, 646–654 (2015).

[14] M. Kadek, L. Konecny, B. Gao, M. Repisky, and K. Ruud, "X-ray absorption resonances near l 2, 3-edges from real-time propagation of the dirac–kohn–sham density matrix," Physical Chemistry Chemical Physics **17**, 22566–22570 (2015).

[15] J. M. Kasper, P. J. Lestrange, T. F. Stetina, and X. Li, "Modeling l2, 3-edge x-ray absorption spectroscopy with real-time exact two-component relativistic time-dependent density functional theory," Journal of Chemical Theory and Computation **14**, 1998–2006 (2018).

[16] F. A. Evangelista, P. Shushkov, and J. C. Tully, "Orthogonality constrained density functional theory for electronic excited states," The Journal of Physical Chemistry A **117**, 7378–7392 (2013).

[17] W. D. Derricotte and F. A. Evangelista, "Simulation of x-ray absorption spectra with orthogonality constrained density functional theory," Physical Chemistry Chemical Physics **17**, 14360–14374 (2015).

[18] D. Hait and M. Head-Gordon, "Highly accurate prediction of core spectra of molecules at density functional theory cost: Attaining sub-electronvolt error from a restricted open-shell kohn–sham approach," The journal of physical chemistry letters **11**, 775–786 (2020).

[19] D. Hait and M. Head-Gordon, "Orbital optimized density functional theory for electronic excited states," The journal of physical chemistry letters **12**, 4517–4529 (2021).

[20] D. Hait, K. J. Oosterbaan, K. Carter-Fenk, and M. Head-Gordon, "Computing x-ray absorption spectra from linear-response particles atop optimized holes," The Journal of chemical physics **156** (2022).

[21] J. E. Arias-Martinez, L. A. Cunha, K. J. Oosterbaan, J. Lee, and M. Head-Gordon, "Accurate core excitation and ionization energies from a state-specific coupled-cluster singles and doubles approach," Physical Chemistry Chemical Physics **24**, 20728–20741 (2022).

[22] L. A. Cunha, D. Hait, R. Kang, Y. Mao, and M. Head-Gordon, "Relativistic orbital-optimized density functional theory for accurate core-level spectroscopy," The journal of physical chemistry letters **13**, 3438–3449 (2022).

[23] D. Hait and M. Head-Gordon, "Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory," Journal of chemical theory and computation **16**, 1699–1710 (2020).

[24] J. Cullen, M. Krykunov, and T. Ziegler, "The formulation of a self-consistent constricted variational density functional theory for the description of excited states," Chemical Physics **391**, 11–18 (2011).

[25] A. Barth and J. Schirmer, "Theoretical core-level excitation spectra of n2 and co by a new polarisation propagator method," Journal of Physics B: Atomic and Molecular Physics **18**, 867 (1985).

[26] J. Wenzel, M. Wormit, and A. Dreuw, "Calculating core-level excitations and x-ray absorption spectra of medium-sized closed-shell molecules with the algebraic-diagrammatic construction scheme for the polarization propagator," Journal of computational chemistry **35**, 1900–1915 (2014).

[27] J. Wenzel, M. Wormit, and A. Dreuw, "Calculating x-ray absorption spectra of open-shell molecules with the unrestricted algebraic-diagrammatic construction scheme for the polarization propagator," Journal of chemical theory and computation **10**, 4583–4598 (2014).

[28] J. Wenzel, A. Holzer, M. Wormit, and A. Dreuw, "Analysis and comparison of cvs-adc approaches up to third order for the calculation of core-excited states," The Journal of Chemical Physics **142** (2015).

[29] B. Peng, P. J. Lestrange, J. J. Goings, M. Caricato, and X. Li, "Energy-specific equation-of-motion coupled-cluster methods for high-energy excited states: Application to k-edge x-ray absorption spectroscopy," Journal of chemical theory and computation **11**, 4146–4153 (2015).

[30] S. Coriani and H. Koch, "Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework," The Journal of chemical physics **143** (2015).

[31] J. P. Carbone, L. Cheng, R. H. Myhre, D. Matthews, H. Koch, and S. Coriani, "An analysis of the performance of coupled cluster methods for k-edge core excitations and ionizations using standard basis sets," in *Advances in Quantum Chemistry*, Vol. 79 (Elsevier, 2019) pp. 241–261.

[32] D. A. Matthews, "Eom-cc methods with approximate triple excitations applied to core excitation and ionisation energies," Molecular Physics **118**, e1771448 (2020).

[33] M. Simons and D. A. Matthews, "Transition-potential coupled cluster," The Journal of Chemical Physics **154**, 014106 (2021).

[34] M. Simons and D. A. Matthews, "Accurate core-excited states via inclusion of core triple excitations in similarity-transformed equation-of-motion theory," Journal of Chemical Theory and Computation **18**, 3759–3765 (2022).

[35] M. Simons and D. A. Matthews, "Transition-potential coupled cluster ii: Optimisation of the core orbital occupation number," Molecular Physics **121**, e2088421 (2023).

[36] S. Sorensen, X. Zheng, S. Southworth, M. Patanen, E. Kokkonen, B. Oostenrijk, O. Travnikova, T. Marchenko, M. Simon, C. Bostedt, *et al.*, "From synchrotrons for xfels: the soft x-ray near-edge spectrum of the esca molecule," Journal of Physics B: Atomic, Molecular and Optical Physics **53**, 244011 (2020).

[37] M. L. Vidal, A. I. Krylov, and S. Coriani, "Dyson orbitals within the fc-cvs-eom-ccsd framework: theory and application to x-ray photoelectron spectroscopy of ground and excited states," Physical Chemistry Chemical Physics **22**, 2693–2703 (2020).

[38] M. L. Vidal, P. Pokhilko, A. I. Krylov, and S. Coriani, "Equation-of-motion coupled-cluster theory to model l-edge x-ray absorption and photoelectron spectra," The journal of physical chemistry letters **11**, 8314–8321 (2020).

[39] J. Liu, D. Matthews, S. Coriani, and L. Cheng, "Benchmark calculations of k-edge ionization energies for first-row elements using scalar-relativistic core-valence-separated equation-of-motion coupled-cluster methods," Journal of chemical theory and computation **15**, 1642–1651 (2019).

[40] S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, "Coupled-cluster theory for excited electronic states: The full equation-of-motion coupled-cluster single, double, and triple excitation method," The Journal of Chemical Physics **115**, 8263–8266 (2001).

[41] R. H. Myhre, S. Coriani, and H. Koch, "Near-edge x-ray absorption fine structure within multilevel coupled cluster theory," Journal of chemical theory and computation **12**, 2633–2643 (2016).

[42] B. O. Roos, P. R. Taylor, and P. E. Sigbahn, "A complete active space scf method (casscf) using a density matrix formulated super-ci approach," Chemical Physics **48**, 157–173 (1980).

[43] P. Å. Malmqvist, A. Rendell, and B. O. Roos, "The restricted active space self-consistent-field method, implemented with a split graph unitary group approach," Journal of Physical Chemistry **94**, 5477–5482 (1990).

[44] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, "Second-order perturbation theory with a casscf reference function," Journal of Physical Chemistry **94**, 5483–5488 (1990).

[45] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, "Second-order perturbation theory with a complete active space self-consistent field reference function," The Journal of chemical physics **96**, 1218–1226 (1992).

[46] P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, "The restricted active space followed by second-order perturbation theory method: Theory and application to the study of cuo2 and cu2o2 systems," The Journal of chemical physics **128** (2008).

[47] I. Josefsson, K. Kunnus, S. Schreck, A. Föhlisch, F. de Groot, P. Wernet, and M. Odelius, "Ab initio calculations of x-ray spectra: Atomic multiplet and molecular orbital effects in a multiconfigurational scf approach to the l-edge spectra of transition metal complexes," The journal of physical chemistry letters **3**, 3565–3570 (2012).

[48] R. V. Pinjari, M. G. Delcey, M. Guo, M. Odelius, and M. Lundberg, "Cost and sensitivity of restricted active-space calculations of metal l-edge x-ray absorption spectra," Journal of computational chemistry **37**, 477–486 (2016).

[49] C. E. de Moura and A. Y. Sokolov, "Simulating x-ray photoelectron spectra with strong electron correlation using multireference algebraic diagrammatic construction theory," Physical Chemistry Chemical Physics **24**, 4769–4784 (2022).

[50] C. E. de Moura and A. Y. Sokolov, "Correction: Simulating x-ray photoelectron spectra with strong electron correlation using multireference algebraic diagrammatic construction theory," Physical Chemistry Chemical Physics **24**, 8041–8046 (2022).

[51] D. Maganas, J. K. Kowalska, M. Nooijen, S. DeBeer, and F. Neese, "Comparison of multireference ab initio wavefunction methodologies for x-ray absorption edges: A case study on [fe (ii/iii) cl4] 2–/1–molecules," The Journal of Chemical Physics **150** (2019).

[52] C. M. Marian, A. Heil, and M. Kleinschmidt, "The dft/mrci method," WIREs Computational Molecular Science **9**, e1394 (2019), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1394.

[53] I. Lyskov, M. Kleinschmidt, and C. M. Marian, "Redesign of the dft/mrci hamiltonian," The Journal of chemical physics **144**, 034104 (2016).

[54] A. Heil and C. M. Marian, "Dft/mrci hamiltonian for odd and even numbers of electrons," The Journal of Chemical Physics **147**, 194104 (2017).

[55] A. Heil, M. Kleinschmidt, and C. M. Marian, "On the performance of dft/mrci hamiltonians for electronic excitations in transition metal complexes: The role of the damping function," The Journal of Chemical Physics **149**, 164106 (2018).

[56] D. R. Dombrowski, T. Schulz, M. Kleinschmidt, and C. M. Marian, "R2022: A dft/mrci ansatz with improved performance for double excitations," The Journal of Physical Chemistry A **127**, 2011–2025 (2023), pMID: 36799533, https://doi.org/10.1021/acs.jpca.2c07951.

[57] T. S. Costain, V. Ogden, S. P. Neville, and M. S. Schuurman, "A dft/mrci hamiltonian parameterized using only ab initio data: I. valence excited states," The Journal of Chemical Physics **160**, 224106 (2024), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0210897/19985974/224106_1_5.0210897.pdf.

[58] I. Seidu, S. P. Neville, M. Kleinschmidt, A. Heil, C. M. Marian, and M. S. Schuurman, "The simulation of x-ray absorption spectra from ground and excited electronic states using core-valence separated dft/mrci," The Journal of Chemical Physics **151**, 144104 (2019).

[59] S. Grimme and M. Waletzke, "A combination of kohn–sham density functional theory and multi-reference configuration interaction methods," The Journal of chemical physics **111**, 5645–5655 (1999).

[60] R. W. Wetmore and G. A. Segal, "?" Chem. Phys. Lett. **36**, 478 (1975).

[61] G. A. Segal, R. W. Wetmore, and K. Wolf, "Efficient methods for configuration interaction calculations," Chemical Physics **30**, 269–297 (1978).

[62] J. D. Wadey and N. A. Besley, "Quantum chemical calculations of x-ray emission spectroscopy," Journal of Chemical Theory and Computation **10**, 4557–4564 (2014).

[63] A. Nakata, Y. Imamura, T. Otsuka, and H. Nakai, "Time-dependent density functional theory calculations for core-excited states: Assessment of standard exchange-correlation functionals and development of a novel hybrid functional," The Journal of Chemical Physics **124**, 094105 (2006), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2173987/15379376/094105_1_online.pdf.

[64] N. A. Besley and A. Noble, "Time-dependent density functional theory study of the x-ray absorption spectroscopy of acetylene, ethylene, and benzene on si(100)," The Journal of Physical Chemistry C **111**, 3333–3340 (2007), https://doi.org/10.1021/jp065160x.

[65] J. Noga and R. J. Bartlett, "The full CCSDT model for molecular electronic structure," The Journal of Chemical Physics **86**, 7041–7050 (1987), https://pubs.aip.org/aip/jcp/article-pdf/86/12/7041/18964408/7041_1_online.pdf.

[66] G. E. Scuseria and H. F. Schaefer, "A new implementation of the full ccsdt model for molecular electronic structure," Chemical Physics Letters **152**, 382–386 (1988).

[67] R. A. Kendall, J. Dunning, Thom H., and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," The Journal of Chemical Physics **96**, 6796–6806 (1992), https://pubs.aip.org/aip/jcp/article-pdf/96/9/6796/18998924/6796_1_online.pdf.

[68] D. E. Woon and J. Dunning, Thom H., "Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties," The Journal of Chemical Physics **100**, 2975–2988 (1994), https://pubs.aip.org/aip/jcp/article-pdf/100/4/2975/19303178/2975_1_online.pdf.

[69] M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. Mezei, and Á. Ganyecz, "The mrcc program system: Accurate quantum chemistry from water to proteins," The Journal of chemical physics **152**, 074107 (2020).

[70] M. Kállay, P. R. Nagy, D. Mester, L. Gyevi-Nagy, J. Csóka, P. B. Szabó, Z. Rolik, G. Samu, J. Csontos, B. Hégely, Á. Ganyecz, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. Mezei, and R. Horváth, "Mrcc, a quantum chemical program suite written by m. kallay et. al," J Chem Phys **139**, 094105 (2013).

[71] M. Kállay and J. Gauss, "Calculation of excited-state properties using general coupled-cluster and configuration-interaction models," The Journal of chemical physics **121**, 9257–9269 (2004).

[72] P.-F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, "A mountaineering strategy to excited states: Highly accurate reference energies and benchmarks," Journal of chemical theory and computation **14**, 4360–4379 (2018).

[73] T. Fransson, I. E. Brumboiu, M. L. Vidal, P. Norman, S. Coriani, and A. Dreuw, "Xaboom: An x-ray absorption benchmark of organic molecules based on carbon, nitrogen, and oxygen 1s$\rightarrow \pi^*$ transitions," Journal of Chemical Theory and Computation **17**, 1618–1637 (2021).

[74] H. Koch, O. Christiansen, P. Jo/rgensen, A. M. Sanchez de Merás, and T. Helgaker, "The CC3 model: An iterative coupled cluster approach including connected triples," The Journal of Chemical Physics **106**, 1808–1818 (1997), https://pubs.aip.org/aip/jcp/article-pdf/106/5/1808/19128978/1808_1_online.pdf.

[75] S. Neville and M. Schuurman, "GRaCI: General Reference Configuration Interaction," (2021).

[76] Q. Sun, "Libcint: An efficient general integral library for g aussian basis functions," Journal of computational chemistry **36**, 1664–1671 (2015).

[77] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, "Pyscf: the python-based simulations of chemistry framework," Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).

[78] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, "Recent developments in the pyscf program package," The Journal of chemical physics **153** (2020).

[79] M. A. Marques, M. J. Oliveira, and T. Burnus, "Libxc: A library of exchange and correlation functionals for density functional theory," Computer physics communications **183**, 2272–2281 (2012).

[80]S. Lehtola, C. Steigemann, M. J. Oliveira, and M. A. Marques, "Recent developments in libxc—a comprehensive library of functionals for density functional theory," SoftwareX **7**, 1–5 (2018).

[81]F. Weigend, "A fully direct ri-hf algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency," Phys. Chem. Chem. Phys. **4**, 4285–4291 (2002).

[82]S. P. Neville and M. S. Schuurman, "Removing the deadwood from dft/mrci wave functions: The p-dft/mrci method," Journal of Chemical Theory and Computation **17**, 7657–7665 (2021).

[83]U. Gelius, E. Basilier, S. Svensson, T. Bergmark, and K. Siegbahn, "A high resolution esca instrument with x-ray monochromator for gases and solids," Journal of Electron Spectroscopy and Related Phenomena **2**, 405–434 (1973).

[84]V. Feyer, O. Plekan, R. Richter, M. Coreno, G. Vall-Llosera, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, T. E. Moskovskaya, E. V. Gromov, *et al.*, "Tautomerism in cytosine and uracil: An experimental and theoretical core level spectroscopic study," The Journal of Physical Chemistry A **113**, 5736–5742 (2009).

[85]V. Feyer, O. Plekan, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, I. L. Zaytseva, and J. Schirmer, "Tautomerism in cytosine and uracil: A theoretical and experimental x-ray absorption and resonant auger study," The Journal of Physical Chemistry A **114**, 10270–10276 (2010).

[86]T. Wolf, R. H. Myhre, J. Cryan, S. Coriani, R. Squibb, A. Battistoni, N. Berrah, C. Bostedt, P. Bucksbaum, G. Coslovich, *et al.*, "Probing ultrafast $\pi\pi^*$/nπ^* internal conversion in organic chromophores via k-edge resonant absorption," Nature communications **8**, 29 (2017).

[87]A. I. Krylov and P. M. Gill, "Q-chem: an engine for innovation," Wiley Interdisciplinary Reviews: Computational Molecular Science **3**, 317–326 (2013).

[88]M. F. Herbst, M. Scheurer, T. Fransson, D. R. Rehn, and A. Dreuw, "adcc: A versatile toolkit for rapid development of algebraic-diagrammatic construction methods," Wiley Interdisciplinary Reviews: Computational Molecular Science **10**, e1462 (2020).

[89]B. T. Pickup, "On the theory of fast photoionization processes," Chemical Physics **19**, 193–208 (1977).

[90]Y. Öhrn and G. Born, "Molecular electron propagator theory and calculations," (Academic Press, 1981) pp. 1–88.

[91]M. Spanner, S. Patchkovskii, C. Zhou, S. Matsika, M. Kotur, and T. C. Weinacht, "Dyson norms in xuv and strong-field ionization of polyatomics: Cytosine and uracil," Phys. Rev. A **86**, 053406 (2012).

[92]F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand, and L. González, "Efficient and flexible computation of many-electron wave function overlaps," Journal of Chemical Theory and Computation **12**, 1207–1219 (2016), pMID: 26854874, https://doi.org/10.1021/acs.jctc.5b01148.

[93]N. A. Besley, M. J. Peach, and D. J. Tozer, "Time-dependent density functional theory calculations of near-edge x-ray absorption fine structure with short-range corrected functionals," Physical Chemistry Chemical Physics **11**, 10350–10358 (2009).

[94]Y. Jin and R. J. Bartlett, "Accurate computation of x-ray absorption spectra with ionization potential optimized global hybrid functional," The Journal of Chemical Physics **149**, 064111 (2018).

[95]T. Koopmans, "Ordering of wave functions and eigenenergies to the individual electrons of an atom," Physica **1**, 104–113 (1933).

[96]D. W. Smith and O. W. Day, "Extension of koopmans' theorem. i. derivation," The Journal of Chemical Physics **62**, 113–114 (1975).

[97]O. W. Day, D. W. Smith, and R. C. Morrison, "Extension of koopmans' theorem. ii. accurate ionization energies from correlated wavefunctions for closed-shell atoms," The Journal of Chemical Physics **62**, 115–119 (1975).

[98]R. C. Morrison, O. W. Day, and D. W. Smith, "An extension of koopmans' theorem iii. ionization energies of the open-shell atoms li and b," International Journal of Quantum Chemistry **9**, 229–235 (1975).

[99]J. C. Ellenbogen, O. W. Day, D. W. Smith, and R. C. Morrison, "Extension of koopmans' theorem. iv. ionization potentials from correlated wavefunctions for molecular fluorine," The Journal of Chemical Physics **66**, 4795–4801 (1977).

[100]J. F. Janak, "Proof that $\frac{\partial e}{\partial n_i} = \varepsilon$ in density-functional theory," Physical Review B **18**, 7165 (1978).

[101]M. K. Harbola, "Relationship between the highest occupied kohn-sham orbital eigenvalue and ionization energy," Physical Review B **60**, 4545 (1999).

[102]P. Verma and R. J. Bartlett, "Increasing the applicability of density functional theory. v. x-ray absorption spectra with ionization potential corrected exchange and correlation potentials," The Journal of chemical physics **145** (2016).

[103]Y. Jin and R. J. Bartlett, "The qtp family of consistent functionals and potentials in kohn-sham density functional theory," The Journal of Chemical Physics **145** (2016).

[104]J. T. Margraf, P. Verma, and R. J. Bartlett, "Ionization potential optimized double-hybrid density functional approximations," The Journal of Chemical Physics **145** (2016).

[105]T. M. Maier, H. Bahmann, A. V. Arbuznikov, and M. Kaupp, "Validation of local hybrid functionals for tddft calculations of electronic excitation energies," The Journal of Chemical Physics **144** (2016).

[106]A. Nakata, Y. Imamura, and H. Nakai, "Hybrid exchange-correlation functional for core, valence, and rydberg excitations: Core-valence-rydberg b3lyp," The Journal of chemical physics **125** (2006).

[107]A. Nakata, Y. Imamura, and H. Nakai, "Extension of the core-valence-rydberg b3lyp functional to core-excited-state calculations of third-row atoms," Journal of Chemical Theory and Computation **3**, 1295–1305 (2007), pMID: 26633203, https://doi.org/10.1021/ct600368f.

[108]P. Verma and R. J. Bartlett, "Increasing the applicability of density functional theory. iv. consequences of ionization-potential improved exchange-correlation potentials," The Journal of chemical physics **140** (2014).

[109]A. A. Fouda and N. A. Besley, "Improving the predictive quality of time-dependent density functional theory calculations of the x-ray emission spectroscopy of organic molecules," Journal of Computational Chemistry **41**, 1081–1090 (2020).

[110]B. G. Janesko, "Replacing hybrid density functional theory: motivation and recent advances," Chemical Society Reviews **50**, 8470–8495 (2021).

[111]S. P. Neville and M. S. Schuurman, "A perturbative approximation to dft/mrci: Dft/mrci (2)," The Journal of Chemical Physics **157**, 164103 (2022).

[112]A. ESCA, "Molecular and solid state structure studied by means of electron spectroscopy," Siegbahn, K , 108 (1967).

[113]M. E. Defonsi Lestard, M. E. Tuttolomondo, D. A. Wann, H. E. Robertson, D. W. Rankin, and A. B. Altabef, "Experimental and theoretical structure and vibrational analysis of ethyl trifluoroacetate, cf3co2ch2ch3," Journal of Raman Spectroscopy **41**, 1357–1368 (2010).

[114]O. Travnikova, K. J. Børve, M. Patanen, J. Söderström, C. Miron, L. J. Sæthre, N. Mårtensson, and S. Svensson, "The esca molecule—historical remarks and new results," Journal of Electron Spectroscopy and Related Phenomena **185**, 191–197 (2012).

[115]S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, J. Šponer, P. Slavíček, and P. Hobza, "Correlated ab initio study of nucleic acid bases and their tautomers in the gas phase, in a microhydrated environment and in aqueous solution part 1. cytosine," Physical Chemistry Chemical Physics **4**, 4192–4203 (2002).

[116]B. McFarland, J. Farrell, S. Miyabe, F. Tarantelli, A. Aguilar, N. Berrah, C. Bostedt, J. Bozek, P. Bucksbaum, J. Castagna, *et al.*, "Ultrafast x-ray auger probing of photoexcited molecular dynamics," Nature communications **5**, 4235 (2014).

[117]T. J. Wolf, R. M. Parrish, R. H. Myhre, T. J. Martínez, H. Koch, and M. Gühr, "Observation of ultrafast intersystem crossing in thymine by extreme ultraviolet time-resolved photoelectron spectroscopy," The Journal of Physical Chemistry A **123**, 6897–6903 (2019).

[118]T. J. Wolf, A. C. Paul, S. D. Folkestad, R. H. Myhre, J. P. Cryan, N. Berrah, P. H. Bucksbaum, S. Coriani, G. Coslovich, R. Feifel, *et al.*, "Transient resonant auger–meitner spectra of photoexcited thymine," Faraday discussions **228**, 555–570 (2021).

[119]B. G. Janesko, "Projected hybrid density functionals: Method and application to core electron ionization," Journal of Chemical Theory and Computation **19**, 837–847 (2023).

[120]B. G. Janesko, "Core-projected hybrids fix systematic errors in time-dependent density functional theory predicted core-electron excitations," Journal of Chemical Theory and Computation **19**, 5112–5121 (2023).