An Ab Initio Parameterization of the DFT/MRCI Method for Valence and Core Excitation

Teagan Costain

Thesis submitted to the University of Ottawa in partial
fulfillment of the requirements for the degree of
Master of Science, Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Supervisor: Michael S. Schuurman

ABSTRACT

With the latest technological advancements related to spectroscopy, experimental capabilities have become ever more sophisticated. Theoreticians must therefore meet ever-growing demands to analyze the experimental data in time-resolved studies of ultrafast molecular dynamics. An indispensable tool in the theoretician's toolbox is electronic structure theory, the study of methods to describe electronic wave functions. With a focus on the combined density functional theory and multi-reference configuration interaction (DFT/MRCI) method, this thesis aims to highlight the importance of developing semi-empirical methods to achieve balance between the predictive accuracy of purely ab initio methodology and the computational efficiency of empirical methodology. To this end, this thesis presents two new parameterizations of the DFT/MRCI method for valence and core-excited states that, respectively, achieve sub-eV error metrics with respect to benchmark vertical excitation energies.

ACKNOWLEDGEMENTS

Thank you to my supervisor Michael Schuurman for his invaluable support, feedback, and guidance throughout. I give additional thanks to Simon Neville, without whom the present work could not be nearly what it is; to Augusto Gerolin, for broadening my understanding of DFT beyond what I could have accomplished alone; and to all my colleagues, who have taught me so much. Finally, thanks to my family and to my partner for their unconditional support over the years.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Publications

Included as part of this thesis:

1. *A DFT/MRCI Hamiltonian Parameterized Using Only Ab Initio Data: I. Valence Excited States*

   Teagan Shane Costain, Victoria Ogden, Simon P. Neville, and Michael S. Schuurman. *J. Chem. Phys.* 160, 224106 (2024)

   Contribution: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing – review & editing (supporting).

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

   Teagan Shane Costain, Jibrael B. Rolston, Simon P. Neville, and Michael S. Schuurman. *J. Chem. Phys.* 161, 114117 (2024)

   Contribution: Conceptualization (equal); Data curation (lead); Methodology (equal); Project administration (lead); Visualization (lead); Writing – original draft (equal); Writing – review & editing (equal).

1

Background Information

1.1 INTRODUCTION

The study of electronic excitation and excited-state processes is driven by photochemistry, which plays a key role in diverse fields, from biology and medicine to material science. Photochemistry is the study of light-matter interaction and is crucial to understanding biological processes such as photosynthesis,[1] visual perception,[2] and photoprotective mechanisms in DNA.[3–6] This biological focus lends itself to medicinal applications, such as photodynamic therapy.[7] In material science, classic applications of photochemistry include photovoltaic cells and organic light-emitting diode (OLED) displays. New and emerging research leads to increasingly diverse applications across optics, photonics, and optoelectronics.[8] For example, photochromic molecules act as molecular photo-switches, amenable to a variety of applications in photonic devices.[9]

In experiment, ultrashort laser pulses can be tuned to probe, or steer, ultrafast dynamics. Processes such as intramolecular vibrational relaxation and motion, excitation energy transfer, charge transfer, and electronic dephasing, occur on a femto- to picosecond timescale.[10] These processes play an important role in photochemical reactions; photosynthesis depends upon excitation energy transfer via exciton coupling,[1] and visual perception relies on retinal isomerization via ultrafast bond torsion driven by photoexcitation.[2] Spectroscopy is the observational tool used to study these ultrafast processes, effectively taking snapshots in time. However, a time-resolved picture of these dynamics requires laser pulses short enough to resolve the femtosecond timescale of vibronic dynamics, which has historically represented a technological limitation. In photosynthetic complexes, energy transfer occurs within a time frame of 100 fs to hundreds of

picoseconds; the advent of light sources capable of sub-50 fs pulses has allowed for the experimental resolution of this process.[11]

Time-resolved studies generally apply the pump-probe technique. An initial excitation pulse drives a photoinduced process, and a secondary pulse delivered after a synchronized time delay measures the changes in an optical property. transient absorption (TA) spectroscopy relies on this technique, capturing the time-dependent variation in the absorption spectrum of a sample over the time evolution of the stimulated dynamics. In recent decades, ongoing research and development has greatly expanded research capabilities in the domain of ultrafast dynamics. Progress in optical technology allows the generation of shorter, brighter pulses of tunable wavelength,[12] and novel spectroscopic techniques yield more sophisticated alternatives to classic pump-probe spectroscopy.[10, 13]

As technology has evolved, so too has the definition of "ultrafast". Whereas picosecond timescales were once considered ultrafast, this domain has since been redefined to mean sub-100 femtosecond time scales.[14] X-ray science in particular has gained traction with the availability of fourth-generation light sources at X-ray free-electron laser (XFEL) facilities.[12, 15] Ultrashort, high-energy X-ray pulses benefit from increased temporal resolution and element-wise chemical specificity, as X-rays can interact with core electrons. Local structural information about a molecule can be extracted from X-ray absorption spectra (XAS) due to the large chemical shifts that can develop between core excitations localized on different atomic sites. For example, this resolution makes it possible to elucidate the photoprotective mechanism of thymine, which is resistant to photochemical dimerization. Upon photoexcitation, ultrafast internal conversion to a nonreactive dark state reduces the probability of dimerization, contributing to the photostability of DNA.[3–6] The existence of this internal conversion is demonstrated by the clear,

unambiguous changes in the oxygen K-edge absorption spectrum.[4]

Advancements in the experimental study of ultrafast dynamics naturally motivate further development of the theoretical methods to study these dynamics. Experimental spectra do not directly yield information about the underlying system being studied; these spectra must be interpreted, and they can be analyzed by theoretical simulations. To model time-resolved processes, this simulation must simultaneously describe (i) the nuclear dynamics, and (ii) the electronic structure at each time coordinate. In many cases, vibronic coupling is negligible and these two problems are separable; the effective decoupling of the nuclear and electronic degrees of freedom in this manner is the Born–Oppenheimer (BO) approximation. A dynamics simulation describes the time evolution of atomic motion throughout a chemical process, whilst the electronic structure problem considers the electron density at a time-independent "snapshot" along the reaction coordinate. The solution to this latter problem contains crucial information about the systems under study, representing the focus of the present work.

Electronic structure theory broadly describes the study of methods for solving the Schrödinger equation to determine electronic wave functions. However, electronic states of doubly-excited, multireference, Rydberg, and charge-transfer (CT) character are challenging to describe, as are core-excited states. Attaining an accurate description of valence- and core-excited states of various electronic character generally implies an ab initio, wave function theory (WFT) based approach. However, such approaches are accompanied by a steep trade-off in computational efficiency, limiting the size of systems that can be realistically studied. In particular, molecules relevant to organic photochemistry can become practically impossible to describe using ab initio methods due to their size. Light-harvesting complexes such as carotenoids and chlorophyll are prime examples.[16] Highly accurate wave function-based methods, such as coupled-cluster

(CC), are an ideal, albeit expensive choice.[17] In contrast, the time-dependent density functional theory (TDDFT) method is an excellent alternative to purely ab initio methodology for studying large molecules; yet, it can be unreliable in predicting CT states,[18] meaning that it is not an ideal choice for modeling photosynthetic charge separation. The crux of electronic structure theory is this fundamental trade-off: computational efficiency versus predictive accuracy.

The aim of closing this gap has led to the development of a wide variety of semi-empirical methods, combining desirable aspects of both DFT and WFT-based approaches.[19] The availability of benchmark-quality excited-state energies and properties in the QUEST Databases[20–28] furthermore paves the way for developing highly accurate semi-empirical methods. The density functional theory and multireference configuration interaction (DFT/MRCI) method[29, 30] is one such method that benefits from the availability of these benchmarking data sets. This method has previously been applied, with success, to a wide variety of organic molecules[31–40] and inorganic transition-metal complexes.[41, 42] Many developments of DFT/MRCI have improved on the formulation since its conception,[41, 43–47] yet core-excited states have nevertheless proven challenging,[48] exhibiting errors an order of magnitude larger than that of the valence-excited manifold.

This thesis aims to develop a new parameterization of the DFT/MRCI Hamiltonian that differs from previous ones in a few fundamental aspects. First, only ab initio benchmark quality data is used in the fitting sets, the vast majority of which is available courtesy of the QUEST databases. Past work has had to rely on the use of experimental transition energies,[41, 43, 44] which are vibronic, rather than purely electronic, in nature. Second, this work explicitly investigates the role of DFT in DFT/MRCI. Understanding the features of exchange-correlation (XC) functionals that translate to an improved one-particle basis for DFT/MRCI calculations is key

to achieving consistent and predictable accuracy across electronic states of diverse character. A new XC functional is selected for the parameterizations developed herein, with the goal that new parameterization strategies may henceforth be informed from the analysis presented in this work. The following sections of Chapter 1 lay the foundation for the subject matter of this thesis. Section 1.2 is an overview of DFT, with particular emphasis on aspects that are relevant to the description of core electrons. Section 1.3 is likewise a review of configuration interaction (CI) theory, and its multireference counterpart multireference configuration interaction (MRCI). This section furthermore compares and contrasts the excited-state CI formalism with its DFT counterpart, TDDFT, paving the way for the introduction of DFT/MRCI in Section 1.4. Following this, Chapter 2 presents the results of parameterizing the DFT/MRCI Hamiltonian for valence excitation energies and additionally discusses the selection of the XC functional for a balanced description of valence- and core-excited states. Chapter 3 builds off of the results of Chapter 2, detailing the parameterization strategy for the core excited states. Finally, Chapter 4 presents concluding remarks and proposes future directions for developing the DFT/MRCI method.

1.2 DENSITY FUNCTIONAL THEORY

1.2.1 HOHENBERG–KOHN DENSITY FUNCTIONAL THEORY

The following review of DFT is not an exhaustive exposition of the underlying theory, but an introduction of the essential concepts which shall prove useful in discussing the results of this thesis. For a general introduction to DFT, readers are referred to ref. [49]. For an in-depth mathematical review, ref. [50] is suggested. Finally, for a comprehensive and multidisciplinary overview of developments in DFT, ref. [51] is suggested.

The simplest representation of a wave function of N indistinguishable fermions is that of a single Slater determinant,[52] which is antisymmetric with respect to the exchange of electrons.

$$|\Psi(x_1,...,x_N)\rangle \equiv |\Psi(r_1,s_1,...,r_N,s_N)\rangle \ (x_i=(r_i,s_i),\ r_i\in\mathbb{R},\ s_i\in\alpha,\beta) \tag{1.1}$$

Applying the Rayleigh–Ritz principle,[53] the ground-state wave function of a system is determined variationally over the set of all admissible N-electron wave functions, $\mathscr{W}_N$. Admissible wave functions are normalized to unity, and are eigenfunctions of the electronic Hamiltonian, $\widehat{H}(v)$, yielding a finite expectation value. The electronic Hamiltonian is a function of the one-body potential v, in the set of point-charge Coulomb potentials, $\mathscr{V}_C$.

$$\mathscr{W}_N = \{\Psi \mid \|\Psi\|_2 = 1,\ |\langle\Psi|\widehat{H}(v)|\Psi\rangle| < \infty,\ v\in\mathscr{V}_C,\ \text{N-electron antisymmetric } \Psi\} \tag{1.2}$$

$$\mathscr{V}_C = \{v \mid v(\mathbf{r}) = -\sum_{I=1}^{N_C} Z_I|\mathbf{r}-\mathbf{R_I}|^{-1},\ N_C\in\mathbb{N},\ Z_I\in\mathbb{R}\} \tag{1.3}$$

The electronic Hamiltonian of an N-electron system (in atomic units) may be generally expressed as:

$$\begin{aligned}\widehat{H}_N = &\ \widehat{T} + \widehat{V}_{ee} + \widehat{V}_{ext} \\ = &-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\mathbf{r_i}-\mathbf{r_j}|} + \sum_{i=1}^{N} v_{ext}(\mathbf{r_i})\end{aligned} \tag{1.4}$$

$$v_{ext}(\mathbf{r_i}) = -\frac{1}{2}\sum_{k=1}^{M}\frac{Z_k}{|\mathbf{r_i}-\mathbf{R_k}|} \tag{1.5}$$

The total electronic energy is expressed in terms of the kinetic energy operator ($\widehat{T}$), the two-body electrostatic (Coulomb) repulsion between electrons ($\widehat{V}_{ee}$), and the electrostatic attraction to nuclei, expressed in terms of a one-body external potential ($\widehat{V}_{ext}$).[a] In DFT, the Hamiltonian is often expressed as an expectation value of the following form:

$$\begin{aligned}\langle\Psi|\widehat{H}_N(v)|\Psi\rangle &= \langle\Psi|\widehat{T}+\widehat{V}_{ee}|\Psi\rangle + \int d^3\mathbf{r}\, v(\mathbf{r})\rho(\mathbf{r}) \\ &= F(\rho) + \int d^3\mathbf{r}\, v(\mathbf{r})\rho(\mathbf{r})\end{aligned} \tag{1.6}$$

This form of the equation (Eqn. 1.6) introduces the Hohenberg–Kohn functional, $F(\rho)$. Within the Levy-Lieb constrained-search formalism,[54, 55] the Hohenberg–Kohn functional is defined as a minimization over single-electron densities, ρ.

$$F_{LL}(\rho) = \min_{\Psi\mapsto\rho}\langle\Psi|\widehat{T}+\widehat{V}_{ee}|\Psi\rangle,\ \Psi\in\mathscr{W}_N \tag{1.7}$$

[a]The internuclear repulsion is a constant at a fixed nuclear geometry (i.e., within the Born–Oppenheimer Approximation). This term is omitted from the electronic Hamiltonian, although it is typically included in the total energy.

Importantly, this form (Eqn. 1.7) of the functional recasts the minimization problem in terms of a density ρ which *maps to* the many-electron wave function, Ψ. While a wave function Φ defined in a given orbital basis is not equivalent to Ψ, the approximate wave function Φ is constrained to reproduce the density of the *exact* many-electron wave function, Ψ. The ground-state minimization problem in terms of the Levy-Lieb functional is defined as a minimization over N-representable densities:

$$\begin{aligned} E_N(v) &= \min_{\Psi \in \mathscr{W}_N} \langle \Psi | \widehat{H} | \Psi \rangle \\ &= \inf_{\rho \in \mathscr{I}_N} \{ \min_{\rho \mapsto \Psi} \langle \Psi | \widehat{H} | \Psi \rangle \} \\ &= \inf_{\rho \in \mathscr{I}_N} \{ F_{LL}(\rho) + \int d^3 v(\mathbf{r}) \rho(\mathbf{r}) \} \end{aligned} \tag{1.8}$$

$$\mathscr{I}_N = \{ \rho \in L^1(\mathbb{R}^3) | \, \rho \geq 0, \int \rho(\mathbf{r}) d\mathbf{r} = N, \rho^{1/2} \in \mathrm{H}_1^1 \} \tag{1.9}$$

Within the Hohenberg–Kohn (HK) formulation,[56], the Hamiltonian is expressed as a function of the external potential, v, and the Hohenberg–Kohn functional ($F(\rho)$) as a function of the density, ρ. The ρ-representable potential corresponds to an N-electron ground-state wave function, Ψ_v. The v-representable ground-state density, ρ, constitutes the unique ground-state solution in the associated ground-state potential, v. In practice, an explicit formula for deriving such a ground-state potential does not exist, nor does an explicit method to derive the kinetic energy directly from the density of a fully interacting system. Hohenberg–Kohn's theory, while formally exact, lacks a practical route to its implementation. The work of Levy and Lieb provided a rigorous theoretical backing to HK-DFT, yet the question of how to implement such a method is unclear without the knowledge of an exact XC functional. The Kohn–Sham (KS), or (more

rarely) Hohenberg–Kohn–Sham (HKS) approach[57] supersedes HK theory, providing a route to solving the Schrödinger equation self-consistently. A technical description of Kohn–Sham theory follows.

1.2.2 THE ADIABATIC CONNECTION

Kohn–Sham density functional theory (KS-DFT)[57] links a real, interacting system of electrons with a fictitious, non-interacting reference system with the same density as the real one. The exact exchange-correlation (XC) functional connects these systems, recovering part of the interacting kinetic energy and the many-electron exchange-correlation effects within the term $E_{xc}(\rho)$. The adiabatic connection theorem[58] is an alternate route to explaining Kohn–Sham (KS) theory, and it facilitates the discussion of the following section. The adiabatic connection employs a coupling constant, λ, which explicitly connects the fully interacting N-electron system to its non-interacting counterpart. These systems correspond to coupling strengths of $\lambda = 1$ (fully-interacting) and $\lambda = 0$ (non-interacting), respectively. The Hohenberg–Kohn functional is expressed in terms of λ as:

$$F_{HK}^{\lambda}(\rho) = \min_{\Psi_\lambda \mapsto \rho} \langle \Psi_\lambda | \widehat{T} + \lambda \widehat{V}_{ee} | \Psi_\lambda \rangle \tag{1.10}$$

The Kohn–Sham functional is defined in terms of a non-interacting kinetic energy, $T_s(\rho)$, and a coupled interaction term, W. This interaction term contains the canonical Coulombic interaction term, denoted $J(\rho)$, as well as the exchange-correlation energy, $E_{xc}^{\lambda}(\rho)$. This latter term captures

all the missing kinetic, exchange, and correlation energy of the real, interacting system.

$$\begin{aligned} F_{KS}^{\lambda}(\rho) &= T_s(\rho) + \lambda W(\rho) \\ &= T_s(\rho) + \lambda (J(\rho) + E_{xc}^{\lambda}(\rho)) \end{aligned} \tag{1.11}$$

$$F_{KS}(\rho) = T_s(\rho) + J(\rho) + E_{xc}(\rho) \tag{1.12}$$

$$\begin{aligned} E_{xc}^{\lambda}(\rho) &= E_x(\rho) + E_c^{\lambda}(\rho) \\ &= F_{KS}^{\lambda}(\rho) - T_s(\rho) - \lambda J(\rho) \end{aligned} \tag{1.13}$$

The Kohn–Sham functional is provided in Eqn. 1.11 in terms of the adiabatic connection, and in Eqn. 1.12 in its general form ($\lambda = 1$). The XC functional may be decomposed into its exchange and correlation contributions, as in Eqn. 1.13, representing the difference between the real, *interacting* system and the non-interacting reference system. As before, there is no known relation to derive the kinetic energy of the fully interacting system from the density. Rather, the non-interacting kinetic energy is computed by applying the kinetic energy operator $\widehat{T}$ (see Eqn. 1.4):

$$\begin{aligned} T_s(\rho) &= F^0(\rho) \\ &= \min_{\Psi \mapsto \rho} \langle \Psi | \widehat{T} | \Psi \rangle \end{aligned} \tag{1.14}$$

The Coulombic interaction term $J(\rho)$ is the standard two-body "Hartree" energy.

$$J(\rho) = \frac{1}{2} \int \int \frac{\rho(\mathbf{r_1})\rho(\mathbf{r_2})}{r_{12}} d\mathbf{r_1} d\mathbf{r_2} \tag{1.15}$$

The exchange-correlation energy in Eqn. 1.11 can also be presented in the following form:

$$E_{xc}^{\lambda}(\rho) = \int_0^1 V_{xc}^{\lambda}(\rho) d\lambda \tag{1.16}$$

Here, the potential of the interacting and non-interacting systems are explicitly linked by integration over the coupling strength parameter, λ. While the exact XC functional is unknown, various density functional approximations (DFAs) exist to describe *approximate* XC functionals. The design of these functionals is often application-specific, where a given functional is developed to reproduce particular physical properties for a class of systems. The subject of XC functionals is revisited in Sec. 1.2.4. In the following section, coordinate scaling is introduced as it relates to the adiabatic connection, and subsequently to the nature of approximate XC functionals.

1.2.3 THE HIGH-DENSITY (WEAKLY-CORRELATED) LIMIT

In the applications considered here, the properties of both core and valence electrons are of interest. However, a simultaneous description of either is challenging given the very different correlation regimes of the localized core electrons and delocalized valence electrons. Furthermore, standard XC functionals are parameterized to describe properties related to the valence-electron density. Such properties do not strongly depend on an accurate description of the core-electron density. To this end, we consider the following limiting case: the high-density (or weak-interaction) limit. The localized core electrons are better described in this regime, where electron-electron interactions are negligible and kinetic energy dominates. Consider a density on $\mathbb{R}^d$ dilated by a factor of α, where $d = 3$ for three-dimensional coordinates. Applying coordinate-scaling,[59–

62] this dilated density is described as:

$$\rho_\alpha(\mathbf{r}) = \alpha^d \rho(\alpha\mathbf{r}),\ \int_{\mathbb{R}^d} \rho_\alpha(\mathbf{r})d\mathbf{r} = \int_{\mathbb{R}^d} \rho(\mathbf{r})d\mathbf{r} = N \tag{1.17}$$

Where we note that:

$$\Psi(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}) \mapsto \rho(\mathbf{r}) \tag{1.18}$$

$$\Psi_\alpha(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}) \mapsto \alpha^d \rho(\alpha\mathbf{r}) \tag{1.19}$$

Scaling the wave function of the density $\rho(\mathbf{r})$ does not necessarily yield a wave function which maps to the scaled density, $\rho_\alpha(\mathbf{r})$,[59] meaning that coordinate scaling does not necessarily commute with constrained search:

$$\alpha^{dN/2}\Psi(\alpha\mathbf{r_1}, \alpha\mathbf{r_2}, ..., \alpha\mathbf{r_N}) \neq \Psi_\alpha(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}),\ \alpha \neq 1 \tag{1.20}$$

However, the wave function which maps to ρ satisfying Eqn 1.10, such that $\alpha = \lambda^{-1}$, does satisfy the scaling relation[59, 61, 63]:

$$\alpha^{dN/2}\Psi_\lambda(\alpha\mathbf{r_1}, \alpha\mathbf{r_2}, ..., \alpha\mathbf{r_N}) = \Psi_\alpha(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N}),\ \alpha = \lambda^{-1} \tag{1.21}$$

An interchange of variables, $T(\Psi_\alpha) = \alpha^2 T(\Psi)$ and $V_{ee}(\Psi_\alpha) = \alpha V_{ee}(\Psi)$, allows for the Hohenberg–Kohn functional to be expressed as[64]:

$$\begin{aligned} F(\rho_\alpha) &= \alpha \min_{\Psi_\alpha \mapsto \rho_\alpha} \langle \Psi_\alpha | \alpha \widehat{T} + \widehat{V}_{ee} | \Psi_\alpha \rangle, \ \Psi_\alpha \in \mathscr{W}_N \\ &= \alpha \min_{\Psi \mapsto \rho} \langle \Psi | \alpha \widehat{T} + \widehat{V}_{ee} | \Psi \rangle, \ \Psi \in \mathscr{W}_N \\ &= \alpha^2 F^{1/\alpha}(\rho) \end{aligned} \tag{1.22}$$

Thus, coordinate scaling and the adiabatic connection are combined to yield Equation 1.22. For the weak-interaction limit, analogous to that derived in the strong-interaction limit,[64] we find that:

$$F(\rho_\alpha) \underset{\alpha \longrightarrow \infty}{\sim} \alpha^2 \min_{\Psi \mapsto \rho} \langle \Psi | \widehat{T} | \Psi \rangle \tag{1.23}$$

$$\lim_{\lambda \longrightarrow 0} [F^\lambda(\rho)] = T_s(\rho) \tag{1.24}$$

Moreover, in terms of the Kohn–Sham formalism, and the adiabatic connection:

$$F^\lambda_{KS}(\rho_\alpha) = T_s(\rho_\alpha) + \lambda (J(\rho_\alpha) + E^\lambda_{xc}(\rho_\alpha)) \tag{1.25}$$

Recall that the exchange-correlation energy may be divided into its respective components: $E^\lambda_{xc} = E_x + E^\lambda_c$. At the high-density limit, density-scaling yields the following relation[59]:

$$\begin{aligned} E_c(\rho_\alpha) &> \alpha \ E_c(\rho), \ (\alpha > 0) \\ E_x(\rho_\alpha) &= \alpha \ E_x(\rho) \end{aligned} \tag{1.26}$$

The correlation energy, E_c, does not scale exactly as the density does. Rather, it approaches a negative constant as $\alpha \rightarrow \infty$.[60] The exchange energy scales linearly, constituting the predominant contribution to the exchange-correlation energy.[65] Noting again the inverse relation between α and λ (i.e., $\lambda = \alpha^{-1}$), a perturbation expansion may be used to present the high-density limit as the limit where $\lambda \rightarrow 0$.[65, 66]

$$\begin{aligned} F_{KS}^{\lambda}(\rho) &= T_s(\rho) + \lambda[J(\rho) + E_{xc}^{\lambda}(\rho)] \\ &= T_s(\rho) + \lambda[J(\rho) + E_x(\rho)] + \lambda^2 E_c^{(2)}(\rho) + \lambda^3 E_c^{(3)}(\rho) + ... \end{aligned} \tag{1.27}$$

The weak dependence on λ in the high-density regime[62, 65] (i.e, when $\alpha >> 1$) follows from Eqn 1.21; inverse proportionality between the scaling factor and the coupling constant is the condition required for density scaling to commute with constrained search. This observation concerning the high-density limit can be related to the core electrons. The K-shell electrons are more strongly influenced by the attractive electrostatic potential of the nuclei than by electrostatic repulsion with other electrons. Being highly localized, they comprise a point-charge-like probability distribution. Whereas the outer-shell electrons are screened from the nuclear charge by inner-shell electrons, causing electrostatic repulsion (and the dynamic correlation associated with it) to dominate, the inner-shell electrons are screened from Coulomb interaction with the outer-shell electrons. The strong repulsion felt between the core electrons themselves is still far smaller than the large electrostatic attraction felt toward the nucleus. The dynamic correlation correction to the Coulomb potential is thus comparatively small (but non-zero) for these electrons, and Fermi interactions (i.e., exchange interactions) are far more important, comprising the leading correction to their energy. This picture of core electrons agrees with Eqn. 1.27, repre-

senting the simplification of F_{KS}^{λ} in the high-density limit (recall that the KS functional excludes the one-body potential).

As illustrated, the primary contributions from the KS functional to KS orbital energies (assuming implicitly a non-degenerate ground-state) in the limit where α is large and λ is small are that of $T_s(\rho)$, $J(\rho)$, and $E_x(\rho)$. The zeroth order term in Eqn. 1.27, $T_s(\rho)$, is simply the limit in Equation 1.24. The latter two terms comprise the first-order correction. Only the exchange energy E_x need be considered in greater detail. Its value is XC-functional dependent; the definitions of $T_s(\rho)$[b] and $J(\rho)$ are exact. Thus, uncertainty in the KS core-orbital energies yielded by a given XC functional must derive primarily from the description of exchange.

1.2.4 Self-Interaction and Electron Exchange

The self-interaction error (SIE) is pervasive at the level of inner-shell electrons. This SIE, derived from an incomplete cancellation of the Coulombic self-interaction[c] by the (semi-local) exchange correction, can scale dramatically for the inner-shell electrons because the magnitude of the self-interaction is larger when a small inter-electronic distance is considered (r_{ij}), and the core electrons comprise a point-charge-like spatial probability distribution. The electron exchange described by density functional approximations (DFAs) is often described as "local", or "semi-local". The local (spin-) density approximation (LDA, or LSDA),[57, 67] is an explicit function of the density, $\rho(r)$, and is based on the *uniform electron gas* model[68] which assumes a homogeneous electron density. Accordingly, LDA exhibits incorrect behaviour at the weak

[b] Recall that this is the *non-interacting* kinetic energy.

[c] *Technical Note.* In Eqn. 1.10, two-body electron integrals are summed over i and j for $i \neq j$, implying that these examples are redundant. Practically, integrals are stored in matrix form, and self-interaction is subject to explicit cancellation.

and strong correlation limits (at its asymptotic limits), where the electron density is either highly concentrated or highly diffuse. In the case of semi-local exchange, additional correction terms are defined. To account for the non-homogeneity of the electron density, the generalized gradient approximation (GGA) uses the density gradient, $|\nabla\rho(r)|$, to compute an additional correction term to the local exchange energy.[69]

$$E_{XC}^{SL}[\rho] = \int e_{XC}[\rho(\mathbf{r}), |\nabla\rho(r)|, \ldots]d\mathbf{r} \tag{1.28}$$

$$E_{X}^{LDA}[\rho] = \int \rho(\mathbf{r})\,\varepsilon_X(\rho)\,d\mathbf{r} \tag{1.29}$$

$$\varepsilon_X(\rho) = -\frac{3}{4}\sqrt[3]{\frac{3\rho(r)}{\pi}} \tag{1.30}$$

A general expression for semi-local exchange-correlation energy is given in Eqn. 1.28, where e_{XC} is the exchange-correlation energy density. The expression for local LDA-type exchange is given in Eqn. 1.29, where ε_X is the particle exchange energy (or "Slater" exchange energy[70]) defined in Eqn. 1.30. This represents the exchange energy per particle of a *uniform electron gas*, having the density $\rho(r)$. The exchange energy E_X^{LDA} is the sum over all space of the particle exchange energy weighted by the probability, $\rho(r)$, that an electron exists at that point, r. In the LDA case, the exchange is only a variable of ρ (as in Eqn. 1.29). The *locality* of LDA is derived from its reference-point-centered approach; the exchange hole is determined for a reference electron at a particular coordinate, r.

$$E_X^{exact}[\rho] = -\frac{1}{2}\sum_{i,j}\int\int \psi_i^*(\mathbf{r_1})\psi_j^*(\mathbf{r_2})\frac{1}{r_{12}}\psi_j(\mathbf{r_1})\psi_i(\mathbf{r_2})dr_1dr_2 \tag{1.31}$$

So-called "exact exchange" (EEX)[d] is a type of non-local exchange which is employed in hybrid XC functionals. EEX is computed in the same manner that Hartree–Fock (HF) exchange is computed; see Equation 1.31. The former is sometimes distinguished from the latter in that *exact* exchange is implied to correspond to a correlated DFT density. The *HF* exchange likewise corresponds to that computed for a HF density, in which the opposite-spin electrons are uncorrelated. Regardless, EEX is not an explicit function of the density, but rather a function of the orbital basis; the EEX hole simply assumes the form of the KS orbital. Exact exchange is "non-local" because it represents a reference-point (r) independent description of electron exchange. While EEX is exact for one-electron systems (more broadly, where $\lambda = 0$), it is not exact for an interacting system of electrons. Although the XC functionals used in DFT are not exact either. LDA functionals are exact for a homogeneous (i.e., *infinite*) electron density. Real systems have a finite number of electrons, and the density they describe is *finite*, truncating at the high- and low-density limits. Notably, the most common scheme of hybrid functional incorporates fractions of both EEX and semilocal exchange in a *global* manner. Range-separated hybrid functionals apply higher fractions of EEX in particular regions, but this range-separation is parameterized and system-dependent.

The nature of non-local (exact) exchange can be illustrated by considering the following three cases: (i) an individual H atom, (ii) a stretched cationic H_2^+ molecule, and (iii) a stretched neutral H_2 molecule. In the first case, the EEX hole of an electron in an *s*-orbital correctly assumes the form of the orbital, exactly canceling out the self-repulsion of the electron in question. In the second case, the single electron has an EEX hole equally centered about either atom. This is logical; its position is not correlated with that of another electron, and it has an equal probability

[d] Note that this term is misleading; the "exact exchange" is only *exact* for one-electron systems.

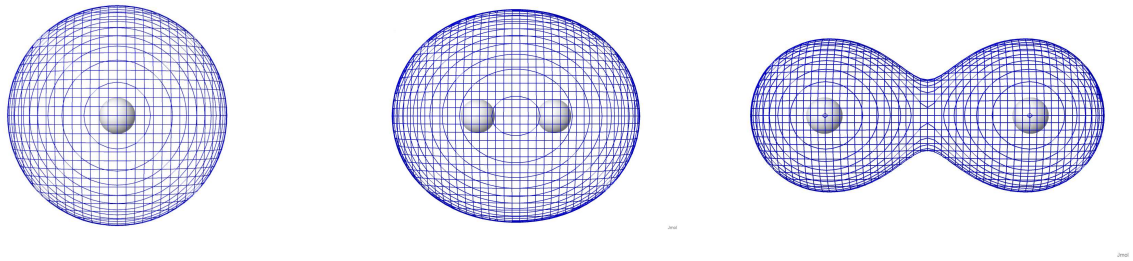


Figure 1.1: Visualization of the 1s-orbital in the H atom (left), and of the bonding σ-orbital in equilibrium H_2 (centre) and stretched H_2 (right).

of being located about either nucleus. Again, the self-repulsion is exactly canceled out by the exchange term. However, in neutral H_2, two electrons share a bonding molecular orbital. The EEX hole of either electron assumes the form of the delocalized bonding orbital, distributed over both nuclei. Here, pair correlation is neglected. In reality, the probability distributions of the electrons are mutually coupled; when one electron is located about one nucleus, there is an increased probability of finding its spin-paired electron about the adjacent nucleus. This "left-right" correlation is a type of long-range *static* correlation, inherently neglected within a single-determinant method based on molecular orbital (MO) theory. This is equivalent to simply stating that Fermi and Coulomb correlation are mutually coupled. These observations about EEX holes can be visualized by directly plotting the molecular orbitals. The HF orbitals are illustrated in Figure 1.1 for the H atom and the H_2 molecule. As the H_2 bond is stretched, the electron pair should be increasingly localized about either atom, which is not reflected in the canonical HF orbitals; in stretched H_2, both electrons equally occupy the delocalized MO depicted on the far right.

In DFT, local and non-local exchange are equally problematic. Whereas EEX over-localizes electrons (under-binding covalent bonds), the self-interaction error inherent to local exchange

leads to over-delocalization (over-binding covalent bonds)[71, 72] and the related fractional charge error.[73] Nevertheless, resolution of this self-interaction does not universally yield improvement. The well-known Perdew–Zunger self-interaction correction (PZ-SIC)[74] was designed to overcome the supposed flaw of self-interaction in local exchange functionals; yet, it was observed that at best it made little difference, and at worst it degraded the accuracy of physical properties.[75] This result suggests that the unphysical self-interaction somehow plays a role in describing certain systems. The concept of pair correlation, as explained previously, is key to understanding this phenomenon. That is, the SIE in exchange functionals is observed to mimic the effects of pair correlation.[76] While DFT describes dynamic correlation well, it cannot, in general, account for static correlation. Thus, the SIE is counter-intuitively advantageous to DFT, wherein the density is assumed to arise from a single electronic configuration. Nevertheless, it is also the source of the many-electron self-interaction (delocalization) error, the solution to which is non-trivial. Furthermore, while SIE indirectly mimics pair correlation in bonding valence orbitals, it simply leads to large errors in the core orbital energies, resulting in differential errors between the core and valence manifolds. Thus, the orbital-energy differences between the core and valence manifolds are greatly underestimated in KS-DFT, especially when the excitation of a 1s electron is considered. The implications of this "orbital-energy SIE" are revisited in the following section, which discusses time-dependent DFT, and once more in Section 1.4, which discusses the DFT/MRCI formalism.

1.2.5 TIME-DEPENDENT DFT AND THE TAMM–DANCOFF APPROXIMATION

The general excited-state extension to the DFT methodology is time-dependent density functional theory (TDDFT), which can be expressed in terms of the real-time[77] (RT) or linear response[78–80] (LR) approaches. LR-TDDFT is far more common for the determination of excitation energies, being more practical than the former; diagonalization of the LR-TDDFT equations scales as $\mathcal{O}(N^4)$[81], and LR is typically faster and more direct than real-time propagation.[e] TDDFT assumes an adiabatic approximation: that the Kohn–Sham XC kernel reacts instantaneously to changes in the electronic density.[85] Thus, temporal information obtained via real-time propagation need not be considered. In the standard approach, the solution of the non-Hermitian eigenvalue equation below yields the excitation and de-excitation energies, for $\omega > 0$ and $\omega < 0$, respectively.

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}* & \mathbf{A}* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \tag{1.32}$$

$$\mathrm{A}_{ia,jb} = (\varepsilon_a - \varepsilon_i)\delta_{ij}\delta_{ab} + (ia|jb) - \mathrm{x}_{HF}(ij|ba) + (1 - \mathrm{x}_{HF})(ia|f_{XC}|jb) \tag{1.33}$$

$$\mathrm{B}_{ia,jb} = (ia|bj) - \mathrm{x}_{HF}(ib|ja) + (1 - \mathrm{x}_{HF})(ia|f_{XC}|bj) \tag{1.34}$$

The matrix elements of the **A** and **B** matrices in Eqn. 1.32 are given by Eqns. 1.33 and 1.34 in terms of compound indices *ia* and *jb*. The coefficient x_{HF} is the fraction of EEX in the chosen

[e] RT-TDDFT can be more efficient for extensive systems with a high density of states, see discussion in ref.[82, 83]. Nevertheless, the RT formalism describes spectral features; it does not explicitly solve for individual excitation energies, which is the object of the present work. See ref. [84] for an extensive review of RT-TD methodology.

XC functional. The XC kernel is defined as:

$$f_{XC} = \frac{\delta^2 E_{XC}(r)}{\delta\rho(r)^2} \tag{1.35}$$

Following standard convention, the two-electron integral V_{ijkl} is simplified to $(ij|kl)$. Applying the Tamm–Dancoff approximation (TDA), [86] excitations and de-excitations are decoupled by neglecting the **B** matrix. This reduces the eigenvalue equation to a form analogous to that of configuration interaction singles (CIS) (the CI method is expanded upon in Sec. 1.3):

$$\mathbf{AX} = \omega\mathbf{X} \tag{1.36}$$

Core excitations in TDDFT can be understood in a sense similar to the case of charge-transfer (CT) excitations, as both involve spatially localized orbitals.[87–89] Consider Eqn. 1.33. For excitations from orbital i to a, we assume, as in the CT case, that the orbitals have zero overlap. The on-diagonal exchange-like term $(ia|ia)$ goes to zero due to this zero-overlap condition. The response of the XC kernel in the final term also goes to zero due to the zero overlap. The remaining term is the scaled, Coulomb-like integral $\mathrm{x}_{HF}(ii|aa)$, describing electrostatic attraction between the excited particle a and the positively charged hole left behind, i.

$$\mathrm{A}_{ia,ia} = (\varepsilon_a - \varepsilon_i) - \mathrm{x}_{HF}(ii|aa),\ \langle\phi_a|\phi_i\rangle = 0 \tag{1.37}$$

This equation demonstrates the so-called electron-hole self-interaction error (*eh*-SIE),[90] or electron-transfer self-interaction error (ET-SIE).[88] Only in the case of 100% exact exchange, $\mathrm{x}_{HF} = 1$, is the electron-hole attraction fully described. This type of SIE may be distinguished

from the SIE of the occupied orbitals, which can contribute to inconsistent orbital energy differences at the level of core electrons. The orbital-energy SIE described in Section 1.2.4 is not expected to have as significant an effect as the *eh*-SIE for core-excited states.[89, 91] Nevertheless, both effects contribute to underestimated excitation energy in TDDFT. Selecting the right balance of exact exchange in the underlying XC functional can compensate for the lack of orbital relaxation and higher-order correlation, leading to excitation energies that are more accurate than their CIS counterpart.

While the TDDFT equations formally include a form of correlation via off-diagonal coupling, the extent of this correlation is limited in TDDFT. Only single excitations are described due to the adiabatic approximation invoked by this formalism; the XC kernel is not suited to describing correlation for higher-order excitations. The result is a diagonally dominant Hamiltonian matrix. Furthermore, errors in the diagonal terms, arising from a combination of *eh*-SIE and orbital-energy SIE, cannot be compensated through the effect of configuration mixing. The adiabatic approximation itself follows naturally from the single-particle picture of KS theory, wherein the KS equations reduce the many-electron system into non-interacting single-particle equations and individual particles are rather subject to the *effective potential* of the fully-interacting system. This fundamental limitation means that double excitations, charge-transfer states, and states with doubly-excited or multireference character are especially challenging to describe within a TDDFT framework.

1.3 CONFIGURATION INTERACTION

1.3.1 CONFIGURATION INTERACTION WORKING EQUATIONS

Hartree–Fock (HF) theory[70, 92–94] is the cornerstone of all single reference WFT-based methodology. Just as ground-state density is determined variationally in DFT, the ground-state wave function is determined variationally in Hartree–Fock theory.[f] The wave function is optimized by applying the self-consistent field (SCF) procedure[95] to the standard one-particle HF equations for a given atomic orbital (AO) basis. By considering Eqn. 1.12, an equivalent form of HF theory can be derived as a simplification of KS-DFT. In the limit where $\lambda = 0$, then $E_{xc}^{0} = E_X$ (see Eqn. 1.16). The Hartree–Fock method does not treat dynamic correlation; the XC energy E_{xc}^{λ} is substituted altogether with E_X.

$$F_{HF}(\Psi) = T_s(\Psi) + J(\Psi) + E_X(\Psi) \tag{1.38}$$

This "Hartree–Fock functional", as given above, is not a concept defined in canonical HF theory. Rather, it is used here to highlight the similarity between the Hartree–Fock and Kohn–Sham formalisms. For a review of the fundamental concepts of HF theory, readers are referred to Chapter 3 in ref. [96].

Many methods have been derived to account for dynamic correlation within the WFT frame-

[f] It is often said that “there is no wave function in DFT”. Recall that KS-DFT is defined in terms of the density that the wave function describes, and not in terms of the wave function itself. The HF wave function is a truncation of the complete many-electron wave function; the KS single-determinant wave function is arguably unphysical, being subject to the constraint that it reproduces the same density that the complete many-electron wave function does.

work, falling under the common umbrella term of "post-Hartree–Fock"; configuration interaction (CI) is conceptually the simplest among them. The HF wave function acts as a starting point, assumed to be a good zeroth-order approximation to the ground state. The ground state configuration itself represents a reference or *anchor* configuration. The CI wave function ansatz yields an N-electron basis of Slater determinants by exciting the N electrons of the reference configuration into K virtual orbitals, where the quantity K is basis-set dependent. The wave function is determined variationally as a linear combination of these determinants.

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{ar} c_a^r|\Phi_a^r\rangle + \sum_{a<b,r<s} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + ... \tag{1.39}$$

By convention, the CI wave function (Eqn. 1.39) is expanded by excitation class, where electrons in occupied orbitals $\{a,b,c,...\}$ are excited into virtual orbitals $\{r,s,t,...\}$. The determinants $|\Phi_a^r\rangle$ represent all singly-excited configurations, and $|\Phi_{ab}^{rs}\rangle$ all doubly-excited configurations, etc. The ground-state wave function $|\Psi_0\rangle$ (see Eqn. 1.39) can be represented as a linear combination of these configurations, where $|\Phi_0\rangle$ represents the single Slater determinant (i.e., Hartree–Fock) ground state. Each determinant refers to a particular occupation pattern (i.e., configuration) in the known MO basis, meaning that a wave function $|\Psi\rangle$ can be represented by the vector of coefficients $[c_0\ c_1\ c_2\ ...\ c_{N_{CI}}]$. In most cases, the HF solution is a good approximation to the ground state wave function, $|\Psi_0\rangle$, such that $c_0 \approx 1$, and $|\Psi_0\rangle \approx |\Phi_0\rangle$. Otherwise, configuration mixing yields variational flexibility. For excited states, this mixing can account for both multireference character and the imperfect nature of the MO basis, being that the orbitals are optimized for the

ground state configuration. The secular equation is given as:

$$\mathbf{Hc} = \omega\,\mathbf{c} \tag{1.40}$$

Above, the CI matrix is formed by expressing the molecular Hamiltonian, **H**, in a basis of Slater determinants. Diagonalizing the CI matrix corresponds to optimization of the coefficients **c** with respect to linear variation, yielding the eigenvalues ω. The matrix **c** contains the column vectors corresponding to the wave function of the ground state $|\Psi_0\rangle$, and excited state wave functions $\{|\Psi_1\rangle,\ |\Psi_2\rangle\ldots\}$. The N^{th} excitation energy is yielded by the eigenvalue difference $\omega_N - \omega_0$. The blocks of excited-state configurations can be converted to a symbolic form. For example, $|\Phi_a^r\rangle$ and $|\Phi_{ab}^{rs}\rangle$, respectively representing single and double excitations, may be simplified to $|S\rangle$ and $|D\rangle$. This yields a simplified block structure of the CI Hamiltonian.

$$\mathbf{H} = \begin{bmatrix} \langle\Phi_0|\widehat{H}|\Phi_0\rangle & 0 & \langle\Phi_0|\widehat{H}|D\rangle & 0 & 0 & \ldots \\ 0 & \langle S|\widehat{H}|S\rangle & \langle S|\widehat{H}|D\rangle & \langle S|\widehat{H}|T\rangle & 0 & \ldots \\ \langle D|\widehat{H}|\Phi_0\rangle & \langle D|\widehat{H}|S\rangle & \langle D|\widehat{H}|D\rangle & \langle D|\widehat{H}|T\rangle & \langle D|\widehat{H}|Q\rangle & \ldots \\ 0 & \langle T|\widehat{H}|S\rangle & \langle T|\widehat{H}|D\rangle & \langle T|\widehat{H}|T\rangle & \langle T|\widehat{H}|Q\rangle & \ldots \\ 0 & 0 & \langle Q|\widehat{H}|D\rangle & \langle Q|\widehat{H}|T\rangle & \langle Q|\widehat{H}|Q\rangle & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{1.41}$$

The block structure of the CI Hamiltonian, illustrated in Eqn. 1.41, has two notable features. First, the HF ground state $|\Phi_0\rangle$ does not couple with the singles block $|S\rangle$. This is the statement of Brillouin's Theorem: by definition, the variational solution of the HF ground state eigenvalue problem requires invariance with respect to linear variation. The ground-state represents

an energetic minimum in variational space, and singly-excited determinants are orthonormal to $|\Phi_0\rangle$. Furthermore, there are at most two-body interactions in the electronic Hamiltonian due to electron-electron electrostatic repulsion. This is described by the operator $\widehat{V}_{ee}$ in Eqn. 1.4. Following from Slater–Condon rules, matrix elements between configurations that differ by more than two electron indices are zero, so coupling between blocks differing by more than two excitation levels are deduced *a priori* to be zero. Elements within the singles block $|S\rangle$ can couple with those in the $|D\rangle$ and $|T\rangle$ blocks, but cannot possibly couple with elements in $|Q\rangle$. From the DFT standpoint, an important observation is that HF includes no Coulomb correlation in the ground state; the first-order correction to the ground state energy is obtained via coupling with $|D\rangle$. Typically, the CI Hamiltonian is truncated by excitation class to reduce computational expense, and the resultant method is termed CIS, CISD, etc., based on this truncation.

The description of excited states is aided by representing matrix elements in terms of the difference of occupation between an arbitrary set of determinants. Here, the CI matrix elements are presented using the convention described by Wetmore and Segal.[97–99] The vector $\mathbf{w}$ specifies the occupation number of each MO in a given configuration. For example, the i^{th} MO contains at most two electrons: $\mathrm{w}_i \in \{0,1,2\}$. The anchor configuration is the occupation vector which corresponds to $|\Phi_0\rangle$, denoted $\overline{\mathbf{w}}$. The ground-state SCF energy and the elements of its associated Fock matrix are:

$$E_{SCF} = \sum_i F_{ii}\overline{\mathrm{w}}_i - \frac{1}{2}\sum_{ij}\left(V_{iijj} - \frac{1}{2}V_{ijji}\right)\overline{\mathrm{w}}_i\,\overline{\mathrm{w}}_j \tag{1.42}$$

$$F_{ij} = h_{ij} + \frac{1}{2}\sum_k \left(V_{ijkk} - \frac{1}{2}V_{ikkj}\right)\overline{\mathrm{w}}_k \tag{1.43}$$

The CI matrix takes the form: $\langle \mathbf{w}'\boldsymbol{\omega}'|\hat{H}|\mathbf{w}\boldsymbol{\omega}\rangle$, where the vector $|\mathrm{w}\boldsymbol{\omega}\rangle$ denotes a given configu-

ration state function (CSF). As before, $\mathbf{w}$ is a spatial occupation vector; ω stores the spin occupations of all singly occupied orbitals. The second quantization notation is convenient for representing excited configurations in a single-particle basis. The creation and annihilation operators $[a^{\dagger}_{i\sigma}\, a_{j\tau}]$ respectively create and annihilate an electron at the indicated molecular orbital indices. By convention, Latin letters correspond to the spatial index, and Greek letters to the spin index, either α (spin-up) or β (spin-down). The excitation operator is defined as:

$$\hat{E}_i^j = \sum_{\sigma=\alpha,\beta} a^{\dagger}_{i\sigma}\, a_{j\sigma} \tag{1.44}$$

This operator only explicitly acts on spatial coordinates, preserving multiplicity. However, multiple spin states are possible for a given spatial configuration. The result of the action of $\hat{E}_i^j$ on the configuration $|\mathbf{w}\omega\rangle$ can thus be expressed as a linear combination of possible spin configurations, weighted by the coefficient η, for one- and two-electron excitations:

$$\begin{aligned} \hat{E}_i^j|\mathbf{w}\omega\rangle &= \sum_{\omega'} \eta_i^j(\mathbf{w}',\omega',\mathbf{w},\omega)|\mathbf{w}'\omega'\rangle \\ \hat{E}_i^j\hat{E}_k^l|\mathbf{w}\omega\rangle &= \sum_{\omega'} \eta_{ik}^{jl}(\mathbf{w}',\omega',\mathbf{w},\omega)|\mathbf{w}'\omega'\rangle \end{aligned} \tag{1.45}$$

The spin-coupling coefficient $\eta_{ik}^{jl}(\mathbf{w}',\omega',\mathbf{w},\omega)$ is shortened to η_{ik}^{jl} for convenience. The on-diagonal matrix element H_{nn} for a given CSF is found as:

$$\begin{aligned} \left\langle \mathbf{w}\omega\middle|\hat{H}\middle|\mathbf{w}\omega\right\rangle =& E_{SCF} + \sum_i F_{ii}\Delta\mathrm{w}_i + \frac{1}{2}\sum_{ij}\left(V_{iijj} - \frac{1}{2}V_{ijji}\right)\Delta\mathrm{w}_i\Delta\mathrm{w}_j \\ &+ \frac{1}{2}\sum_{i\in S_\mathrm{w}}\sum_{j\in S_\mathrm{w}} V_{ijji}\left[\eta_{ij}^{ji}(1-\delta_{ij}) - \frac{1}{2}\right] \end{aligned} \tag{1.46}$$

Above, $\Delta \mathrm{w}_i = \mathrm{w}_i - \overline{\mathrm{w}}_i$ is the occupation difference of the i^{th} MO with respect to the anchor configuration. By considering the Slater–Condon rules once again, the off-diagonal Hamiltonian matrix elements can be simplified into three principal cases. In case (i), matrix elements $\langle w\omega'|\widehat{H}|w\omega\rangle$ differ only by spin coupling. The matrix elements $\langle w'\omega'|\widehat{H}|w\omega\rangle$ yield two distinct cases: (ii) $\sum_i \mathrm{w}_i - \mathrm{w}'_i = 1$, and (iii) $\sum_i \mathrm{w}_i - \mathrm{w}'_i = 2$. These off-diagonal matrix elements $\mathrm{H}_{nn'}$ are simplified below.

(i): Same spatial configuration. CSFs differ by spin-coupling.

$$\left\langle \mathbf{w}\boldsymbol{\omega}' \middle| \hat{H} \middle| \mathbf{w}\boldsymbol{\omega} \right\rangle = \frac{1}{2} \sum_{i \in S_{\mathrm{w}}} \sum_{j \in S_{\mathrm{w}}} V_{ijji} \eta_{ij}^{ji} \left(1 - \delta_{ij}\right) \tag{1.47}$$

(ii): One-electron difference. CSFs differ by a single excitation from orbital i to orbital a.

$$\begin{aligned} \left\langle \mathbf{w}'\boldsymbol{\omega}' \middle| \hat{H} \middle| \mathbf{w}\boldsymbol{\omega} \right\rangle = & \left[F_{ia} + \sum_k \left(V_{iakk} - \frac{1}{2} V_{ikka} \right) \Delta \mathrm{w}_k \right] \eta_a^i \\ & + \sum_{\substack{k \in S_{\mathrm{w}} \\ k \neq i,a}} V_{ikka} \left[\eta_{ak}^{ki} + \left(\frac{1}{2} \mathrm{w}_k - 1 \right) \eta_a^i \right] \\ & + \left[V_{aaai} \mathrm{w}_a + V_{aiii} (\mathrm{w}_i - 2) \right] \eta_a^i \end{aligned} \tag{1.48}$$

(iii): Two-electron difference. CSFs differ by a double excitation from orbitals i and j to orbitals a and b.

$$\left\langle \mathbf{w}'\boldsymbol{\omega}' \middle| \hat{H} \middle| \mathbf{w}\boldsymbol{\omega} \right\rangle = \left[V_{aibj} \eta_{ab}^{ij} + V_{ajbi} \eta_{ab}^{ji} \right] \left[(1 + \delta_{ab})(1 + \delta_{ij}) \right]^{-1} \tag{1.49}$$

CI theory, unlike DFT, can be exact in practice. In full configuration interaction (FCI), the many-electron wave function is expanded in terms of all N_{CI} excited-state configurations, span-

ning the complete Hilbert space. FCI corrects the errors introduced by the single Slater determinant approximation in HF theory, which neglects both dynamic and static correlation. Dynamic correlation arises from uncanceled terms in the two-body integrals, while the single reference nature of HF neglects static correlation, leading to broken degeneracy in strongly correlated systems. However, while an FCI (or equivalent) treatment fully recovers the missing correlation energy in the complete basis set limit, it is rarely a feasible option due to steep computational scaling. Therefore, truncated CI schemes must be used.[g] However, truncating CI excitations from a single reference configuration reintroduces the single reference character formerly corrected by FCI, leading to the need for a more sophisticated approach.

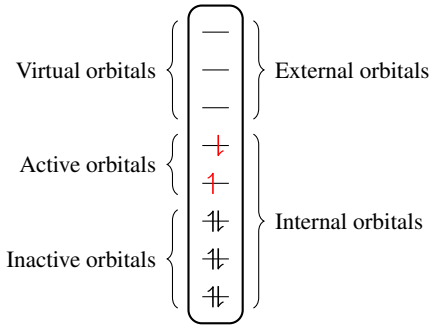
1.3.2 MULTIREFERENCE CONFIGURATION INTERACTION

Multireference configuration interaction (MRCI) is advantageous for the description of excited states when their zeroth-order description is not well described by a single configuration. This is generally the case for excited electronic states. Single reference (SR) methods are inconsistent in that they do not treat dynamic and non-dynamic correlation equally. Thus, they perform poorly for systems with strong static correlation. In contrast, multireference (MR) methods are designed to treat both types of correlation on equal footing while retaining computational savings. This section comprises a brief review of the MRCI method.[100–102] For an in-depth introduction to MRCI, readers are referred to ref. [103]. A broad overview of MR methodology in electronic structure theory is captured within references [104] and [105].

In MR methods, the orbital space is divided into subspaces. These are illustrated and labeled

[g] Technically, truncated CI is superseded by the coupled-cluster (CC) methods, which include higher-order correlation at similar computational scaling.

Figure 1.2: Division of molecular orbital spaces in MRCI.



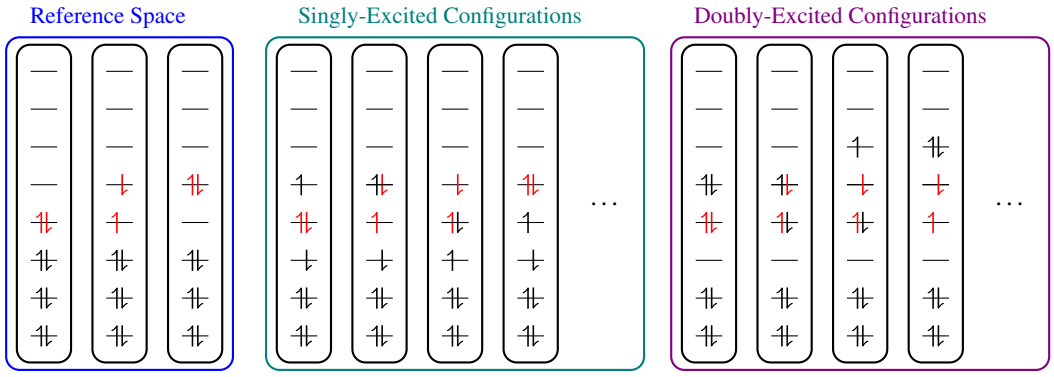
in Fig. 1.2. The active orbital space defines a subset of frontier orbitals in which a complete set of occupation patterns is considered. The anchor configuration represents the sole reference determinant in truncated CI. In contrast, the active space defines a *set* of reference configurations *including* the anchor configuration. Compare the following MRCISD wave function to that in Eqn. 1.39, wherein a set of reference determinants is defined:

$$|\Psi\rangle = \sum_{k=1}^{N_{ref}} c_k^{ref} |\Phi_k^{ref}\rangle + \sum_l c_l^S |\Phi_l^S\rangle + \sum_m c_m^D |\Phi_m^D\rangle \tag{1.50}$$

Inactive orbitals are doubly-occupied in all reference configurations, and all together, the active and inactive spaces make up the set of "internal" orbitals. Likewise, the virtual orbitals are the set of "external" orbitals that are unoccupied in all reference configurations.

In Fig. 1.3, a set of reference configurations is illustrated alongside subsets of the singly- and doubly-excited configurations. In the MRCI procedure, the singly-excited subspace will include configurations that are doubly-excited and greater with respect to the anchor configuration. Likewise, the doubly-excited subspace includes configurations that are triply-excited and greater with respect to the anchor configuration. Thus, MRCI effectively includes higher-order

Figure 1.3: MRCI configuration space depicting the reference configurations (left), singly-excited configurations (centre), and doubly-excited configurations (right).



correlation effects than standard, truncated CI.

$$|\Psi_I\rangle = \sum_{\Omega \in \mathscr{R}} C_{\Omega I} |\Omega\rangle + \sum_{\Omega \in \mathscr{F}} C_{\Omega I} |\Omega\rangle. \tag{1.51}$$

There are different possible selection schemes for generating the Hilbert subspace of configurations in MRCI. In this work, the first-order interacting space (FOIS) is generated via single and double excitations from the active and inactive orbitals. Equation 1.51 explicitly reflects this scheme. The CSF space is partitioned into a reference space $\mathscr{R}$ and a FOIS $\mathscr{F}$, where $\{|\Psi_I\rangle\}$ denotes the set of eigenfunctions of the electronic Hamiltonian $\hat{H}$, and $|\Omega\rangle$ denotes the spin-adapted CSFs.

1.3.3 Linking the TDA and CIS Methods

In the following section, the DFT/MRCI method is formally introduced. However, a more intuitive way to explain DFT/MRCI may be to first derive density functional theory and configuration

interaction singles (DFT/CIS),[106] explicitly drawing parallels between TDA-TDDFT and CIS. The matrix elements of the ab initio formalism presented in the previous section are expressed in a multiplicity agnostic form. By ignoring spin-flip configurations and multiply-excited configurations, Eqn. 1.46 and 1.48 can be simplified for a closed-shell reference. A single-excitation from orbital i to a yields $\Delta w_i = -1$, and $\Delta w_a = 1$. All other occupation differences are zero, and these zero terms are discarded. The matrix elements F_{ii} translate to the HF orbital energies ε_i^{HF}. By summing over all MO indices and grouping equivalent integrals, the on- and off-diagonal matrix elements simplify as:

$$\begin{aligned} \langle \Psi_i^a | \hat{H}_{CIS} - E_{SCF} | \Psi_i^a \rangle &= (\varepsilon_a^{HF} - \varepsilon_i^{HF}) + V_{iaia} - V_{iiaa} \\ \langle \Psi_i^a | \hat{H}_{CIS} | \Psi_j^b \rangle &= V_{iajb} - V_{ijba} \end{aligned} \tag{1.52}$$

The CIS Hamiltonian can be expressed in a general form:

$$\langle \Psi_i^a | \hat{H}_{CIS} - E_{SCF} | \Psi_j^b \rangle = (\varepsilon_a^{HF} - \varepsilon_i^{HF}) \delta_{ij} \delta_{ab} + V_{iajb} - V_{ijba} \tag{1.53}$$

Following the convention used previously for TDA-TDDFT, we retrieve a CIS Hamiltonian matrix with elements:

$$\mathrm{H}_{ij}^{ab} = (\varepsilon_a^{HF} - \varepsilon_i^{HF}) \delta_{ij} \delta_{ab} + (ia|jb) - (ij|ba) \tag{1.54}$$

This equation is analogous to the $\mathbf{A}$ matrix elements defined in Eqn. 1.33, with $x_{HF} = 1$, and KS orbital energies replaced by their HF counterparts. Next, one may consider using the ab initio formulation of CI, but using KS-DFT rather than HF to describe the anchor configuration. The on-diagonal orbital energies are the KS eigenvalues, and the two-electron integrals are subject to

scaling parameters c_X and c_C.

$$\langle \Psi_i^a | \hat{H}^{DFT} - E_{DFT} | \Psi_j^b \rangle = (\varepsilon_a^{KS} - \varepsilon_i^{KS} + \Delta_{ia})\delta_{ij}\delta_{ab} + \mathrm{c}_X(ia|jb) - \mathrm{c}_C(ij|ba) \tag{1.55}$$

Since the exchange-correlation kernel f_{XC} is discarded, an on-diagonal correction Δ is included. This correction is empirically defined in terms of the exchange-like integral.[106] The scaling parameter c_C is introduced in place of x_{HF} to maintain consistency. The parallel between the TDA and CIS equations implies that $\mathrm{c}_C \approx \mathrm{x}_{HF}$, and that the scaling parameter c_X be set to 1 at the DFT/CIS level. Both parameters are empirically fit and depend on the XC functional and basis set. Substituting the coefficient c_C with $1 - p_C$, where $p_C = 1 - \mathrm{c}_C$, and doing likewise for c_X, the ab initio CIS matrix elements can be combined into a term $\langle \Psi_i^a | \hat{H}^{CIS} - E_{SCF} | \Psi_j^b \rangle$. Then, the DFT/CIS matrix elements may be expressed in terms of their difference from the standard CIS Hamiltonian. Below, the on-diagonal HF orbital energy difference is replaced with its KS counterpart, and the two-electron integrals are subject to scaled *correction* terms.

$$\begin{aligned}\langle \Psi_i^a | \hat{H}^{DFT} - E_{DFT} | \Psi_j^b \rangle =& \langle \Psi_i^a | \hat{H}^{CIS} - E_{SCF} | \Psi_j^b \rangle - (\varepsilon_a^{HF} - \varepsilon_i^{HF})\delta_{ij}\delta_{ab} + (\varepsilon_a^{KS} - \varepsilon_i^{KS})\delta_{ij}\delta_{ab} \\ & - p_X(ia|jb) + p_C(ij|ba) + (\Delta_{ia})\delta_{ij}\delta_{ab}\end{aligned} \tag{1.56}$$

The parallel between TDA and MRCI is less clear; correlation energy is described beyond the single-excitation framework of CIS and TDA-TDDFT. Thus, a more complex scheme is required to include electron correlation consistently.

1.4 DENSITY FUNCTIONAL THEORY/MULTIREFERENCE CONFIGURATION INTERACTION

1.4.1 MOTIVATION

The computational bottleneck of MRCI calculations is the diagonalization of the Hamiltonian matrix; the existence of a reference space implies a larger first-order interacting space of excited configurations. However, a vast number of these do not contribute to any particular state. Nevertheless, in purely ab initio CI, accuracy increases as the size of the Hamiltonian does; off-diagonal coupling between configurations, H_{ij}, recovers the dynamic correlation missing in the HF ground state. To loosen the strict dependence that off-diagonal coupling recover the dynamic correlation correction, KS orbital energies are introduced to the on-diagonal elements in Eqn. 1.46. These KS orbital energies, ε_{ii}^{KS}, replace the Fock matrix elements, F_{ii}. Then, with an improved zeroth-order description of the excitation energy, empirical parameters are introduced to avoid double-counting the correlation energy. These terms scale the on-diagonal Coulomb and exchange integrals as well as damp the off-diagonal coupling between configurations, H_{ij}. This latter measure naturally lends itself to truncation of the Hamiltonian matrix, improving the computational expediency of the resultant method, DFT/MRCI. The details of this approach are explained further in the following section, which introduces the working equations of DFT/MRCI in terms of the CI matrix elements defined in Section 1.3.

1.4.2 DFT/MRCI WORKING EQUATIONS

The DFT/MRCI method is presented here following the convention used in ref. [107, 108], with Hamiltonian matrix elements expressed with respect to their purely ab initio counterparts defined in Section 1.3.1.

$$\begin{aligned}\left\langle \mathbf{w}\omega\middle|\hat{H}^{DFT} - E_{DFT}\middle|\mathbf{w}\omega\right\rangle = \left\langle \mathbf{w}\omega\middle|\hat{H} - E_{SCF}\middle|\mathbf{w}\omega\right\rangle \\ + \sum_i (\varepsilon_{ii}^{KS} - F_{ii})\Delta \mathrm{w}_i + \Delta E_x + \Delta E_c.\end{aligned} \tag{1.57}$$

The correction terms to the on-diagonal elements are the Coulomb- and exchange-like two-electron integrals scaled by the parameters p_C and p_X.

$$\Delta E_c = -p_C \left[\sum_{i<j} V_{iijj}\Delta \mathrm{w}_i \Delta \mathrm{w}_j + \frac{1}{2}\sum_{i\in S_{\overline{\mathrm{w}}}} V_{iiii}|\Delta \mathrm{w}_i| + \sum_i V_{iiii}\delta_{\Delta \mathrm{w}_i,2} + \frac{1}{4}\sum_{i\in S_{\overline{\mathrm{w}}}} V_{iiii} \right] \tag{1.58}$$

$$\Delta E_x = p_X \left[-\sum_{\substack{i,j \\ j\in S_{\mathrm{w}}}} V_{ijij}|\Delta \mathrm{w}_i| + \frac{1}{2}\sum_{\substack{i\in C_{\mathbf{w}} \\ j\in A_{\mathbf{w}}}} V_{ijij}\Delta \mathrm{w}_i \Delta \mathrm{w}_j + \sum_{\substack{i<j \\ i,j\in S_{\mathrm{w}}}} V_{ijij}\eta_{ij}^{ji} \right] \tag{1.59}$$

The off-diagonal matrix elements can again be divided into three cases, with an additional consideration. The ab initio MRCI recovers dynamic correlation via coupling between the reference and FOIS CSFs. In DFT/MRCI, there is already an on-diagonal correction to the matrix elements derived from the replacement of HF orbital energies with their KS-DFT counterpart. To avoid double-counting the dynamic correlation described in this manner, the off-diagonal Hamiltonian matrix elements are damped according to an energy-based criterion. The parameter $\Delta E_{\mathbf{ww}'}$ is introduced to describe the spin-coupling-averaged energy difference between a pair of configu-

rations, $\mathbf{w}$ and $\mathbf{w}'$.

$$\Delta E_{\mathbf{ww}'} = \frac{1}{n_\omega}\sum_{\omega}^{n_\omega} H^{DFT}_{\mathbf{w}\omega,\mathbf{w}\omega} - \frac{1}{n_{\omega'}}\sum_{\omega'}^{n_{\omega'}} H^{DFT}_{\mathbf{w}'\omega',\mathbf{w}'\omega'} \tag{1.60}$$

The three cases are as follows below.

(i): Same spatial configuration. CSFs differ by spin-coupling.

$$\left\langle \mathbf{w}\omega\middle|\hat{H}^{DFT} - E_{DFT}\middle|\mathbf{w}\omega'\right\rangle = (1-p_X)\left\langle \mathbf{w}\omega\middle|\hat{H} - E_{SCF}\middle|\mathbf{w}\omega'\right\rangle, \tag{1.61}$$

(ii): One-electron difference. CSFs differ by a single excitation.

$$\left\langle \Omega\middle|\hat{H}^{DFT} - E_{DFT}\middle|\Omega'\right\rangle = D(\Delta E_{\mathbf{ww}'})\left(\left\langle \Omega\middle|\hat{H} - E_{SCF}\middle|\Omega'\right\rangle - F_{ia}\right), \tag{1.62}$$

(iii): Two-electron difference. CSFs differ by a double excitation.

$$\left\langle \Omega\middle|\hat{H}^{DFT} - E_{DFT}\middle|\Omega'\right\rangle = D(\Delta E_{\mathbf{ww}'})\left\langle \Omega\middle|\hat{H} - E_{SCF}\middle|\Omega'\right\rangle, \tag{1.63}$$

The form of the damping function $D(\Delta E_{\mathbf{ww}'})$ differs in different versions of DFT/MRCI.[29, 43, 44] The form of the function used in this work is specified in Chapter 2. This damping results in the decoupling of a subset of the FOIS from the reference space, leading to many redundant configurations. The size of the FOIS is therefore reduced using an orbital-energy-based selection criterion:[29]

$$d_{\mathbf{w}} = \sum_i \Delta \mathrm{w}_i \varepsilon_i^{KS} - \delta E_{sel} \tag{1.64}$$

Typically, δE_{sel} is set to either 0.8 or 1.0 E_h. This selection criterion, along with the specific

details that pertain to the damping function, is further discussed in Chapter 2.

The concept of *eh*-SIE does not clearly transfer from TDDFT to DFT/MRCI. The exchange-correlation kernel (f_{XC}) is discarded, and the inclusion of higher-order correlation via the MRCI *ansatz* increases variational flexibility in the resultant CI matrix due to configuration mixing. Furthermore, the semi-empirical parameterization is fit to maximize agreement with benchmark-quality excitation energies. Nevertheless, the issue of orbital-energy SIE applies to the DFT/MRCI equations, meaning that a single parameterization is insufficient to describe both valence and core excitation with comparable accuracy.[48] As a result, an improved parameterization of DFT/MRCI for both valence and core excitation must consider the underlying XC functional in addition to considering the fit of the empirical parameters themselves. The following chapters respectively consider each case: valence-excited states and core-excited states.

2

A DFT/MRCI Hamiltonian Parameterized Using Only Ab Initio Data: I. Valence Excited States

2.1 INTRODUCTION

The combined density functional theory and multireference configuration interaction (DFT/MRCI) electronic structure method,[29, 30, 41, 43, 44] in which compact MRCI expansions are supplemented with DFT-specific corrections, has a number of desirable properties. First, it is amenable to "black box" implementations, as the reference space is generated automatically and iteratively, rather than via a rigid active orbital subspace specified *a priori*. Secondly, it can furnish accurate vertical excitation energies (VEEs) at low computational cost through the use of DFT-specific Hamiltonian corrections to account for the preponderance of dynamic electron correlation. Thirdly, it can describe a large range of electronic states, including those of multireference, Rydberg, and intramolecular charge-transfer character.

In part driven by this set of desirable and comparatively unique characteristics, the DFT/MRCI method has seen a recent increase in the pace of development over the past years. In particular, the original multiplicity-dependent formulation has been expanded to a spin-state agnostic parameterization,[43] extended to transition metal complexes,[41] and a new formulation was recently presented that focused on redressing previously observed errors in the description of doubly-excited electronic states.[47] Additionally, new methods based on the original DFT/MRCI approach, but with even more favorable computational scaling, have recently been described. This includes a “pruned” variant, *p*-DFT/MRCI,[45] and a perturbative approximation, DFT/MRCI(2),[46] in which the interaction of the reference and first-order interacting space (FOIS) is treated semi- and fully-perturbatively, respectively. Lastly, a core-valence separated DFT/MRCI method (CVS-DFT/MRCI) has been introduced[48] for the description of core-

excited states and has been subsequently used to simulate both static and time-resolved X-ray spectra.[109]

The origin of the computational efficiency of DFT/MRCI lies in the use of short, truncated MRCI expansions to capture the static electronic correlation, in conjunction with DFT-specific Hamiltonian corrections aimed at the recovery of the remaining dynamic correlation. The DFT/MRCI Hamiltonian directly incorporates Kohn–Sham (KS) orbital energy differences, long-recognized to yield a good zeroth-order description of electronic transition energies, into the diagonal matrix elements, and thus effectively incorporates a significant amount of dynamic correlation in a highly efficient manner. Accordingly, the off-diagonal elements of the Hamiltonian are damped to avoid a double-counting of dynamic correlation. Furthermore, the reference space is generated iteratively via a selected configuration interaction algorithm, thereby ensuring that the description of static correlation, captured via the interaction of the reference configurations, is also highly computationally efficient, with only those configurations contributing significantly to the wave functions of interest being retained.

The semi-empirical parameters that arise in the working equations for DFT/MRCI Hamiltonians serve two primary purposes: *i*) to scale the Coulomb and exchange integral contributions to the diagonal Hamiltonian matrix elements, and *ii*) to damp the off-diagonal Hamiltonian matrix elements using a configuration energy-based criterion, as discussed above. The values of these parameters have historically been chosen so that DFT/MRCI vertical excitation energies reproduce the peak maxima in ultraviolet/visible (UV/Vis) absorption spectra, although more recent parameterizations have employed mixed ab initio and experimental fitting data.[47] However, previous work has demonstrated that, in many cases, the positions of peak maxima in UV/Vis absorption spectra can exhibit significant deviations from electronic vertical excitation

energies.[110–114] In particular, structural relaxation, excited state anharmonicity, and vibronic coupling can all give rise to absorption band profiles for which the maximum of the band is shifted from the corresponding electronic vertical excitation energy. In short, the underlying transitions are *vibronic* in nature, not electronic. For organic chromophores, these effects generally conspire to give rise to positions of peak maxima that are red-shifted by around 0.1–0.3 eV compared to the corresponding electronic VEEs. As such, one can anticipate a significant underestimation of electronic vertical excitation energies by DFT/MRCI Hamiltonians whose fitting data sets contain such experimentally-derived excitation energy estimates.

In contrast to the DFT/MRCI Hamiltonian *ansatz* and parameterization, the choice of the underlying exchange-correlation (XC) functional has not been the subject of significant study in the literature. In fact, all current DFT/MRCI implementations employ the same functional: BHLYP.[115] The large quantity of exact exchange (i.e. 50%) has been previously noted empirically to yield more accurate excitation energies relative to, for example, B3LYP.[116] However, no systematic exploration of the choice of XC functionals has yet been performed.

This work reports a new DFT/MRCI formulation that differs from previous iterations in two fundamental characteristics: *i*) the values of the Hamiltonian parameters are optimized to reproduce benchmark ab initio estimates of electronic excitation energies only, and *ii*) the choice of XC functional is made to give a balanced description of both core and valence excitation and ionization energies. The motivation for the first property is to develop a method that reproduces as well as possible the eigenvalues of the *electronic* Hamiltonian; by incorporating experimental UV/Vis absorption peak maxima into fitting sets, previous parameterizations can be viewed as having a bias to instead reproducing *vibronic* eigenvalues. As for the second property, the previously implemented CVS-DFT/MRCI method,[48] based on the R2017 Hamiltonian, has

been observed to yield reasonably accurate core excitation energies, albeit after applying large (multi-eV), edge-dependent shifts. Given that the KS orbital energy differences make up the leading contribution to the diagonal matrix elements of the DFT/MRCI Hamiltonian, the present study considers whether the use of a different KS orbital basis will be able to provide improved accuracy for vertical core excitation energies, while retaining an accurate description of valence-excited states. This will lead the way for the development of a DFT/MRCI Hamiltonian tailored to the accurate description of core-excited states, which is the topic of Chapter 3.

Continuing from Section 1.4, in which the DFT/MRCI matrix elements are described, Section 2.2 begins with a brief overview of the semi-empirical parameters that characterize the DFT/MRCI method. This is followed by an articulation of the parameterization strategy, including the choice of ab initio data included in the fitting and validation sets. Section 2.3 describes technical details of the electronic structure computations and Section 2.4 presents the results of the parameter optimization, including the choice of a new XC functional employed to generate the underlying KS orbital basis. This new parameterization is denoted "QE8", and it shows incredibly favorable error metrics for the excitation energies of organic molecules. Section 2.5 showcases a set of representative applications of the QE8 parameterization including UV/Vis absorption spectra and vertical excitation energies for large molecular systems, which are reported alongside the results of a comparative benchmark. Finally, Section 2.6 summarizes the work addressed in this chapter, including a prospectus for future refinements.

2.2 THE DFT/MRCI METHOD

2.2.1 DFT/MRCI PARAMETERIZATION

For MRCI wave functions (see Eqn. 1.51), the relatively small reference space $\mathscr{R}$ is primarily associated with a description of static correlation. In contrast, the extensive FOIS $\mathscr{F}$ primarily captures dynamic correlation energy. Due to the size of the FOIS, the diagonalization of the electronic Hamiltonian projected onto the space spanned by the total configuration state function (CSF) basis, $\mathscr{R} \cup \mathscr{F}$, quickly becomes an extremely demanding computational task.

In Eqn. 1.46, the on-diagonal Hamiltonian matrix elements may be interpreted in a particle-hole picture with respect to excitations relative to the base configuration $\overline{\mathbf{w}}$. The leading contribution to each element is the sum of differences of on-diagonal Fock matrix elements between the excited (particle) and annihilated (hole) orbitals. Supplementing these zeroth-order terms are exchange and Coulomb interactions between the particle and hole orbitals as well as spin-couplings between the open shell orbitals. The key idea in DFT/MRCI is to replace the zeroth-order Fock-matrix element difference term with a quantity that more closely correlates with ground-to-excited-state excitation energies. To achieve this, a KS orbital basis is adopted and these terms are replaced with the corresponding differences between KS orbital *energies*, which more closely approximate vertical excitation energies. Then, as the zeroth-order description is now improved, the Coulomb and exchange contributions must be down-scaled. In this way, one may effectively incorporate a large amount of dynamic electron correlation into the on-diagonal Hamiltonian matrix elements that would otherwise be accounted for by the coupling of the reference and FOIS CSFs. The correction terms ΔE_c and ΔE_x respectively down-scale the

Coulomb and exchange contributions. The exact form of these corrections has varied across the different DFT/MRCI parameterizations.[29, 41, 43, 44] In this work, the selected form of the correction terms ΔE_c and ΔE_x is that defined by Marian and co-workers[44], featuring the scalar parameters $p_X, p_C \in [0,1]$.[41, 44] These corrections are defined in Eqns. 1.58 and 1.59, and they are shown again here.

$$\begin{aligned}
\Delta E_c &= -p_C \left[\sum_{i<j} V_{iijj} \Delta \mathrm{w}_i \Delta \mathrm{w}_j + \frac{1}{2} \sum_{i \in S_{\overline{\mathrm{w}}}} V_{iiii} |\Delta \mathrm{w}_i| + \sum_i V_{iiii} \delta_{\Delta \mathrm{w}_i, 2} \right] + \frac{1}{4} \sum_{i \in S_{\overline{\mathrm{w}}}} V_{iiii} \\
\Delta E_x &= \; p_X \left[- \sum_{\substack{i,j \\ j \in S_{\mathrm{w}}}} V_{ijij} |\Delta \mathrm{w}_i| + \frac{1}{2} \sum_{\substack{i \in C_{\mathbf{w}} \\ j \in A_{\mathbf{w}}}} V_{ijij} \Delta \mathrm{w}_i \Delta \mathrm{w}_j + \sum_{\substack{i<j \\ i,j \in S_{\mathrm{w}}}} V_{ijij} \eta_{ij}^{ji} \right]
\end{aligned} \tag{2.1}$$

For the off-diagonal Hamiltonian matrix elements in Eqs. 1.62 and 1.63, $D(\Delta E_{\mathbf{ww}'})$ denotes a damping function that is chosen to decay rapidly with increasing $\Delta E_{\mathbf{ww}'}$: in practice either an exponential[29, 41] or inverse arctangent[44, 116] function. In this work, a flexible three-parameter exponential damping function is adopted, defined as follows:

$$D(\Delta E_{\mathbf{ww}'}) = d_1 \exp\left(-d_2 \Delta E_{\mathbf{ww}'}^{d_3} \right) \tag{2.2}$$

The damping function of this form thus introduces three parameters: d_1, d_2, and d_3. For formal simplicity, and to ensure well-behaved non-linear optimizations, the parameter d_3 is constrained to positive integer values, i.e., $d_3 \in \mathbb{N}$. The damping of off-diagonal matrix elements decouples a large part of the FOIS from the reference space. This now-redundant subset of FOIS configurations may be removed from the calculation entirely by using a simple orbital energy-based selection criterion: $d_{\mathbf{w}}$ (see Eqn. 1.64). For each FOIS configuration $\mathbf{w}$, the quantity $d_{\mathbf{w}}$ is com-

puted, 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 **w** are selected for inclusion, else they are discarded. As mentioned in Sec. 1.4.2, the value of δE_{sel} has previously been chosen to be either 1.0 or 0.8 E_h, with the latter corresponding to a more aggressive truncation of the FOIS, leading to increased computational savings at the cost of modestly reduced accuracy. This configuration selection step results in a massive reduction of the size of the CSF basis - typically by many orders of magnitude - and yields huge speedups relative to an ab initio MRCI calculation.

2.2.2 Reference Space Generation

As an individually selected CI method, DFT/MRCI is a black box approach to multi-reference calculations; the reference space is refined iteratively via successive calculations. The initial guess reference space, $\mathscr{R}_0$ is defined using an automatic selection procedure[45] termed AutoRAS. Analyzing the eigenvectors of the initial calculation, configurations with large CI coefficients that generate one or more CSFs contributing to the states of interest are selected to construct a refined reference space, $\mathscr{R}_1$. This process is repeated until convergence. A description of this procedure is also found in ref. [46].

Generally, a good guess for the initial reference space is made using the restricted active space CI (RASCI)[117] set of configurations. However, an automated generation of $\mathscr{R}_0$ which consistently yields a good zeroth-order description of the states of interest, $|\Psi_I^{(0)}\rangle$, is important to achieve reliable results with the approximative variants *p*-DFT/MRCI[45] and DFT/MRCI(2).[46] In restricted active-space (RAS) methods, the active space selection is divided into the RAS1 space, the subset of orbitals defined in terms of a maximum number of holes, and the RAS3

space, a subset of orbitals defined in terms of a maximum number of electrons. AutoRAS obviates the need to manually select these RAS orbitals by selecting particle/hole molecular orbitals (MOs) based on the dominant configurations of a preliminary density functional theory and configuration interaction singles (DFT/CIS) calculation.[106] In cases where the preliminary DFT/CIS calculation becomes large, an integral pre-screening is introduced to identify dominant configurations at a reduced cost, as described in ref. [45].

2.3 COMPUTATIONAL DETAILS

There are two distinct aspects to the optimization of the parameters that arise in the DFT/MRCI Hamiltonian: *i)* the calculation of vertical excitation energies for multiple molecules and electronic states using a given set of parameters, and *ii)* the iterative variation of those parameters so as to minimize the error between the computed excitation energies and those of a pre-determined fitting data set, calculated using benchmark-level ab initio quantum chemistry methods. In a subsequent step, optimized parameter values are validated via the computation of the errors that result for a separate sub-set (the validation set) of benchmark excitation energies not included in the fitting (training) set. Below is an overview of the molecules and electronic states included in the training and validation sets, as well as the technical elements of the computations. This is followed by a brief discussion of the considerations involved in the optimization of the semi-empirical parameters that arise in the DFT/MRCI Hamiltonian.

The nuclear geometries and benchmark vertical excitation energies comprising the training and validation sets were obtained from the QUEST databases.[20–26, 28] In all cases, geometries optimized at the aug-cc-pVTZ/CC3[118] level of theory were used. At the time of writing, the QUEST database comprises seven data sets, grouped by the type of electronic excitation and/or molecule size. The fitting set employed here included vertical excitation energies from six of the data sets (DSs): 51 from small molecules (DS1), 9 from double excitations (DS2), 50 from medium molecules (DS3), 27 from exotic molecules and radicals (DS4), 24 from larger molecules (DS5), and 4 from charge transfer excitations (DS6), for a total of 93, 19, and 60 singlet, doublet, and triplet vertical excitation energies, respectively. A complete list of molecules

and excitation energies used in the training and validation sets is given in the appendices.

Where available, near full configuration interaction (FCI) benchmark results were used, computed using the configuration interaction using a perturbative selection made iteratively (CIPSI) method.[119–122] For the remaining molecules, the QUEST theoretical best estimate (TBE), computed using higher-order coupled-cluster theory methods, was employed as the benchmark value. The computation of the DFT/MRCI excitation energies was always performed using the same basis set as that employed in the benchmark computation: typically aug-cc-pVTZ[123, 124] when comparing to the TBE results, and aug-cc-pVDZ for the near-FCI values. All DFT/MRCI computations were performed using the GRaCI software package,[125] which employs PySCF modules to perform KS-DFT computations and integral transformations.[126–130] Supplementary calculations using the equation of motion coupled cluster singles and doubles (EOM-CCSD)[131, 132] method were performed using QChem5.4[133]. Benchmark core excitation energies were computed at the aug-cc-pCVTZ/CCSDT level of theory[134, 135] using the MRCC program package.[136–138]

In all DFT/MRCI calculations, the difference dedicated CI (DDCI)[139] configuration reduction procedure was used. Here, the size of the FOIS is reduced via the exclusion of configurations that do not contribute through second-order in perturbation theory to transition energies. Consistent with previous implementations, the number of configurations in the CSF basis was further reduced by constraining the number of open shells to be at most 10, as well as limiting the maximum excitation degree relative to the base configuration to 8. As described in Sec. 2.2.2, the initial guess reference spaces were generated using the AutoRAS procedure. Finally, all *p*-DFT/MRCI calculations were performed using a pruning threshold α_p of 0.90 (see Reference [45] for a definition of this quantity).

The optimization of the semi-empirical parameters in the DFT/MRCI Hamiltonian was performed using a dedicated module in GRaCI[125] that employs a Nelder–Mead non-linear optimization algorithm[140, 141] as implemented in SciPy.[142] At each iteration of the optimization procedure, those electronic transitions corresponding to the training data were identified via the computation of wave functions overlaps between a large number of roots of the DFT/MRCI Hamiltonian and previously determined reference wave functions of the correct electronic character.

2.4 RESULTS

2.4.1 SELECTION OF XC FUNCTIONAL

The primary motivation of the present work is the determination of a DFT/MRCI parameterization that reproduces benchmark ab initio vertical excitation energies. However, a secondary consideration is to develop a class of methods that can provide a consistent description of both valence and core excitation processes. While a previous implementation of the CVS approximation applied to the valence-parameterized R2017 DFT/MRCI Hamiltonian[48] generally produced good *relative* core excitation energies, the absolute errors were found to be quite large: around 5 eV for K-edge transitions. The main source of error is found to be attributable to the previous choice of XC functional, BHLYP[115]. While the BHLYP KS orbital-energy differences yield good zeroth-order approximations of valence excitation energies, there is a significant degradation when considering core excitations. 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.[143]

This motivates investigating the use of other XC functionals to generate the KS orbitals and orbital energies that enter into the DFT-specific corrections of the DFT/MRCI Hamiltonian. In lieu of performing numerous computationally expensive parameter optimizations for each XC functional, the expected performance of each XC functional was evaluated via a comparison of the errors for valence and core excitation energies using a related method, TDA-TDDFT. Refer to Sections 1.2.5 and 1.3.3; the use of this surrogate method was motivated by the fact that the leading contribution to the diagonal matrix elements of the corresponding matrix eigenvalue

problem within TDA-TDDFT is that of the KS orbital energy differences, analogous to the form of the diagonal matrix elements of the DFT/MRCI Hamiltonian (Eq. 1.57). The evaluation done here naturally ignores the contribution of the *eh*-SIE to the error metrics of each XC functional. Regardless, while the similarities here are more evocative than quantitative, the following results are sufficiently useful to identify candidate XC functionals.

TDA-TDDFT core and valence excitation energy calculations were performed for a series of small molecules, the full results of which are presented in App. A.2. 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 tables A.4 and A.5 of Appendix A 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. To evaluate various XC functionals' simultaneous ability to describe core and valence excitation energies, the mean absolute error (MAE) of core and valence excitation energies are considered, shown in Figure 2.1a), and are also tabulated in Table A.6 of Appendix A. By using a combination of point group symmetry, natural transition orbitals, and optical properties (e.g. oscillator strengths), it was confirmed, case-by-case, that the computed excitation energies for each functional correspond to the appropriate reference values. As Figure 2.1a) shows, the errors in the core excitation energies can be as large as 20 eV (for the functionals shown), while all of the XC functionals reproduce the lowest valence excitations to better than 1 eV. Note that the TDA time-dependent Hartree–Fock (i.e., CIS) results are included (labeled as "HF" in Figure 2.1a)) for reference, and yield MAE errors in the valence and core excitation energies of 0.7 eV and 12.6 eV, respectively. XC functionals that will be potentially useful in the development of new DFT/MRCI parameterizations, namely those that provide a

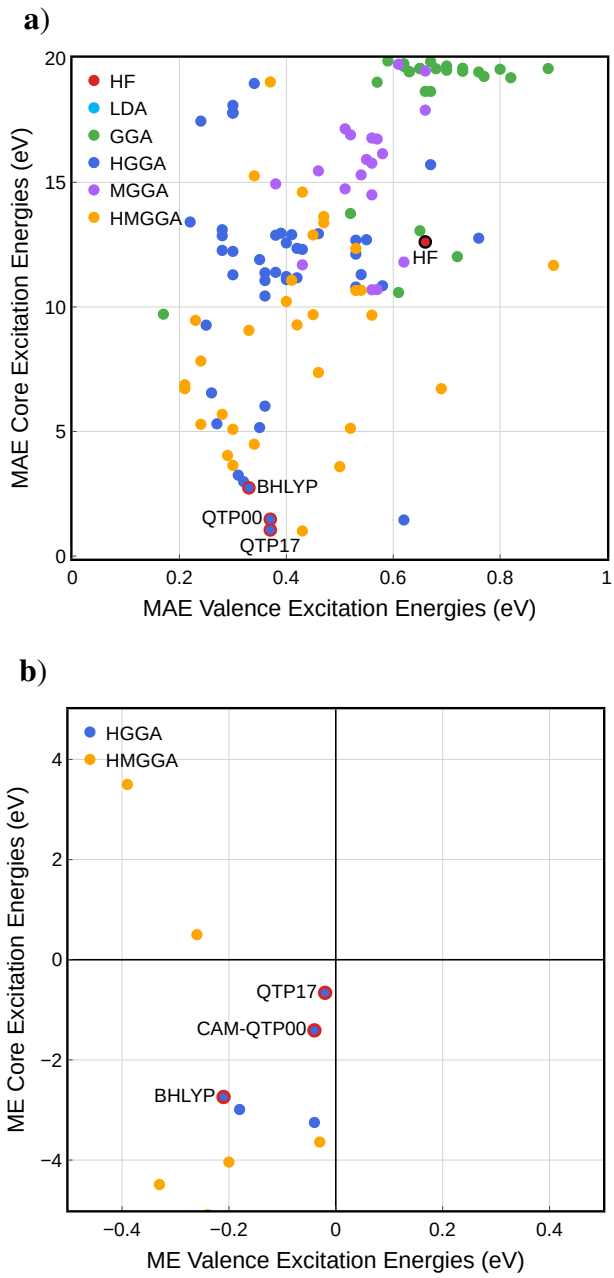


Figure 2.1: Mean absolute errors (MAEs) for core and valence excitation energies for a variety of functionals computed using TDA-TDDFT are given in panel **a**. Panel **b** presents a subset of functionals exhibiting the smallest mean errors (MEs) for valence and core excitation energies. The numerical values for the data in this figure are given in Table A.6 of Appendix A.

consistent (and accurate) description of core and valence excitation energies, are in the lower left quadrant Figure 2.1a). This region of the plot is populated by hybrid functionals: hybrid generalized gradient approximation (GGA) and meta-GGA functionals. Shown plotted in Figure 2.1b) are the mean errors (MEs) for the subset of XC functionals with the smallest MAEs. In general, the same XC functionals perform best using either error metric.

BHLYP, the up-to-now de facto standard XC functional for the DFT/MRCI approach, exhibits MAEs of 0.3 and 2.6 eV for valence and core excitations, respectively. These values correspond well with the previously observed accuracies for BHLYP-based DFT/MRCI Hamiltonians for valence and core excitation energies,[44, 48] which provides a post-hoc sanity check of the indirect XC functional assessment procedure used here. The performance of BHLYP with respect to these metrics is impressive and is validated by the accuracy of the results furnished by previous parameterizations that have been based on this XC functional. However, it is clear that a more balanced description of core and valence excited states is provided by other functionals.

The best performing XC functionals based on this analysis are QTP17[144] and CAM-QTP00.[145] The former is found to perform the best, yielding MAEs of 0.36 and 0.62 eV for valence- and core excitations, respectively. The XC functionals dlDF[146] and WC04[147] also exhibit relatively balanced errors with respect to valence and core excitation. The former is a functional optimized to determine intermolecular interaction energies via the calculation of a separate dispersion correction added to the (dispersionless) dlDF result. The latter is optimized to reproduce ^{13}C NMR chemical shifts. Given the desire to employ a general XC functional, these functionals are not further considered for the present purposes, but the role of "non-standard" functionals for the generation of orbital bases may warrant subsequent study.

The design principle of the QTP17 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 exchange. 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,[148] which is known as the Koopmans' IP. The extended Koopmans' theorem (EKT) correlates each orbital energy to a corresponding IP.[149–152] Equivalently, Janak's theorem is commonly cited as a proof that Koopmans' theorem extends equally to DFT.[153] 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 KS-DFT.[154] The Quantum Theory Project (QTP) family of functionals employs as a design principle rigorous adherence to a complete IP theorem.[144, 155–157] 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.[144]

Given that QTP17 performs best under the current analysis, and the desire to keep the method as computationally efficient as possible by avoiding the use of range-separated functionals, this XC functional was selected to form the basis of the present DFT/MRCI Hamiltonian parameterization. Additionally, although BHLYP actually exhibits a smaller MAE for valence excitation energies than QTP17, the QTP17 description of valence and core excitation energies is more *balanced*. Furthermore, the ME plot in Figure 2.1(b) indicates smaller systematic errors in the QTP17 results than the BHLYP ones.

2.4.2 NEW PARAMETERIZATION: THE QE8 HAMILTONIAN

FORM OF THE DAMPING FUNCTION

As mentioned in Sec. 2.2.1, the DFT/MRCI Hamiltonian employs a semi-empirical damping of the off-diagonal Hamiltonian matrix elements (see Eqns. 1.62 and 1.63). The form of the damping function $D(\Delta E_{\mathbf{ww'}})$, is subsequently given in Eqn. 2.2. The quality of the Hamiltonian parameterization was not found to be strongly dependent on the functional form of the damping function, beyond the requirement that it decay rapidly as a function of $\Delta E_{\mathbf{ww'}}$, and the current form was chosen for its generality. Specifically, the parameter d_1 determines the value of the damping function in the limit of zero spin-coupling-averaged energy differences between CSFs, while the values of d_2 and d_3 (and the exponential functional form) ensures that the both the value of $\Delta E_{\mathbf{ww'}}$ at which the damping function begins to significantly decrease, and the rapidity with which the damping function tends to zero, can be independently varied.

Lastly, an additional constraint is enforced in the present parameterization as a consequence of the iterative generation of the reference space. Specifically, a selected CI algorithm is used that tailors the reference space $\mathscr{R}$ to provide good support for all the reference space states of interest. A consequence of this is that the dimension of the reference space Hamiltonian will depend on the number of roots, N_{root}. As N_{root} increases, the span of the reference space eigenspectrum will correspondingly broaden. Thus, the damping function, which is a function of $\Delta E_{\mathbf{ww'}}$, will exhibit a different effective scaling for different roots of the DFT/MRCI Hamiltonian. This can be observed to manifest as a moderate dependence of the DFT/MRCI state energies on N_{root}, leading to changes in excitation energies of up to 0.2 eV depending on the number of roots computed.

This is particularly a problem if a large number ($\gtrsim 20$) of roots are sought. To ameliorate this effect, d_2 is constrained such that $D(\Delta E_{\mathbf{ww'}}) \leq 0.01$ for $\Delta E_{\mathbf{ww'}} = 1.0$ $\mathrm{E_h}$ which corresponds to $d_2 \geq -\ln(0.01) \approx 4.605$. This choice was found to strike a balance between maintaining the accuracy of the VEEs, while mitigating the N_{root} dependence of the excitation energies.

Analysis of Fitted Parameters

In the following analysis, the R2017 Hamiltonian[43] is used as the primary DFT/MRCI method of comparison. The parameters in R2017 were optimized employing experimental UV/Vis absorption spectra peak maxima, in clear contrast to the current approach. Furthermore, the fit set used in Ref. [43] to optimize the parameters for this Hamiltonian was comprised exclusively of organic chromophores, which is also the emphasis of the present work. Subsequent DFT/MRCI parameterizations focused on an improvement of the description of transition metal complexes[41] and doubly-excited states[47]. More specific comparisons for these classes of excitations may be the subject of future work. The current DFT/MRCI Hamiltonian, described in Section 1.4.2, and using the damping function given in Eqn. 2.2, is hereafter denoted as QEn. This choice is meant to signify a parameterization obtained by fitting to the QUEST databases and a Hamiltonian employing an exponential decay function in which the exponential argument is raised to the power $d_3 = n$. For reasons detailed below, the final choice of the value of d_3 was 8. As such, the final DFT/MRCI Hamiltonian is denoted QE8.

Table 2.1 presents a summary of the optimized QE8 DFT/MRCI Hamiltonian parameter values, as well as the corresponding MAEs relative to the benchmark ab initio excitation energies of the fitting set. Parameter sets for configuration selection parameter values δE_{sel} of 1.0 and 0.8 E_h have been included. For comparison, the same results are shown for the R2017 DFT/MRCI

Table 2.1: Optimized QE8 Hamiltonian parameter values and MAEs (in eV) for both the fitting and validation sets. Values of the configuration selection parameter δE_{sel} are given in units of E_h. For comparison, the parameters of, and MAEs yielded by the R2017 Hamiltonian are shown alongside. The MAEs are split into subsets for the different spin multiplicities present in the fitting and validation sets, as well as the total across all multiplicities.

Hamiltonian	δE_{sel}	p_C	p_X	d_1	d_2	d_n
QE8	1.0	0.426	0.255	0.690	4.61	8
	0.8	0.419	0.258	0.712	4.69	8
	δE_{sel}	p_C	p_X	p_1[a]	p_2	
R2017	1.0	0.503	0.359	0.564	1.857	
	0.8	0.501	0.357	0.574	1.927	

Fitting Data MAEs

Hamiltonian	δE_{sel}	Singlet	Triplet	Doublet	Total
QE8	1.0	0.18	0.14	0.17	0.16
	0.8	0.22	0.15	0.20	0.20
R2017	1.0	0.37	0.29	0.28	0.33
	0.8	0.37	0.30	0.28	0.34

Validation Data MAEs

Hamiltonian	δE_{sel}	Singlet	Triplet	Doublet	Total
QE8	1.0	0.15	0.15	0.22	0.16
	0.8	0.21	0.18	0.23	0.20
R2017	1.0	0.27	0.25	0.38	0.27
	0.8	0.28	0.26	0.40	0.28

[a] 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.

Hamiltonian.[44] The subsequent discussion will exclusively address the $\delta E_{sel} = 1.0$ E_h results, but the slightly less accurate $\delta E_{sel} = 0.8$ E_h results are shown for consistency with previous parameterizations.[29, 41, 43, 44, 47]

As noted previously, the Coulomb integral scaling term, p_C, primarily scales the integrals corresponding to the hole-particle attraction. In the TDDFT formalism, integrals of this form are explicitly scaled by x_{HF}, the proportion of exact exchange present in the XC functional. A broadly similar relation can be expected to be followed by the parameter p_C, in line with previous studies.[29, 44] Indeed, we find an optimized value of $p_C = 0.426$, which is to be compared with a value of $1 - x_{HF} = 0.38$ for the QTP17 XC functional. In the Hamiltonian optimization

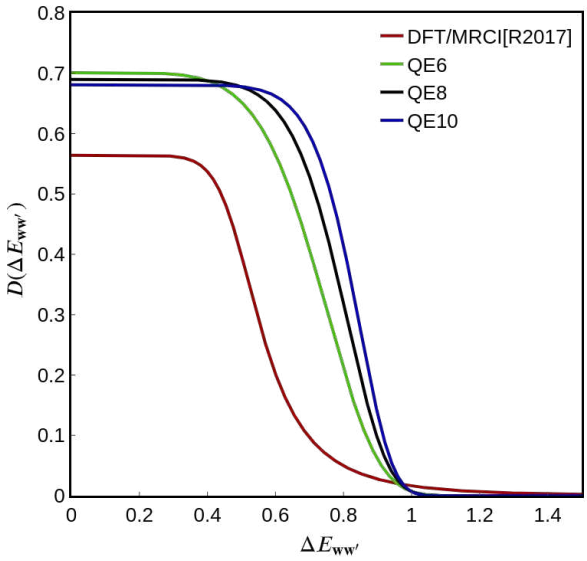


Figure 2.2: The damping function applied to the off-diagonal matrix elements of the DFT/MRCI Hamiltonian for the current (QE8) and previous (R2017) parameterizations.

procedure, the damping function parameters d_1 and d_2 were optimized simultaneously using the Nelder–Mead optimization algorithm, while the exponent d_3 was scanned over integer values

between 4 and 12. The resulting optimized QE8 damping functions, as well as the damping function of the R2017 Hamiltonian, are shown plotted in Figure 2.2. Also included for comparison are the damping functions that result from setting $d_3 = 6$ and 10, which correspond to the Hamiltonians QE6 and QE10, respectively. It is immediately apparent that the new QEn parameterizations incorporate contributions from the off-diagonal matrix elements of the DFT/MRCI Hamiltonian at a much higher level than the previous R2017 formulation. The limiting value of the damping function as $\Delta E_{\mathbf{ww'}} \rightarrow 0$, as well as the value of $\Delta E_{\mathbf{ww'}}$ at which the damping function begins to decrease significantly, are markedly greater than that for the damping function R2017. Furthermore, as d_3 increases, the onset of the damping occurs correspondingly later. While the QE8 ($d_3 = 8$) error metrics were significantly better than QE4 ($d_3 = 4$) and QE6 ($d_3 = 6$), the QE10 ($d_3 = 10$) parameterization resulted in no appreciable improvement (ΔMAE ≤ 0.01 eV) in the error metrics. Hence, the value of $d_3 = 8$ was adopted for the final Hamiltonian parameterization.

As shown in Table 2.1, the QE8 Hamiltonian is able to achieve highly impressive agreement with the theoretical best estimates of VEEs. The fitting set is comprised primarily of near-FCI VEEs for small molecules and CC3 values for larger molecules. For these values, not only was the MAE 0.16 eV, but it was largely independent of multiplicity. This value is approximately half the error of 0.33 eV yielded by the R2017 Hamiltonian.

A notable by-product of employing a different XC functional to generate the KS orbital basis is the significant decrease in the size of the FOIS for the QTP17-based QE8 parameterization in comparison to the BHLYP-based R2017 Hamiltonian. The QE8 Hamiltonian resulted in CSF bases that were roughly half as large as the corresponding R2017 expansion, with an average ratio of 0.53 being found for the molecules contained in the fitting set. Furthermore, this ratio is

found to decrease with increasing molecular size; the QE8 CSF expansions become *more* compact relative to the R2017 equivalent as the dimension of the orbital basis increases. This can be attributed to the lower occupied valence orbital energies furnished by the QTP17 functional, in concert with the orbital energy-based selection criterion given in Eq. 1.64, yielding a smaller effective orbital basis from which to generate FOIS configurations. Of course, simply employing the QTP17 functional the existing R2017 parameterization yields dramatically larger error metrics (i.e. MAE of 0.88 eV for the validation set) than either the optimized R2017 or QE8 Hamiltonians.

While it was previously noted that the fitting set was comprised of electronic excitations involving states of varied electronic character (e.g., optically allowed single excitations, Rydberg, intramolecular charge-transfer, etc.), it is emphasized here that VEEs involving doubly-excited electronic character were also included. A recent work[47] attempted to improve the performance of DFT/MRCI with respect to this historically problematic class of transitions via a new parameterization that introduces configuration-specific hole-particle and two-hole-two-particle Coulomb and exchange scaling parameters. Nevertheless, the present results show that a comparable improvement in performance can be achieved without the introduction of additional parameters. Table 2.2 summarizes the performance of the QE8 and R2017 Hamiltonians for the doubly-excited states included in the fitting set. The MAE furnished by QE8 for these excitations is only 0.26 eV, compared to 0.61 eV for R2017. The largest error by a significant margin is for the 3^1A_1 state of formaldehyde (0.69 eV) for which the TBE of 10.34 eV differs dramatically from the CCSDT (10.79 eV) and CC3 (11.20 eV) results; a challenging excitation to converge. Conversely, the $\pi^2 \rightarrow \pi^{*2}$ transition in butadiene is reproduced to 0.02 eV.

PERFORMANCE ON VALIDATION DATA

Like the training set, the validation set is also culled from transitions across all 6 QUEST

Table 2.2: Errors of the QE8 and R2017 DFT/MRCI Hamiltonians for doubly-excited states relative to theoretical best estimates (TBEs). All values are given in units of eV.

			Error	
Molecule	Transition	TBE	R2017	QE8
butadiene	$1^1A_g \rightarrow 2^1A_g$	6.51	-0.26	0.02
ethylene	$1^1A_g \rightarrow 2^1A_g$	13.07	-0.12	0.39
glyoxal	$1^1A_g \rightarrow 2^1A_g$	5.48	-0.53	0.30
formaldehyde	$1^1A_1 \rightarrow 3^1A_1$	10.45	-1.24	-0.69
nitrosomethane	$1^1A' \rightarrow 2^1A'$	4.84	-0.66	-0.07
nitroxyl	$1^1A' \rightarrow 2^1A'$	4.40	-0.86	-0.04

databases (Table 2.2); it is comprised of 165 singlet, 102 triplet, and 14 doublet vertical excitation energies. Impressively, the MAE across all excitation types is identical to that for the fitting set: 0.16 eV. Figure 2.3 breaks down the MAEs for the singlet, triplet and doublet excitation classes for both the training and validation sets. As the figure evinces, the errors for the QE8 Hamiltonian are consistent across all excitation classes (< 0.2 eV) and roughly half the magnitude of the errors for the R2017 parameterization. Further insight is gained from the histograms of errors for the training and validation sets for both the QE8 and R2017 Hamiltonians, as shown in Figure 2.4. Firstly, the QE8 error distribution is approximately centered around zero, exhibiting mean errors of -0.001 and 0.06 eV for the training and validation sets, respectively. In contrast, the R2017 Hamiltonian yields excitation energies that are evidently systematically red-shifted from the ab initio benchmark values, and yields mean errors of -0.32 and -0.26 eV, for the training and validation sets, respectively. The most likely origin of this is the previously noted empirical observation that for most organic chromophores, UV/Vis band maxima will gen-

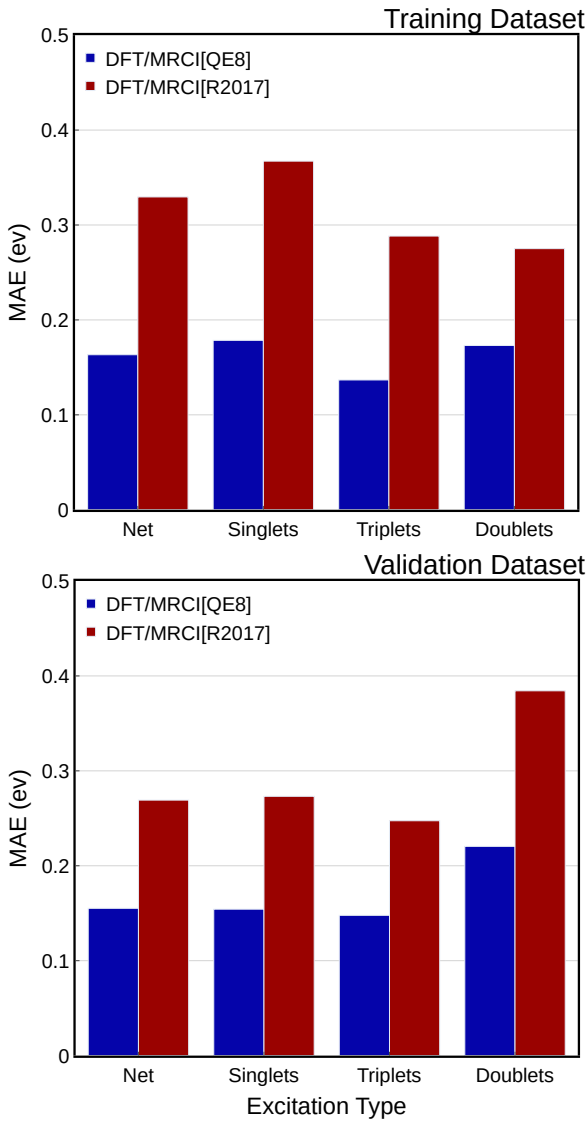


Figure 2.3: Mean absolute errors for singlets, triplet, and doublets in the fitting and validation sets for the new (QE8) and a previous (R2017) DFT/MRCI parameterization.

erally be red-shifted relative to the corresponding electronic vertical excitation energies.[111, 112] The inclusion of such data in the fitting set for the R2017 Hamiltonian thus manifests as a systematic underestimation of VEEs.

PERFORMANCE OF APPROXIMATE DFT/MRCI METHODS

Recent years have seen the development of approximate DFT/MRCI methods that exhibit an even higher level of computational efficiency, termed *p*-DFT/MRCI[45] and DFT/MRCI(2).[46] The *p*-DFT/MRCI method prunes the FOIS by removing those configurations that do not contribute appreciably to a second-order perturbation theory estimate of the energies of the states of interest. The surviving configurations are used to construct the DFT/MRCI Hamiltonian, while the contributions of the pruned ones are treated approximately at the level of second-order Rayleigh-Schrödinger perturbation theory. In the DFT/MRCI(2) approach, the construction of Hamiltonian matrix elements between FOIS configurations is entirely obviated. Instead, a small effective Hamiltonian is constructed using second-order generalized van Vleck perturbation theory (GVVPT2)[158], the eigenvalues of which provide approximations of those of the DFT/MRCI Hamiltonian. It has previously been demonstrated that both the *p*-DFT/MRCI and DFT/MRCI(2) approaches can yield computational savings of upwards of two orders of magnitude for large molecules while retaining the accuracy of the parent DFT/MRCI method.[45, 46]

Table 2.3 shows the MAEs furnished by the *p*-DFT/MRCI and DFT/MRCI(2) approaches in combination with the QE8 Hamiltonian with respect to fitting and validation data sets discussed above. The DFT/MRCI results are shown alongside for ease of comparison. For both the fitting and validation sets, the MAEs are reproduced to within 0.01 eV for both *p*-DFT/MRCI and

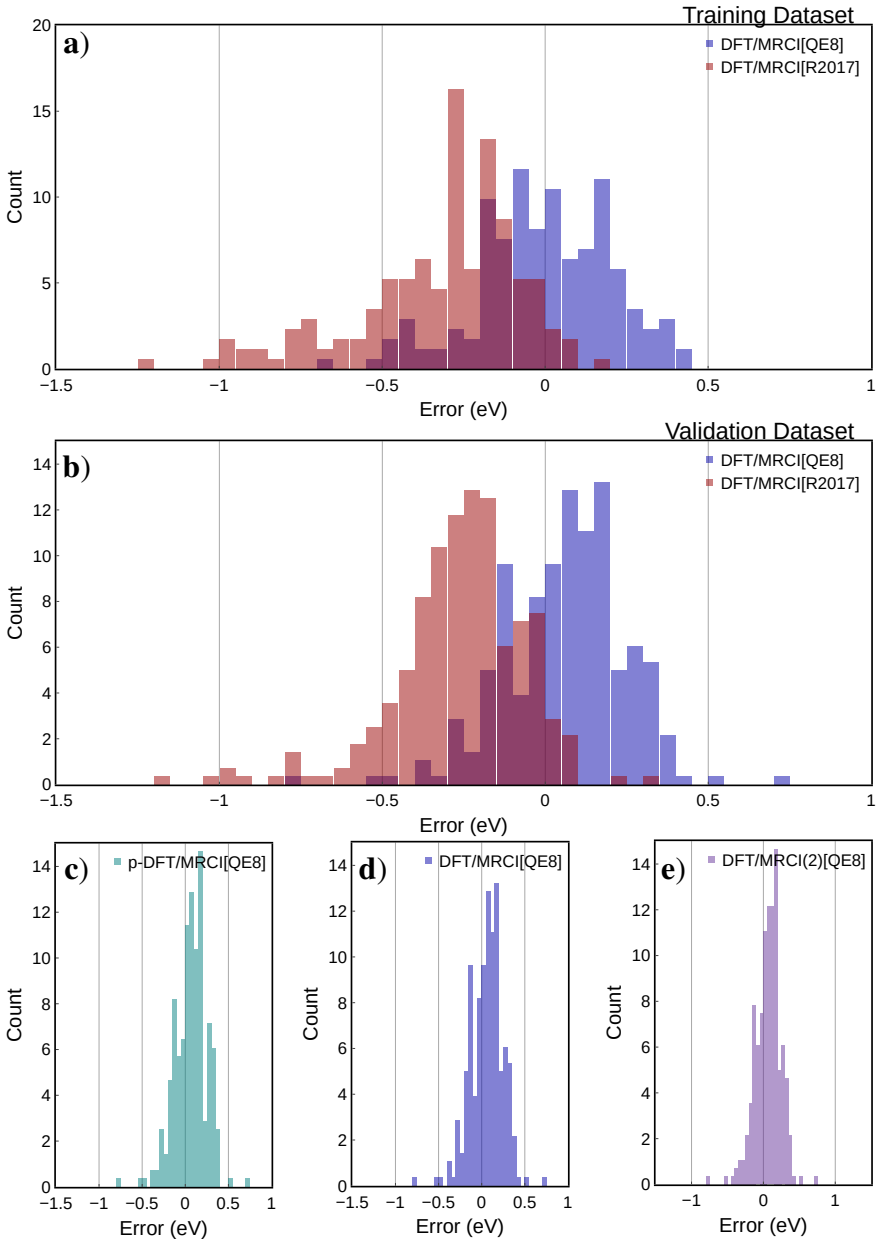


Figure 2.4: Signed vertical excitation energy errors for the QE8 and R2017 Hamiltonians relative to both the fitting and validation data sets. The performance of the *p*-DFT/MRCI and DFT/MRCI(2) methods (panels (c) and (e)) are also compared to the parent DFT/MRCI method (panel (d). In all calculations a value of the configuration selection parameter $\delta E_{sel} = 1.0$ E$_h$ was used.

Table 2.3: Comparison of errors for DFT/MRCI, *p*-DFT/MRCI and DFT/MRCI(2) QE8 Hamiltonian vertical excitation energies relative to both the fitting and validation data sets. All values are given in units of eV. All values were computed using a configuration selection parameter value of $\delta E_{sel} = 1.0$ E$_h$

Fitting Data MAEs				
Method[QE8]	Singlet	Triplet	Doublet	Total
DFT/MRCI	0.18	0.14	0.17	0.16
p-DFT/MRCI	0.18	0.14	0.17	0.16
DFT/MRCI(2)	0.17	0.14	0.17	0.16
Validation Data MAEs				
Method[QE8]	Singlet	Triplet	Doublet	Total
DFT/MRCI	0.15	0.15	0.22	0.16
p-DFT/MRCI	0.16	0.15	0.22	0.16
DFT/MRCI(2)	0.15	0.14	0.22	0.15

DFT/MRCI(2) relative to the parent DFT/MRCI method. Additionally, the error histograms in Figure 2.4 are particularly illustrative in demonstrating how closely *p*-DFT/MRCI and DFT/MRCI(2) reproduce the DFT/MRCI results. The mean errors for the fitting (validation) sets are 0.004 (0.06) eV for *p*-DFT/MRCI and 0.006 (0.07) eV for DFT/MRCI(2), which is in excellent agreement the DFT/MRCI value of -0.001 (0.06) eV.

Both of these approximate methods employ the same Hamiltonian parameterization as the parent method. Separate parameter optimizations were performed for both *p*-DFT/MRCI and DFT/MRCI(2), but the improvement in the error metrics, ≈ 0.02 eV in the MAE, was minimal. Therefore, the simplicity of a single parameterization for both DFT/MRCI and the approximate methods has been adopted.

In summary, the QE8 Hamiltonian does not negatively affect the previously observed agreement between parent DFT/MRCI and related *p*-DFT/MRCI and DFT/MRCI(2) methods. The

consistency in the computed excitation energies is impressive, with *p*-DFT/MRCI and DFT/MRCI both exhibiting a mean difference of ≤ 0.01 eV with DFT/MRCI overall excitation classes in the validation set.

2.5 Representative Applications

To conclude this work, two representative applications are included that illustrate the range of applicability of the DFT/MRCI family of methods in conjunction with the QE8 Hamiltonian: (i) the calculation of electronic spectra of the five nucleobases, and (ii) the calculation of the low-lying states of chlorophyll *a*. The results are compared to established high-level ab initio electronic structure methods for each example. The purpose of the former application is to show the ability of the QE8 Hamiltonian to yield oscillator strengths of good quality, in addition to accurate VEEs. The second application highlights the ability of the DFT/MRCI family of methods to treat large molecules at low cost while retaining high levels of accuracy.

2.5.1 Electronic Spectra of Nucleobases

One important application area for DFT/MRCI is the simulation of optical absorption spectra of molecules. The ability to reliably compute vertical excitation energies and oscillator strengths for a potentially large number of valence electronic states is crucial for these simulations. Accurate simulations of absorption spectra that explicitly describe non-adiabatic vibronic coupling effects have been carried out previously using DFT/MRCI potentials and non-adiabatic couplings,[159, 160] and the present QE8 parameterization has been conceived with applications such as these

in mind.

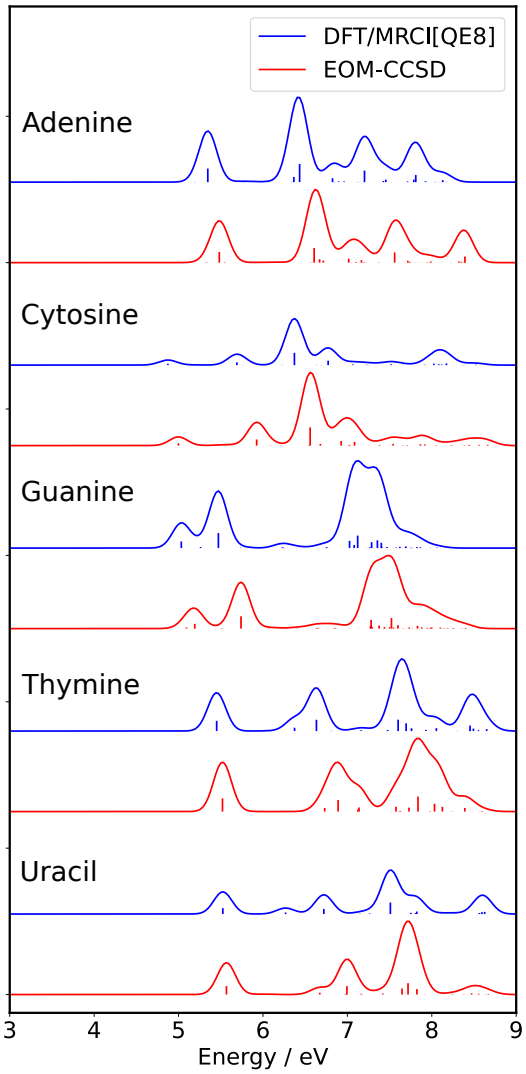


Figure 2.5: Vertical excitation energies and oscillator strengths for nucleobases adenine, cytosine, guanine, thymine, and uracil. DFT/MRCI[QE8] is compared to EOM-CCSD results.

The QE8 parameterization was performed using benchmark vertical excitation energies only, typically including ≤ 5 states per molecule. Although there has been recent work to construct equivalent benchmark QUEST databases for oscillator strengths[27, 161], such information was

not included in the parameter optimization process. In order to be useful in the simulation of absorption spectra, the QE8 Hamiltonian needs to be capable of furnishing both accurate excitation energies and oscillator strengths. To assess this capability, we consider the calculation of purely electronic absorption spectra for the five nucleobases (adenine, cytosine, guanine, thymine, and uracil). The results are compared to a well-established ab initio method: EOM-CCSD. This method, in combination with the aug-cc-pVTZ basis, has a similar accuracy to the QE8 Hamiltonian for valence excitation energies of organic molecules, exhibiting an MAE of 0.13 eV relative to the theoretical best estimates for the 442 excitation energies that span all of the QUEST data sets. Additionally, this method is capable of furnishing accurate oscillator strengths.[27, 161]

For each molecule, VEEs and oscillator strengths were computed at both the EOM-CCSD and DFT/MRCIQE8 levels of theory using the aug-cc-pVDZ basis. The resulting electronic spectra are shown in Fig. 2.5. Here, in order to aid a visual comparison, the spectra are also shown convoluted with a Gaussian lineshape with a full width at half maximum (FWHM) of 0.2 eV. Broadly speaking, the agreement between the DFT/MRCI and EOM-CCSD vertical excitation energies is excellent, with differences between excitations involving equivalent states observed to be $\leq$ 0.3eV; none of the spectra shown in Fig. 2.5 have had any shifts applied. Moreover, the relative oscillator strengths are found to be in good agreement. Notably, however, the percent difference for some of the bright transition oscillator strengths was observed to be up to 40% and may warrant subsequent study. For present purposes, we wish only to demonstrate that the vertical excitation energies of the optically bright electronic states and relative oscillator strengths agree well with the corresponding EOM-CCSD results over a range of multiple eV.

2.5.2 APPLICATION TO LARGE MOLECULES

The development of perturbative DFT/MRCI approaches enables the application of these methods to very large molecular system at low computational cost. This section considers the calculation of the VEEs of the low-lying singlet excited states of chlorophyll a using DFT/MRCI(2) and the QE8 Hamiltonian. In particular, considered here are the four "Gouterman-type" states, corresponding to excitation between the HOMO-1, HOMO, LUMO and LUMO+1 π and π^* orbitals localized on the chlorin ring.[162, 163] In order of increasing energy, these are labeled the Q_y, Q_x, B_x, and B_y states. The ground state minimum energy geometry is shown in Figure 2.6 b). It is commonly accepted that the phytyl chain is essentially electronically decoupled from the chlorin ring on which the Q and B state excitations are localized. As such, when computing these states, it is common practice to replace the phytyl chain with a methyl group in order to reduce computational costs. The geometry of this reduced system is shown in Figure 2.6 a). The computational expediency of the DFT/MRCI(2) method, however, allows for the treatment of the full, 137-atom system. In a recent study, Sirohiwal and co-workers[16] published benchmark

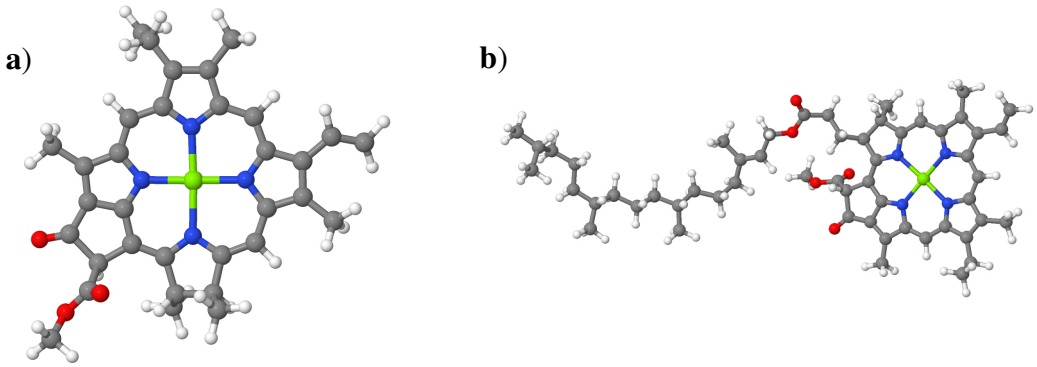


Figure 2.6: Molecular geometry for a model of chlorophyll *a* (top, a)) in which the hydrocarbon tail is terminated with a methyl group, as well as the complete structure of chlorophyll *a* (bottom, b))

domain local pair natural orbital similarity transformed EOM-CCSD (DLPNO-STEOM-CCSD) calculations of the VEEs of the Q and B states for the model system in which a methyl group is substituted for the phytyl chain. In that work, the UV/Vis absorption spectrum of the Q and B bands was simulated in the Franck–Condon approximation, in which the vibrational structure of the excited states was incorporated into the simulation while neglecting the non-adiabatic coupling between relevant electronic states. On the basis of these calculations, and engendered by good agreement with experiment, the authors could report the best estimates of the VEEs that resulted in the best agreement between the simulated and experimental spectra.

Table 2.4: Vertical excitation energies of four Gouterman-type excited states of chlorophyll a for both the full and model structures. In all cases, the def2-TZVP basis set was used. All values are given in units of eV.

	STEOM-CCSD	DFT/MRCI(2)		
State	model[a]	model	full	TBE
Q_y	1.75	2.10	2.06	1.99
Q_x	2.24	2.40	2.36	2.30
B_x	3.17	3.21	3.11	3.12
B_y	3.40	3.40	3.33	3.38

[a] Ref. [16].

In Table 2.4, VEEs of the Q and B states are reported for both the full chlorophyl a structure, including the phtyl chains, as well as the reduced model geometry, with the phytyl chain removed. These energies are computed at the def2-TZVP/DFT/MRCI(2) level of theory using the QE8 Hamiltonian. Shown alongside are the def2-TZVP/DLPNO-STEOM-CCSD results of Sirohiwal et al.[16] as well as the current theoretical best estimates. It is found that DFT/MRCI(2) furnishes results in very close agreement with the DLPNO-STEOM-CCSD results and the TBE values. The computations on the full molecule, in particular, are generally in excellent agree-

ment with the TBEs, with a maximum deviation of only 0.07 eV observed for the Q_y state. We thus conclude that the QE8 Hamiltonian is capable of producing excellent quality results for this prototypical system, albeit at very modest computational costs.

2.5.3 COMPUTATIONAL EFFICIENCY: TIMINGS

To close, a comment is made on the computational efficiency of the DFT/MRCI family of methods: DFT/MRCI, *p*-DFT/MRCI, and DFT/MRCI(2). Table 2.5 shows the computation CPU times for each of the applications presented in Sec. 2.5 for each of these methods. These timings include only the post-DFT components of the calculations. The results shown here were generated using the GRaCI package[125], which uses the PySCF[126–130] program package to solve the KS-DFT equations. However, this aspect of the computation is highly modular and other packages may be employed in this respect, the computational efficiency of which will vary widely.

Considering first the nucleobase calculations, Table 2.5 clearly demonstrates how computationally tractable even extremely large computations are using the DFT/MRCI family of methods. No computation of the UV/Vis spectra shown in Fig. 2.5, which involved the determination of the 30 lowest-energy roots, took longer than 45 minutes using the DFT/MRCI method running in serial. Using the DFT/MRCI(2) method results in large computational savings, with an average CPU time of just 39 seconds. Moreover, the resulting spectra are found to be essentially identical to the more expensive DFT/MRCI results, consistent with the agreement in vertical excitation energies presented in Table 2.3. Likewise, the *p*-DFT/MRCI method also yields significant savings relative to the parent DFT/MRCI method, whilst furnishing VEEs within 0.01 eV of the

Table 2.5: CPU times for representative applications (in units of seconds).[a]

Molecule	DFT/MRCI	*p*-DFT/MRCI	DFT/MRCI(2)	N_{roots}
Adenine	1851	194	50	30
Cytosine	735	47	27	30
Guanine	2604	273	61	30
Thymine	880	101	35	30
Uracil	276	38	22	30
chlorophyll *a* (model)	–	–	265	6
chlorophyll *a* (full)	–	–	2044	6

[a] Values in the table are CPU times (excluding KS-DFT computation) and correspond to calculations run using a single thread of a Intel Xeon Gold 6130 processor. The aug-cc-pVDZ basis set was employed for the nucleobases. The aug-cc-pVDZ basis set was used for the nucleobases, and the def2-TZVP basis was used for both chlorophyll *a* systems.

DFT/MRCI values. However, the *p*-DFT/MRCI CPU times are found to be significantly greater than those for DFT/MRCI(2). Although, due to the increased amount of correlation captured by the QE8 Hamiltonian relative to previous DFT/MRCI Hamiltonians, the configuration pruning algorithm used in *p*-DFT/MRCI is rendered less effective than previously reported[45], and DFT/MRCI(2) now emerges as the clear method of choice for reduced cost DFT/MRCI calculations.

Finally, the timings for the chlorophyll *a* calculations, as performed at the def2-TZVP/ DFT/MRCI(2) level of theory, are considered here. For the model structure, the required CPU time is less than 5 minutes to compute six roots. For the full structure calculation, including the phytyl chain, this increases to 34 minutes, which is still a remarkably low computational cost for a system of this size, especially given the impressive accuracy of the results (see Table 2.4).

2.6 CONCLUSION

Presented here is a newly-developed DFT/MRCI Hamiltonian, termed QE8, parameterized using only benchmark-level ab initio data in the form of VEEs taken from the QUEST databases. With an eye to the development of a DFT/MRCI Hamiltonian that gives a balanced description of core-excited states, a new XC functional was adopted, namely, the QTP17 functional. The performance of a Hamiltonian designed to yield highly accurate core excitation energies, CVS-QE*n*, will be discussed in Chapter 3. The present Hamiltonian yields VEEs with an MAE of 0.16 eV with respect to theoretical best estimates for valence-excited states of organic molecules. Importantly, by fitting the Hamiltonian parameters to only ab initio data, the systematic underestimation of VEEs exhibited by previous DFT/MRCI Hamiltonians is ameliorated. Furthermore, doubly-excited states, a previously problematic class of states for DFT/MRCI methods, can now be computed with high accuracy. In combination with the newly-introduced DFT/MRCI(2) variant, the QE8 Hamiltonian opens the door to the accurate simulation of excited states of large organic molecules at very low cost, as exemplified by the current application to the chlorophyll *a* molecule.

The Hamiltonian design principle employed in the present work involved only the reproduction of benchmark VEEs. Future efforts will examine, and potentially improve, the performance of DFT/MRCI methods for the computation of other quantities relevant for simulations in spectroscopy and dynamics. Specifically, benchmark electronic properties, such as oscillator strengths and electronic moments, are becoming increasingly available[27, 161, 164]. Thus, the degree to which DFT/MRCI wave functions furnish accurate electronic properties will be the

subject of future investigation.

The ability of DFT/MRCI to produce accurate potential energy surfaces has been studied only indirectly via, for example, the parameterization of vibronic Hamiltonians. The DFT/MRCI reference space is iteratively generated anew at each nuclear configuration, meaning that the underlying potential is not entirely smooth and complicates the evaluation of energy gradients and non-adiabatic couplings. However, recent work employing Gaussian process and kernel ridge regression approaches to learn potential energy surfaces[165] is likewise applicable to DFT/MRCI potentials,[166] providing a route to structure optimization and dynamics simulations. The accuracy of excited structural minima derived from DFT/MRCI potential surfaces, including minimum energy conical intersections and the corresponding branching spaces, will likewise be evaluated in an upcoming work.

3

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

3.1 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 (XFEL) facilities 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.[12, 15] Furthermore, the high-order harmonic generation (HHG) process, which can be applied in laboratory settings, can extend up to the soft X-ray range and can be used to generate ultrashort pulses in the attosecond domain.[12] 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.[12]

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 shortcomings, including difficulty in the simultaneous description of states of different electronic character (i.e., charge-transfer, Rydberg, multireference, etc.), although the extent of these problems is

exchange-correlation (XC) functional dependent.[89, 167]

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,[168] 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-dependent density functional theory (LR-TDDFT) is currently the most common choice for computing excitation spectra, being far more computationally efficient than WFT-based methods.[169] Early applications of the CVS approximation to TDDFT[170–172] 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.[173] Nevertheless, TDDFT suffers from systemic limitations; it struggles to describe charge-transfer states, double-excitations, and Rydberg states.[82, 84] 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).[82, 174–176] 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.[82]

A time-independent alternative to TDDFT is orthagonality-constrained DFT (OCDFT).[177, 178] 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 a 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.[179–183] Employing an orbital optimized approach, one may target and directly solve for individual states, including doubly-excited states,[179] with sub-eV error.[179–183] 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. However, recent developments[177, 179, 180, 184, 185] have sought to address 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. To this end, Carter-Fenk et al. describe the electron-affinity formulation of TDDFT within the Tamm–Dancoff approximation (TDA), so-called EA-TDA.[90] Comprising a simpler approach, the optimized KS basis of the cationic ($n-1$ electron) reference system is used to recover orbital relaxation effects and the particle–hole interaction correction. EA-TDA has an RMSD of 0.5 eV with respect to the static exchange approximation (STEX) method for K-shell excitation energies.

The Bethe–Salpeter Equation[186] (BSE) describes a framework to solve the particle-hole Green's function, which has been applied together with the GW Green's Function method[187] to describe core excitation energies within the BSE@GW formalism.[188] This work follows recent developments that have applied the GW and BSE frameworks to simulate core-electron excitation

and ionization.[189–191] The GW scheme has been shown to yield core ionization energies within 0.3 eV of experimental values,[192] and the BSE@GW formalism yields an impressive 0.6 eV MAE with respect to experimental core excitation energies. While past work has cited ease-of-use and computational expense as barriers to greater adoption of the BSE scheme,[193] developments have since ameliorated these former drawbacks.[189, 193] The limitations of this method, as in TDDFT, stem from its single-particle description of electrons.

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[168] with the introduction of a CVS-ADC(2) implementation by Barth and Schirmer in 1985.[194] 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 $\mathscr{O}(N^5)$ 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).[195–197] 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.[196, 197] However, the ADC(2)-x, and ADC(3) methods formally scale as $\mathscr{O}(N^6)$, 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.[198, 199] This method formally scales as $\mathcal{O}(N^6)$ and has been demonstrated capable of producing highly accurate vertical excitation energies,[200–204] as well as absorption,[198–200, 204] and photoelectron spectra.[205–207] 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, converges within 0.1 eV of some experimental vertical core excitation energies at the complete basis set limit.[200] The ionization potential variant, CVS-EOMIP-CCSDT, is found to be similarly accurate, furnishing ionization energies within 0.1 eV of experimentally determined values.[208] However, CVS-EOM-CCSDT has a steep asymptotic scaling of $\mathcal{O}(N^8)$,[209] 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-black box nature.[210]

The description of static correlation in multireference 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 multireference states typically falls to the complete and restricted active-space self-consistent field (CASSCF and RASSCF) family of methods.[211, 212] Dynamic correlation energy is generally accounted for using second-order perturbation theory, most commonly via the use of the CASPT2 and RASPT2 methods.[213–215] 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 computa-

tion 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[216, 217] 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).[218, 219] 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.[218] Likewise, MR-EOM-CC theory is similarly implemented.[220]

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. The combined density functional theory and multireference configuration interaction (DFT/MRCI) approach is a method that, somewhat uniquely, has the potential to simultaneously satisfy all of these requirements.[30] Here, short, automatically tailored CI expansions aimed at the capturing of static correlation are paired with DFT-specific corrections to recover the preponderance of the remaining dynamic correlation efficiently. 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.[41, 43, 44, 47, 107] In particular, Chapter 2 presents 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[48] and found to furnish reasonably 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 observation can be attributed to the use of an XC functional (BHLYP) that yields KS orbital energies whose differences give an unbalanced zeroth-order approximation of valence and core excitation energies. In Chapter 2, with an eye to this work, a new XC functional was identified that gives a significantly more balanced description of valence and core excited states: QTP17. This chapter explores 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. Consistent with the approach taken in Ch. 2, the most accurate benchmark excitation energies were employed to perform the parameterization, which results in the training and validation data sets containing only ground-state core excitation and ionization energies. However, these efforts will enable the application of these DFT/MRCI methods to the determination of spectra from electronically excited states as might be realized in an ultrafast, time-resolved

X-ray pump-probe experiment.

In Sec. 3.2, a brief description is given of the CVS-specific correction to the DFT/MRCI Hamiltonian. Much of the background information pertinent to this chapter is already contained within Ch. 1 and 2. Sections 3.3 and 3.4 respectively describe the technical details of the parameterization and an analysis of the results. Section 3.5 presents the results of some representative applications of the CVS-QE12 Hamiltonian, and Section 3.6 provides a summary of results as well as a prospectus for future work.

3.2 THE DFT/MRCI METHOD

3.2.1 IMPROVED DESCRIPTION OF CORE-EXCITED STATES

The original CVS-DFT/MRCI Hamiltonian[48] 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 exchange, x_{HF}.[221–223] 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 hole and 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, this has 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:

$$\begin{aligned}\Delta E_c = &- p_C^{(v,v)} \left[\sum_{\substack{i<j \\ i,j\in\mathscr{V}}} V_{iijj}\Delta \mathrm{w}_i \Delta \mathrm{w}_j + \frac{1}{2} \sum_{i\in S_{\overline{\mathrm{w}}}} V_{iiii}|\Delta \mathrm{w}_i| + \sum_i V_{iiii}\delta_{\Delta \mathrm{w}_i,2} \right] \\ &- p_C^{(c,v)} \left[\sum_{\substack{i<j \\ i\in\mathscr{C}}} V_{iijj}\Delta \mathrm{w}_i \Delta \mathrm{w}_j \right] + \frac{1}{4} \sum_{i\in S_{\overline{\mathrm{w}}}}^{n_{mo}} V_{iiii}\end{aligned} \tag{3.1}$$

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

3.3 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 squared 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 precomputed set of the correct target character are computed at each iteration, allowing for an unambiguous assignment of the states of interest. For a detailed discussion of the optimization procedure, the reader is referred to Chapter 2. 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 data set and the optimization and validation steps. Following this, computational details of selected representative applications are provided.

3.3.1 Training and Validation Set Data

The training and validation data are comprised of vertical core excitation and ionization energies computed at the CVS-EOM-CCSDT and CVS-EOMIP-CCSDT [134, 135] (hereafter, CCSDT) levels of theory using the aug-cc-pCVTZ basis set.[123, 124] The MRCC package[136–138] was used to perform these calculations. The training set used is similar to that reported previously by Matthews et al.[202]; however, to facilitate state assignments, it was deemed expedient to recompute these excitations to identify the symmetry labels of each state unambiguously.

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 data set 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. 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 relies 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. 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. The aim here is to reproduce the CVS-EOM-CCSDT excitation energies for a single, common basis set. Therefore, it may be acknowledged that while these single basis set, non-relativistic calculations may be accurately called "benchmark", they do not generally correspond to "theoretical best estimates". 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.[200, 208] A few molecules in the training set, such as hypofluorous acid, nevertheless feature a high density of low-lying two-electron transitions, which are excluded from this analysis.

A summary of the training and validation data sets and their content is provided in Table B.1 of Appendix B. The optimized ground state geometries of the molecules in the data sets are taken from literature sources (generally, from the QUEST electronic structure databases)[20, 202, 224] and are additionally provided in Appendix B. In general, these correspond to aug-cc-pVTZ/CC3 optimized structures.[118]

The DFT/MRCI computations were performed using the GRaCI software package,[125] which employs PySCF for KS-DFT calculations and integral transformations.[126–130] Density-fitting was employed for the evaluation of the two-electron integrals and employed an optimized auxiliary basis set.[225] For each molecule in either data set, 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 generally involves the computation of a large number of roots per molecule. Consequently, determining an appropriate initial reference space is non-trivial, and the AutoRAS procedure introduced in Chapter 2 was employed to automate this process.

3.3.2 REPRESENTATIVE APPLICATIONS

Four sets of X-ray spectra are used to demonstrate the ability of the re-parameterized Hamiltonian to reproduce selected experimental data: (i) the C K-edge XPS of the ethyl trifluoroacetate molecule, (ii) the C, N, and O K-edge NEXAFS spectra of cytosine, (iii) the ground and excited state O K-edge NEXAFS spectra of thymine, and (iv) the C K-edge X-ray absorption spectra (XAS) of the allyl free radical. Application (iv) also includes the CVS-QE12 results for a set of 1s to SOMO excitation energies for neutral and charged radical species. Each application is compared to experimental data.[4, 226–231] 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, thymine, and allyl 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,[133] and the ADC-Connect code[232] 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 ref. [4].

For the computation of XPS of ethyl trifluoroacetate, the electronic ionization cross-sections were approximated by the squared norms of the corresponding Dyson orbitals,[233, 234] $\phi^D_{if}(r)$:

$$\phi^D_{if}(r) = \int \Psi^N_i(r, r_2, ...r_N)\Psi^{N-1}_f(r_2, ..., r_N)dr_2...dr_N \tag{3.2}$$

Above, 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. 3.2 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}:

$$G^D_{if} = \langle \phi^D_{if} | \phi^D_{if} \rangle \tag{3.3}$$

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.[235] In the present application, the neutral and cationic wave functions are expanded in a non-orthogonal orbital basis 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.[236]

3.4 RESULTS

3.4.1 SELECTION OF XC FUNCTIONAL

This section serves as a reiteration of the results presented in Sec. 2.4.1 of Chapter 2. All previous DFT/MRCI Hamiltonians have been parameterized for use with the BHLYP XC functional. As noted in Chapter 2, 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 1.57). 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. As a result, the orbital-energy differences between the core and valence regions represent underestimated excitation energies for the transitions that they correspond to. As discussed in Sec. 2.4.1, selected XC functionals were evaluated using TDA-TDDFT as a surrogate excited-state method. The error metrics for a subset of valence and core VEEs were tabulated for each functional considered. From the analysis of these error metrics, the QTP17 functional[144] was chosen to generate the underlying KS orbital basis of the present parameterization. The selected functional yields *absolute* errors for core excitation that are of the same magnitude as the valence excitation energies, despite the former corresponding to energy differences that are $100\times$ larger.

3.4.2 NEW PARAMETERIZATION: CVS-QE12

FORM OF THE DAMPING FUNCTION

Employed here is the same general form of the damping function presented in Chapter 2 (see Eqn. 2.2). 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-QE*n* Hamiltonians (where $n = d_3$), as well as the damping function used in the CVS-R2017 Hamiltonian, are shown plotted in Figure 3.1.

Consistent with the QE8 parameterization, it appeared 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 is 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, it was 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, a constraint $d_3 \geq 2.303$ is imposed 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 the errors in the lowest core transition to ≤ 0.5 eV, irrespective of the value of

N_{root}, while maintaining accurate error metrics in the excitation energies.

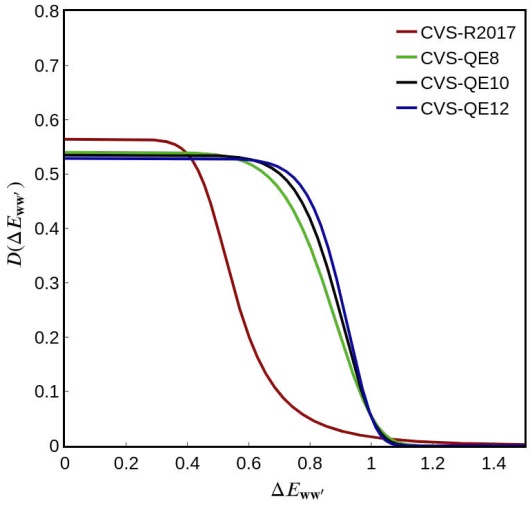


Figure 3.1: The damping function applied to the off-diagonal matrix elements of the DFT/MRCI Hamiltonian for current the (CVS-QE*n*) and previous (CVS-R2017) parameterizations.

The damping functions resulting from this constrained optimization procedure are shown in Figure 3.1 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 Sec. B.2 of Appendix B). Thus, a value of $d_3 = 12$ was adopted, corresponding to the CVS-QE12 Hamiltonian.

Analysis of Optimized Parameters

Table 3.1 presents the optimized parameter values of the CVS-QE12 Hamiltonian and the cor-

responding 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. The subsequent discussion 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 3.1: 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

To begin, a comment is made on the optimized value of the Coulomb integral scaling terms, $p_C^{(v,v)}$ and $p_C^{(c,v)}$. These coefficients scale the integrals corresponding to the hole-particle Coulomb interaction. Equivalent terms in the TDDFT formalism are explicitly scaled by x_{HF}, where x_{HF}

denotes the proportion of exact exchange (EEX) present in the XC functional (see Sec. 1.2.5). 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 $p_C^{(v,v)} = 0.426$. 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. To analyze this result, the Coulomb scaling term can be related to the scaled interaction term in the TDA-TDDFT equations (see Eqn. 1.37): $x_{HF}(ii|aa)$. Fully describing the particle-hole attraction in Eqn. 1.37 (i.e., when $x_{HF} = 1$) correctly increases the energy of the corresponding excited state. In TDDFT, the description of the particle-hole attraction is often incomplete, an effect termed *eh*-SIE, leading to greatly underestimated excitation energies[90] (see Sec. 1.3.3). Increasing $p_C^{(c,v)}$ is similar, in effect, to the down-scaling of an interaction term, $c_{HF}(ii|aa)$. The damping of this interaction lowers the resultant excitation energies. The Coulomb scaling parameter $p_C^{(c,v)}$, as an additional degree of freedom, could be compensating for correlation energy that is damped out via the damping function. Otherwise, the core-valence orbital energy differences described in the KS basis may be over-corrected by the QTP17 functional. Thus, the parameters of the CVS-QE12 Hamiltonian are constrained to lower the resultant excitation energies to compensate.

As mentioned in Sec. 1.4, the general relation holds in DFT/MRCI that $p_C \approx 1 - x_{HF}$. Regardless of the interpretation of the present parameters, it can be stated unambiguously that the differential scaling of the Coulombic particle-hole interaction via the separate terms $p_C^{(v,v)}$ and $p_C^{(c,v)}$ is a result of the description of exchange in QTP17, being a hybrid XC functional. In TDDFT, a combination of global hybrid[144, 145, 173, 221, 222, 237, 238] and short-range

corrected (SRC)[143, 145, 173, 221] functionals have been designed to correct for the self-interaction error (SIE) of core electrons by employing increased fractions of EEX. The balance of exact and semi-local exchange in these approaches can nevertheless have drawbacks,[169, 239, 240] which has been characterized as a "zero-sum trade-off" between the over-localization and over-delocalization of electrons,[72] 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 necessary due to the *global* nature of the EEX admixture in the QTP17 functional. The increased fraction of EEX arguably improves the description of the core-electron density while deteriorating that of the valence-electron density. While the differential error attributed to orbital-energy SIE is improved upon, the underlying KS orbital basis is nevertheless non-ideal.

PERFORMANCE ON TRAINING AND VALIDATION SET DATA

Figure 3.2 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 data sets. For singlet vertical core excitation energies, the CVS-QE12 Hamiltonian yields an 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

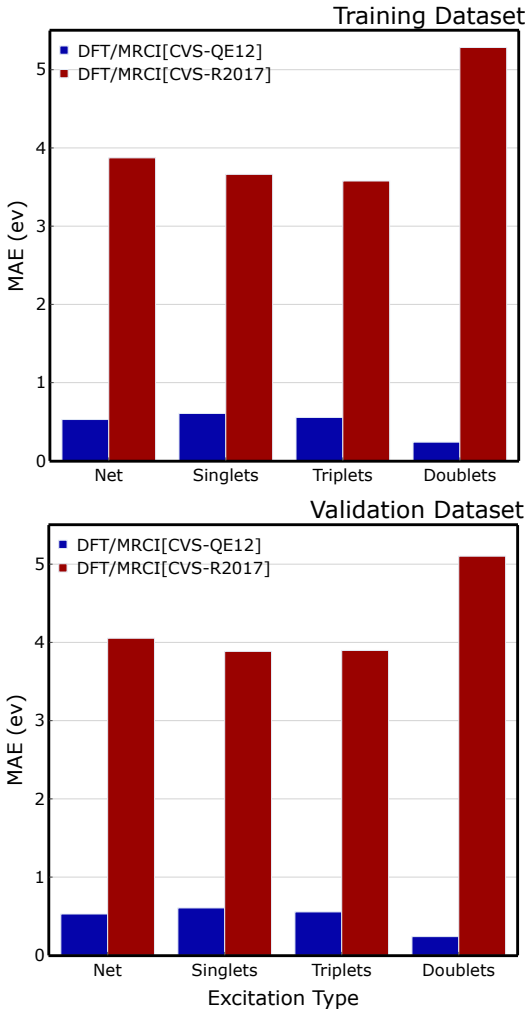


Figure 3.2: MAE of DFT/MRCI with respect to the reference data using the CVS-QE12 and CVS-R2017 Hamiltonian. The performance of each Hamiltonian is compared for the data in the training set and the external validation data set.

Hamiltonians with respect to the training set, see Table 3.1. 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 data sets is tabulated in Table 3.1. The ability to reproduce spectra is arguably better conveyed by errors in the "term values", the excitation energies relative to the low-lying excitation. These results are provided in tables A.4 and A.5 of Appendix A. The MAE in these energies is less than that in the absolute energies, determined as ≈ 0.3 eV and 0.4 eV across the training and validation sets, respectively.

Figure 3.3 illustrates the distribution of errors for the CVS-QE12 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-QE12 Hamiltonian relative to the old CVS-R2017 Hamiltonian.

PERFORMANCE OF APPROXIMATE DFT/MRCI METHODS

Finally, this section showcases 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.[45, 46] 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 3.3, 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 distribution width and center.

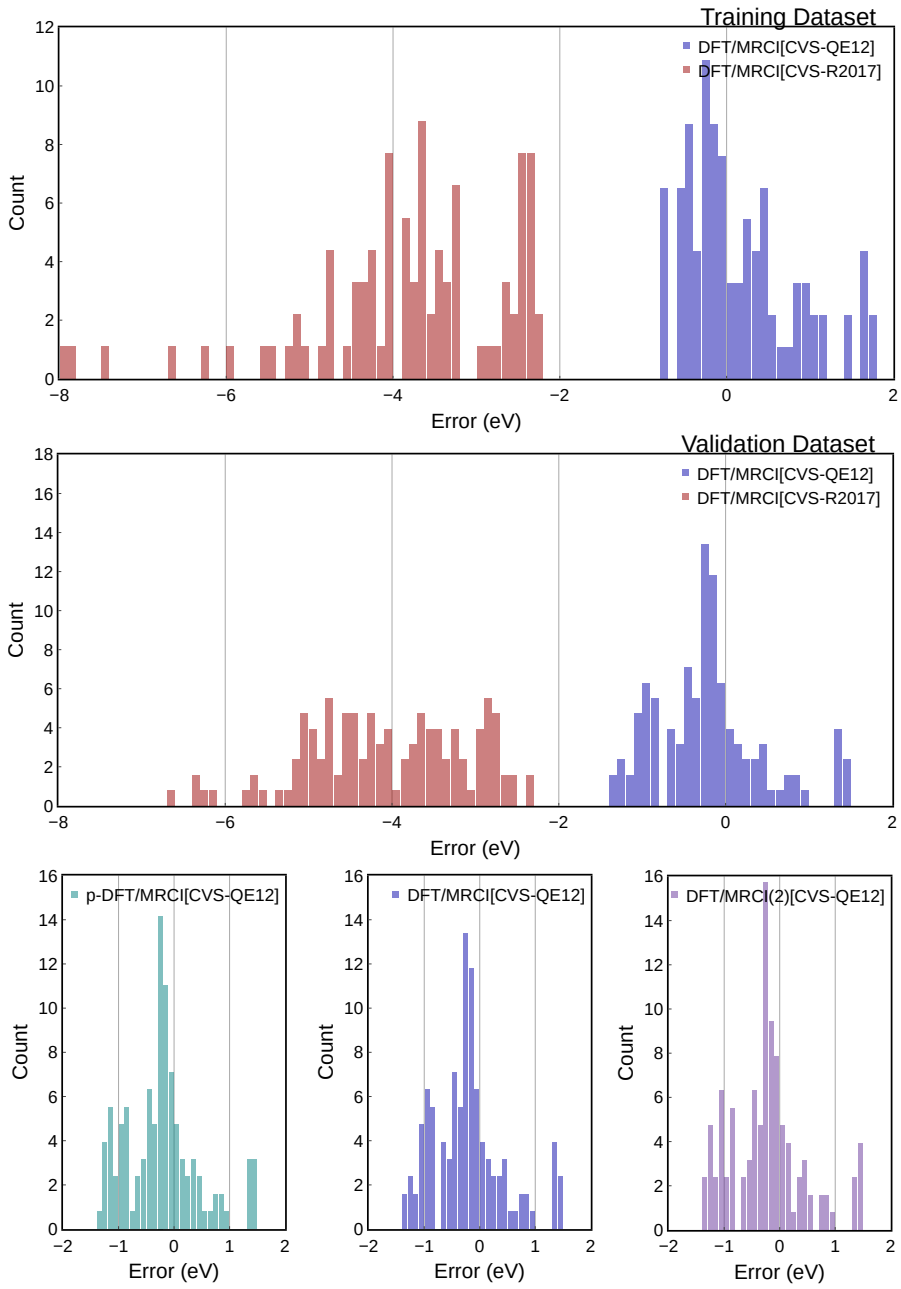


Figure 3.3: Error distribution histograms for the CVS-R2017 (red) and the CVS-QE12 Hamiltonians (blue) with respect to the training (top), and the validation (middle) data sets. Error distribution histogram of the CVS-QE12 Hamiltonian with respect to the validation data set (bottom) for the *p*-DFT/MRCI (teal), DFT/MRCI (blue), and DFT/MRCI(2) (purple) variants.

The MAEs for the *p*-DFT/MRCI and DFT/MRCI(2) variants are given in Table 3.2 for both the training and validation data sets. 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 3.3 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 errors 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 3.2: 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 twofold. 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-QE12 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 0.01 eV for all excitation classes in the validation set.

3.5 Representative Applications

3.5.1 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 XPS of ethyl trifluoroacetate, also known as the ESCA molecule.[241] This is a benchmark spectrum, with the four carbon 1s 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 1s 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. It should be noted that ethyl trifluoroacetate exists in two major conformations: *anti-anti* and *anti-gauche*. The difference in the binding C 1s energies of either conformer are small ($\sim$ 0.1 eV) and do not significantly change the resultant spectrum.[227] Thus, only the *anti-anti* geometry is considered here.

The carbon 1s XPS was simulated using both the CVS-QE12 and CVS-R2017 Hamiltonians and aug-cc-pCVTZ basis. The simulated spectra are shown in Figure 3.4 alongside experimental gas-phase spectra from references [226] and [227]. Two experimental references are included herein due to slight differences between either regarding the third and fourth peaks in the XPS. The stick spectra, determined from vertical ionization energies, are convoluted with a Lorentzian function with a full width at 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 spec-

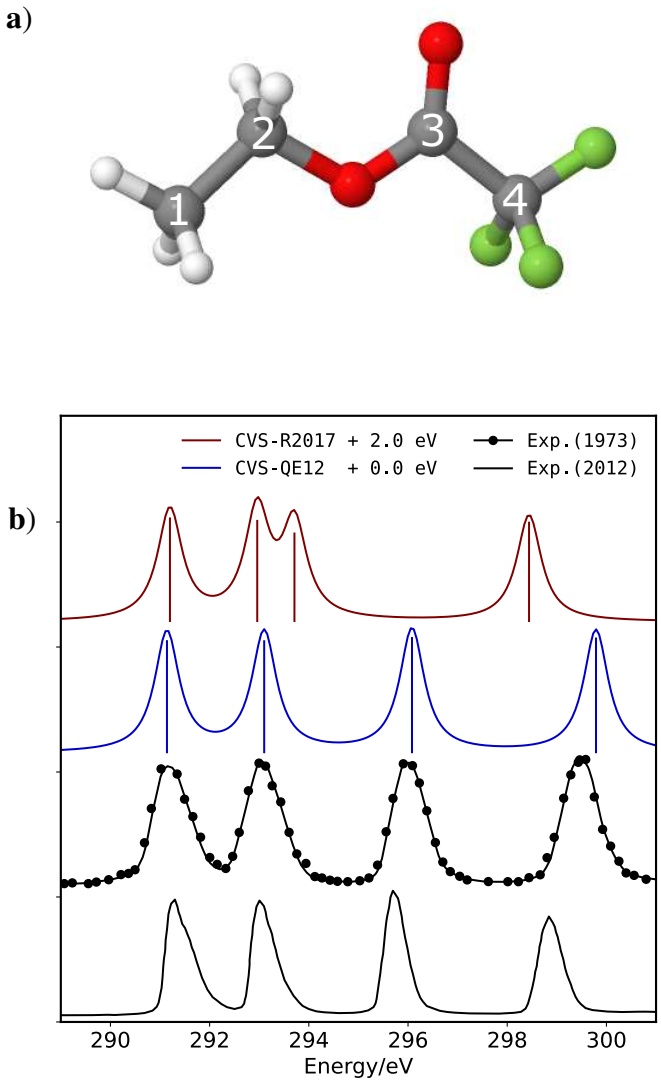


Figure 3.4: Ethyl trifluoroacetate molecule (**a**) and carbon XPS of ethyl trifluoroacetate (**b**) simulated with CVS-DFT/MRCI, using the CVS-R2017 Hamiltonian (red) and the CVS-QE12 Hamiltonian (blue), with comparison to experimental gas-phase X-Ray photoelectron spectra (black). The simulated spectra consider only the anti-anti-conformer. The carbon atoms in (**a**) are numbered per the ordering of the corresponding peaks in the spectrum. Experimental data (1973) shown with line of best fit are digitized from source.[226] Experimental spectrum (2012) digitized from source.[227]

tra into maximum agreement. Although the relative positions of the first two peaks are well reproduced, the position of the third peak, corresponding to the carbon atom labeled number t in Figure 3.4 (a), is drastically 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. Furthermore, the relative positions of all four peaks are correctly described with respect to ref. [226].

3.5.2 X-RAY ABSORPTION SPECTRA: CYTOSINE

Figure 3.5 shows 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 [228] and [229] were used here. The individual spectra computed for the four tautomer structures were weighted and combined using the appropriate Boltzmann population ratio (BPR) at 450K, as computed by Trygubenko et al.[228, 242] 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 3.5.

The spectra computed with CVS-QE12 Hamiltonian notably require smaller shifts to be aligned with the experimental data than do those simulated with ADC(2)-x and EOM-CCSD; the shifts applied to the DFT/MRCI spectra range from -0.8 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.

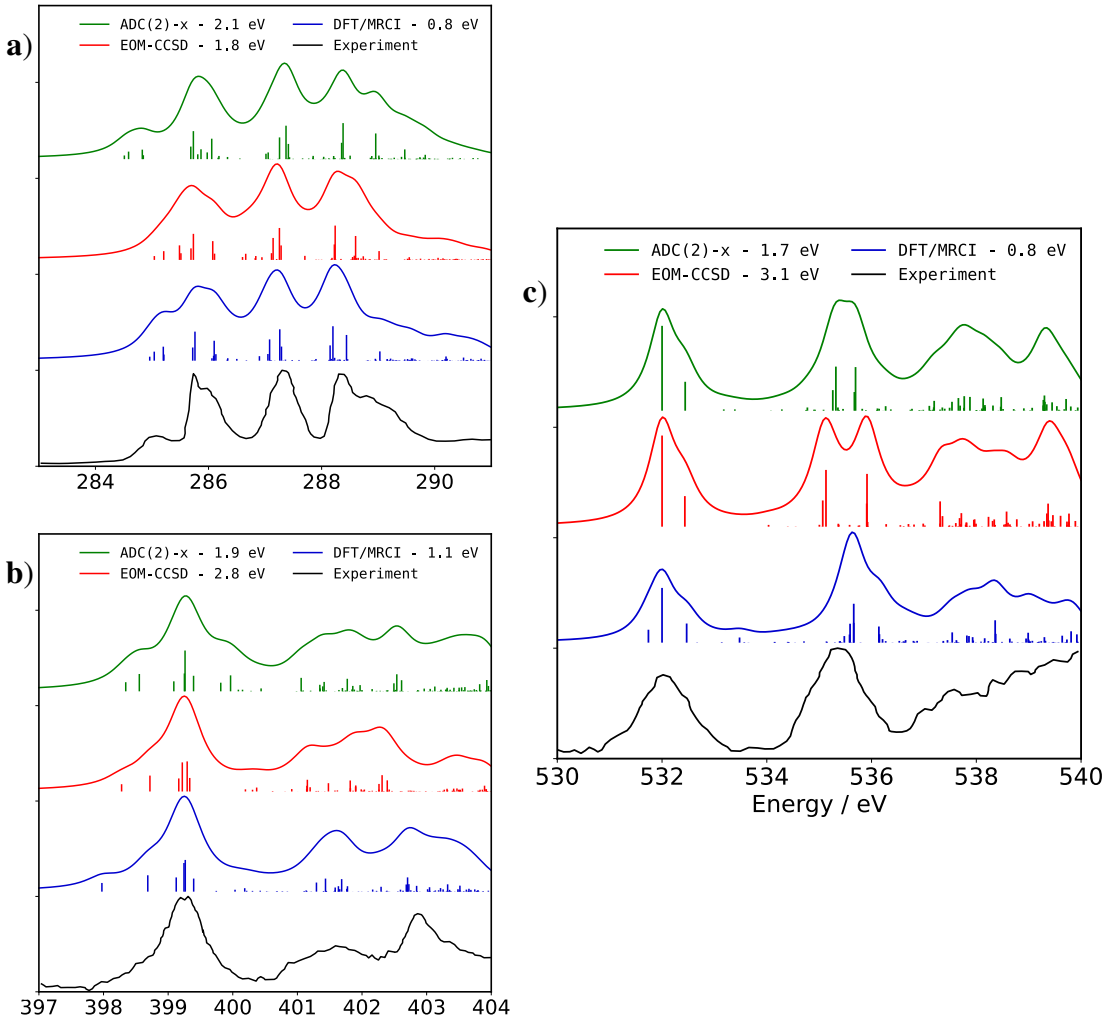


Figure 3.5: NEXAFS Spectrum of carbon (**a**), nitrogen (**b**), and oxygen (**c**) K-shells of cytosine simulated with CVS-ADC(2)-x (green), CVS-EOM-CCSD (orange), and CVS-DFT/MRCI (CVS-QE12) (blue). Experimental spectra (black) digitized from source.[228, 229]

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 in Figure 3.5 (c) 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 in the ground-state structure of cytosine. This state appears to exhibit significant doubly-excited character. The peak at 533.7 eV corresponds to an analogous transition from the carbonyl oxygen in the tautomer, labeled structure "3" in ref. [228] (see Fig. B.7 in App. B).

3.5.3 EXCITED-STATE SPECTRA: THYMINE

One application that motivates the present work is the simulation of time-resolved pump-probe spectroscopies. This application considers 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, the previously studied[3] photoprotective mechanism of the nucleobase thymine is considered. 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.[3] 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.[4–6] 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.[4]

Shown in Figure 3.6(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 ref. [4] are shown alongside in Figure 3.6(b). 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,[4] and this scaling factor was applied to excited state oscillator strength to obtain the DFT/MRCI results in Figure 3.6(a).

First, it should be noted that the agreement between the experimental result and the NEXAFS simulations strongly supports the original assignment made in ref. [4]. Specifically, that the time-delayed feature at 526.5 eV in Figure 3.6(b) is attributable to population of the $n\pi^*$ state. This is clearly seen in Figure 3.6(a) from the NEXAFS spectrum computed at the minimum energy

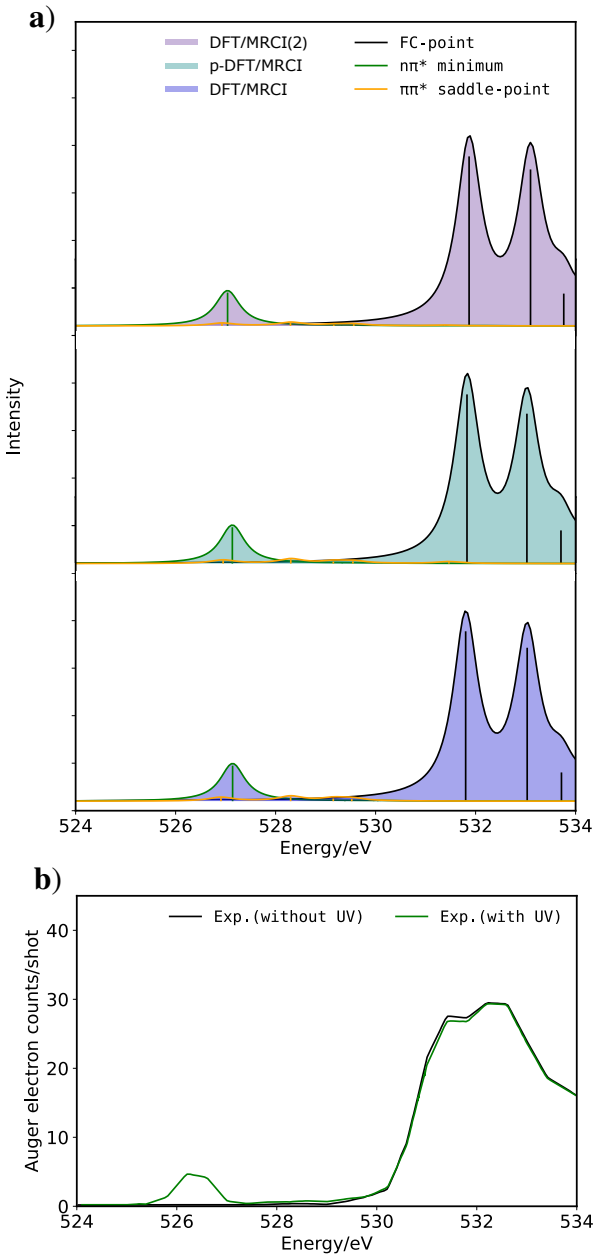


Figure 3.6: Simulated oxygen 1s NEXAFS spectrum of thymine (**a**) and experimental oxygen 1s Auger Spectrum (**b**). The simulated spectra (unshifted) are computed with DFT/MRCI(2) (purple), *p*-DFT/MRCI (teal), and DFT/MRCI (blue) using CVS-QE12 at the Franck–Condon point geometry (ground-state spectrum 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.[4]

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 3.6(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. Furthermore, 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.

3.5.4 Open-Shell Systems: Neutral and Charged Radicals

Recent developments in the parameterization of DFT/MRCI Hamiltonians have enabled a unified and consistent description of electronic excitation between different electronic spin states.[43, 44] Consistent with these previous parameterizations, the CVS-QE12 Hamiltonian employs an underlying KS orbital basis that is constructed using the restricted (RKS) formulation for molecules with an even number of electrons and restricted open-shell (ROKS) formulation for molecules with an odd number of electrons. The training and validation sets employed here included core-ionization energies, which involved the computation of ground-state doublet cationic electronic states. However, examined here is the ability of CVS-QE12 parameterization to describe core excitation energies of neutral and charged radicals, an excitation class not explicitly included in the training data. The motivation for examining the performance of the CVS-QE12 Hamiltonian

for this excitation class is to demonstrate its utility for core excitation in states that exhibit multireference character. A test set similar to that constructed by Hait et al.[230] is employed here.

Table 3.3: 1s $\rightarrow$ SOMO excitation energies of neutral and charged radicals computed with aug-cc-pCVTZ/DFT/MRCI [CVS-QE12] and taken from experimental references. (energies in eV)

1s $\rightarrow$ SOMO Energies from ref. [230]				
Molecule	Atom	Exp.	DFT/MRCI	ΔE
CH_3	C	281.4	281.03	-0.37
CH_3CH_2	C	281.7	282.67	-0.03
i-propyl	C	282.2	282.18	-0.02
t-butyl	C	282.6	282.60	0.00
Allyl	C	282.0	282.07	0.07
CO^+	C	282.0	281.96	-0.04
	O	528.5	527.01	-1.49
CH_2Br	C	282.6	282.38	-0.22
CH_2Cl	C	282.8	282.60	-0.20
NH_2	N	394.3	394.55	0.25
N_2^+	N	394.3	394.29	-0.01
NH_3^+	N	395.2	395.49	0.29
OH	O	525.8	526.19	0.39
HO_2	O	528.6	528.78	0.18
1s $\rightarrow$ SOMO Energies from ref. [243] & [244]				
Molecule	Atom	Exp.	DFT/MRCI	ΔE
Bz^+	C	281.3	281.31	0.01
O_2^-	O	529.0	529.16	0.16

The CVS-QE12 Hamiltonian, like all DFT/MRCI parameterizations, is constructed via the determination of difference configurations relative to an "anchor" or "base" configuration.[29, 30, 107] This formulation is problematic for degenerate ground states which are inherently multiconfigurational. For this reason, NO_2 and O_2 (which are included in ref. [230]) are excluded

from the present analysis. To aid in diversifying the test set, some additional examples are included: the 1s to SOMO excitation energies of the cationic benzene radical (Bz^+) and anionic superoxide radical (O_2^-), from experimental references [243] and [244], respectively. The reference data and results obtained in this work are tabulated in Table 3.3.

The CVS-QE12 Hamiltonian yields an MAE of 0.25 eV for the subset of data reproduced from Hait et al.,[230] and an MAE of 0.23 eV over the entire reference set in Table 3.3. Likewise, the mean error and RMSD across the subset are -0.09 and 0.45 eV. Across the whole data set, these errors are determined as -0.06 and 0.42 eV. The oxygen K-shell excitation energy of CO^+ is an outlier, with an error of -1.5 eV. Interestingly, the O 1s $\rightarrow \pi^*$ has an experimental excitation energy of 533.44 eV,[245] which is exactly reproduced by CVS-QE12. Likewise, the experimental C 1s $\rightarrow \pi^*$ has an associated excitation energy of 289.85 eV,[245] and that predicted by CVS-QE12 is 289.84 eV. The magnitude of the error of the 1s $\rightarrow$ SOMO excitation, being so state-specific in CO^+, warrants further investigation in future work. Overall, the results presented in Table 3.3 demonstrate that the good performance of the CVS-QE12 Hamiltonian for organic photochemistry is consistent for open-shell references.

In theme with the previous applications, the carbon X-ray absorption spectrum of the allyl free radical is simulated here, and compared to the experimental result. Hait et al. include this spectrum in their study on neutral and cationic radicals.[230] They account for the inherently multi-configurational character of the molecules within their test set using state-specific OO-DFT with a self-consistent recoupling scheme, demonstrating that their approach yields an accurate description of the allyl XAS. Only one outstanding detail is reported: the terminal carbon 1s to SOMO excitation energy is consistently over-estimated by 0.5 eV. Notably, the accuracy of this method is not degraded in the high-energy region of the spectrum. By comparison, using

standard linear response methods TDDFT and fc-CVS-EOM-CCSD reveals an inadequate overall description of this spectrum, particularly for the higher-energy transitions. The latter method nevertheless recovers an accurate peak position for the 1s to SOMO transition, likely due to the inclusion of doubles. In Figure 3.7, the CVS-QE12 Hamiltonian is used to compute the core-

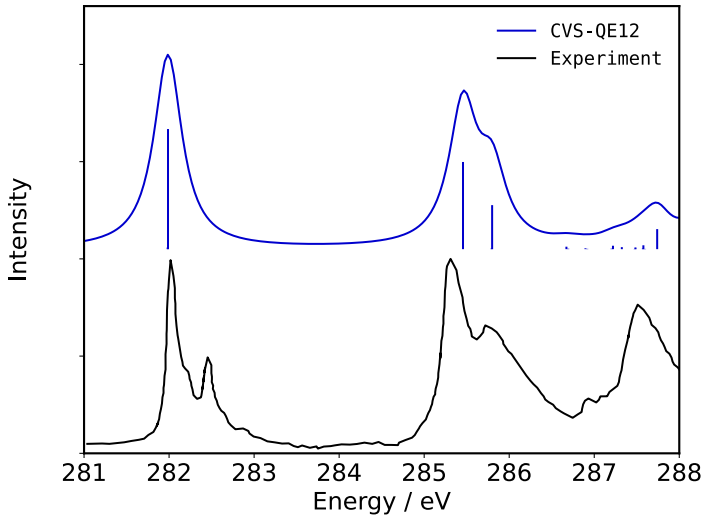


Figure 3.7: Carbon 1s NEXAFS spectrum of the allyl free radical. The simulated spectrum (blue) is computed using DFT/MRCI [CVS-QE12] with the aug-cc-pCVTZ basis set. The experimental spectrum (black) is digitized from its source publication.[231]

excited states of allyl with the aug-cc-pCVTZ basis set (and aug-cc-pVTZ basis on H atoms). As in the previous applications, the spectral peaks are artificially broadened using a Lorentzian convolution; here, a FWHM of 0.40 eV is applied, and no empirical shift is used to align the spectra. The simulated XAS of allyl shows good agreement with the experiment, exclusive of the vibrational structure evinced in the origin band. A complete description of the NEXAFS spectrum would require the inclusion of vibronic coupling effects, which are not presently considered. As Figure 3.7 shows, CVS-QE12 accurately predicts the 1s to SOMO transition peak at 282.0 eV,

and the 1s to LUMO transition peaks between 285 and 286 eV. However, there appears to be a pronounced underestimation of the oscillator strengths associated with a large peak between 287 and 288 eV. While this detail is important, the analysis of physical properties derived from DFT/MRCI computations is the subject of a future study.

In summary, CVS-QE12 parameterization performs quite well for the preponderance of radical excitation energies, as well as the simulation of the previously discussed problematic core excitation spectrum of allyl radical. The ability of the method to readily describe multireference electronic character in both the valence- and core-excited manifolds is one of the primary drivers for the development of the approach. This property of the approach makes it an attractive option for the simulation of, for example, time-resolved spectra involving X-ray pulses.[109, 246]

3.5.5 Computational Efficiency: Timings

To conclude, computational timings are presented for the representative applications of this section. Table 3.4 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,[125] which uses the PySCF[126–130] library to solve the KS-DFT equations. However, other packages may be employed in this respect, the computational efficiency of which will vary.

Figure 3.6, along with figures B.6, B.8, and B.9 in Appendix B, 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. This computational efficiency is

Table 3.4: CPU times (in seconds) of the DFT/MRCI, *p*-DFT/MRCI, and DFT/MRCI(2) calculations for the representative applications.[a]

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
7017	292	131	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
Allyl				
DFT/MRCI	*p*-DFT/MRCI	DFT/MRCI(2)	Atom	*N* Roots
141	106	106	C	50

[a] 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, thymine, and allyl, and the aug-cc-pCVDZ basis was used for cytosine.

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 1s 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 3.5. The DFT/MRCI approach, being a selected CI method, has a certain advantage for the calculation of large numbers of states. 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.

3.6 CONCLUSION

This chapter reports 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 Chapter 2, the QTP17 XC 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, this work shares with Chapter 2 a core design philosophy in which the fitting of the Hamiltonian parameters is performed 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[247, 248] 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.

4

CONCLUDING REMARKS

This thesis has presented two new parameterizations of the DFT/MRCI Hamiltonian for the computation of valence-excited states, and K-edge core-excited states. These parameterizations are found to yield sub-eV error metrics for the VEEs of organic molecules. Chapter 2 presents the QE8 Hamiltonian, which is found to yield an MAE of 0.16 eV with respect to the theoretical best estimates of vertical excitation energies. Chapter 3 likewise presents the CVS-QE12 Hamiltonian, which is found to yield a median 0.54 eV MAE across the benchmark VEEs computed for core-excited states. Chapters 2 and 3 additionally report computational timings of the post-DFT

calculation component for all examples used in the representative applications of either chapter. These results illustrate that the accuracy reported for the parent DFT/MRCI method is retained for the approximative variants *p*-DFT/MRCI, and DFT/MRCI(2), while computational timings are reduced by an order of magnitude. The chosen applications furthermore highlight that the optical properties and relative peak positions show good agreement with comparable electronic structure methods, and with experimental data.

In semi-empirical methodology, it is imperative to link elements of the underlying mathematics to the physical properties of the matter they describe. To this end, Chapter 1 focuses on technical aspects of SIE and the high-density limit in DFT as they pertain to KS orbital energies, and therefore, to excited-state energies. Preliminary results indicated that refitting the parameters of the DFT/MRCI Hamiltonian, as an exclusive measure, was insufficient to achieve comparable error metrics between valence and core-excited states. Thus, an evaluation of the XC functional in the DFT component led to the selection of the QTP17 functional, resulting in more balanced errors between the valence and core-excited states. Nevertheless, it was found to be necessary to introduce separate, but related parameterizations for either case. This was primarily achieved by partitioning the Coulomb scaling parameter. However, the damping function was additionally subject to a separate optimization for core-excited states; this latter choice was partly motivated by the fact that core-excited states exhibit greater orbital relaxation effects in response to the strongly attractive core hole. While this combined effort has furnished impressive results for either class of excitations, it remains an open question whether the innovation of a single, *unified* description of the DFT/MRCI Hamiltonian is feasible.

Chapter 3 of the present work focuses on the K-edge excitation energies for second-row elements in organic molecules. Nevertheless, X-ray spectroscopy has a broad applicability be-

yond second-row elements, and indeed beyond the K-edge. In particular, L-edge spectra, and the spectra of inorganic molecules such as transition metal complexes, are of interest. Beyond the K-edge, however, the description of electronic states becomes more complicated, and spin-orbit coupling must be taken into account. Furthermore, a more flexible approach to generating the underlying KS orbital basis may become necessary. One possible avenue is to explore the use of core-projected hybrid functionals, as applied in TDDFT.[247, 248] In this approach, specific fractions of exact exchange can be projected onto the core orbitals, resulting in accurate core excitation and ionization energies without degradation of the properties related to the valence electron density. Another recent work, using the DFT/CIS method, employs a modified on-diagonal correction scheme for core excitation energies.[249] Exploring the application of different correction schemes in DFT/MRCI will be the subject of future work.

A main motivation in the present work was to parameterize the DFT/MRCI Hamiltonian to reproduce highly accurate ab initio vertical excitation energies, made possible in large part due to the availability of benchmarking data in the QUEST databases.[28] Using a semi-empirical method to reproduce the ab initio Hamiltonian is made difficult when fitting to experimental excitation energies, as these are *vibronic*, rather than purely electronic, in nature. While the properties associated with electronic excitation were not explicitly evaluated, the absorption spectra simulated as a part of the representative applications indicate that the DFT/MRCI oscillator strengths are in good overall agreement with experimental spectra, and with spectra simulated by comparable electronic structure methods. Future work can take advantage of new benchmarking data for oscillator strengths.[27, 161] In addition, a new database of benchmark double excitation energies has become available since the present work was completed,[250] expanding upon the present test set. However, future work would still benefit from a more diverse set of

benchmarking data for core-excited states.

The ability of the DFT/MRCI Hamiltonian to reproduce ab initio VEEs, and therefore the ab initio Hamiltonian, is of interest for its application beyond the Born–Oppenheimer approximation. In cases where this approximation breaks down, or else where the coupling of electronic and nuclear motion is explicitly considered, it is necessary to evaluate energy gradients and non-adiabatic couplings. Previous work has leveraged a propagative variant of the block diagonalization diabatization (BDD)[251–253] scheme (i.e., P-BDD) together with DFT/MRCI[159] to construct the vibronic coupling Hamiltonian using wave function overlaps. Furthermore, the DFT/MRCI(2) method[46] has been extended to the calculation of quasi-diabatic states using an alternate approach, termed the QD-DFT/MRCI(2) method.[254] Being more efficient yet, this latter approach is tractable not only to the construction of the vibronic coupling Hamiltonian, but also to the simulation of on-the-fly dynamics. At first glance, describing dynamics is problematic for a method based on selected CI, as this requires the description of potential energy surfaces. In a selected CI *ansatz*, the reference space is regenerated at each nuclear configuration, so the underlying potential is not smooth. However, machine learning approaches to describe potential energy surfaces can be applied to DFT/MRCI to overcome this issue.[165, 166] The ability to describe surfaces lends itself to structure optimization and minimum energy conical intersection optimization,[166] as well as to dynamics.

As the simulation of photochemical processes can imply high-dimensional coupled potential energy surfaces, methods that are both computationally efficient and predicatively accurate are indispensable. This is especially relevant given the barriers that have been overcome in recent decades, launching the field of femtochemistry[255] and subsequently the field of attochemistry.[256] The work presented herein, and related developments[45, 46, 48, 159, 165, 166, 246,

254] within the same research group, have a shared vision. Supported by the present work, the DFT/MRCI(2) method opens the door to studying molecular systems that once posed a prohibitively expensive computational task.

APPENDICES

A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 Detail on the Training and Validation Data Sets

The QUEST databases[20–25, 28] of ab initio excitation energies were used as a reference for both the training and validation sets in the present parameterization of the DFT/MRCI Hamiltonian, QE8. This data is comprised primarily of full configuration interaction (FCI) and near-FCI results, or else the theoretical best estimate (TBE), which is otherwise specified for each of the databases consulted. Each database in QUEST provides exemplar transition energies for a specified type of chemical environment, making up overall a robust benchmarking database for any electronic structure method.

Table A.1: Summary of the contents of training and validation sets of ab initio reference data from the QUEST Databases.

Database	Theme	Level of Theory
QUEST 1	Small Molecules	FCI/aDZ
QUEST 2	Double Excitation Transitions	FCI/aDZ, FCI/6-31+G(d)
QUEST 3	Medium Molecules	FCI/6-31+G(d), TBE/aTZ[a]
QUEST 4	F, P, Cl, Si-containing Molecules, Radicals	FCI/aDZ, TBE/aTZ[b]
QUEST 5	Large Molecules	TBE/aTZ[c]
QUEST 6	Charge Transfer (CT) Transitions	TBE/aTZ[d]

[a] The TBE (QUEST 3) is at least CCSDT or CCSDTQ; SCI is used when possible.[22]
[b] The TBE (QUEST 4) is obtained using SCI via the CIPSI algorithm.[23]
[c] The TBE (QUEST 5) is obtained using either FCI, or by using an *ONIOM*-type approach with CC theory (up to CCSDTQ)[24]
[d] The TBE (QUEST 6) is obtained by an additive CC scheme, using CCSDT alongside CC3 and CCSD corrections.[25]

Each QUEST database concerns a specific type of molecular system or a specific class of transitions; this general “theme" is summarized for each subset in Table A.1. The data in each

database are split between the training and validation sets, except for QUEST 2, which concerns double excitations. QUEST 2 comprises a smaller set of excitations and only makes up a part of the fit set. The level of theory used differs slightly between the databases, and necessarily so; larger molecules are considered in databases 3 through 6, which results in an insurmountably large computational expense should FCI be applied in each case. For each corresponding DFT/MRCI computation, the basis set is selected to match that of the reference. The level of theory used for the reference data used in this work is specified in column 3 of Table A.1. The TBE is defined differently among the databases and is noted accordingly in Table A.1.

Tables A.2 and A.3 tabulate the results of the fitted QE8 Hamiltonian for the training and validation set reference data, respectively. Two sets of parameters are canonically reported for the DFT/MRCI Hamiltonian, corresponding to an $\delta E_{sel} = 0.8\ E_h$, and $1.0\ E_h$, respectively. Furthermore, the present work reports on the performance of the *p*-DFT/MRCI and DFT/MRCI(2) methods[45, 46] with respect to standard DFT/MRCI. The results of the $\delta E_{sel} = 1.0\ E_h$ Hartree parameter set using standard DFT/MRCI are reported below for QE8.

Table A.2: Training set data from the QUEST Databases and DFT/MRCI results. Energy in eV.

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Acetylene					
	$1\,^1\Sigma_g$	$1\,^1\Sigma_u^-$	7.20	6.75	-0.45
	$1\,^1\Sigma_g$	$1\,^1\Delta_u$	7.51	7.00	-0.51
	$1\,^1\Sigma_g$	$1\,^1\Delta_u$	7.51	7.02	-0.49
	$1\,^1\Sigma_g$	$1\,^3\Sigma_u^+$	5.50	5.54	0.04
	$1\,^1\Sigma_g$	$1\,^3\Delta_u$	6.46	6.10	-0.36
	$1\,^1\Sigma_g$	$1\,^3\Delta_u$	6.46	6.21	-0.25
	$1\,^1\Sigma_g$	$1\,^3\Sigma_u^-$	7.14	6.73	-0.41
Ammonia					
	$1\,^1A_1$	$1\,^1A_2$	6.48	6.70	0.22
	$1\,^1A_1$	$1\,^1E$	8.08	8.25	0.17
	$1\,^1A_1$	$2\,^1A_2$	9.68	9.86	0.18
	$1\,^1A_1$	$1\,^3A_2$	6.19	6.37	0.18
Carbon Monoxide					
	$1\,^1\Sigma^+$	$1\,^1\Pi$	8.57	8.40	-0.17
	$1\,^1\Sigma^+$	$1\,^3\Pi$	6.29	6.23	-0.06
	$1\,^1\Sigma^+$	$1\,^3\Sigma^+$	8.46	8.22	-0.24
Dinitrogen					
	$1\,^1\Sigma_g$	$1\,^1\Pi_g$	9.41	9.23	-0.18
	$1\,^1\Sigma_g$	$1\,^3\Sigma_u^+$	7.70	7.70	-0.00
	$1\,^1\Sigma_g$	$1\,^3\Pi_g$	8.05	7.98	-0.07
Ethylene					
	$1\,^1A_g$	$1\,^1B_{3g}$	7.31	7.24	-0.07
	$1\,^1A_g$	$1\,^1B_{1u}$	7.93	7.94	0.01
	$1\,^1A_g$	$1\,^1B_{1g}$	8.00	7.89	-0.11
	$1\,^1A_g$	$1\,^3B_{1u}$	4.55	4.54	-0.01
	$1\,^1A_g$	$1\,^3B_{3u}$	7.16	7.12	-0.04
	$1\,^1A_g$	$1\,^3B_{1g}$	7.93	7.83	-0.1
Formaldehyde					
	$1\,^1A_1$	$1\,^1A_2$	3.99	3.87	-0.12
	$1\,^1A_1$	$1\,^1B_2$	7.11	7.15	0.04
	$1\,^1A_1$	$2\,^1B_2$	8.04	8.23	0.19
	$1\,^1A_1$	$2\,^1A_1$	8.12	8.18	0.06
	$1\,^1A_1$	$2\,^1A_2$	8.65	9.03	0.38
	$1\,^1A_1$	$1\,^1B_1$	9.29	9.14	-0.15
	$1\,^1A_1$	$3\,^1A_1$	9.53	9.37	-0.16

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
	$1\,^1A_1$	$3\,^3A_1$	3.58	3.50	-0.08
	$1\,^1A_1$	$1\,^3A_2$	6.10	5.85	-0.25
	$1\,^1A_1$	$1\,^3B_2$	6.95	6.95	0.00
	$1\,^1A_1$	$2\,^3B_2$	7.87	8.00	0.13
	$1\,^1A_1$	$2\,^3A_1$	8.01	8.00	-0.01
	$1\,^1A_1$	$1\,^3B_1$	8.48	8.41	-0.07
Hydrogen Sulfide					
	$1\,^1A_1$	$1\,^1A_2$	6.10	6.21	0.11
	$1\,^1A_1$	$1\,^1B_1$	6.29	6.11	-0.18
	$1\,^1A_1$	$1\,^3A_2$	5.75	5.82	0.07
	$1\,^1A_1$	$1\,^3B_1$	5.90	5.75	-0.15
Ketene					
	$1\,^1A_1$	$1\,^1A_2$	3.84	3.89	0.05
	$1\,^1A_1$	$1\,^1B_1$	5.88	6.11	0.23
	$1\,^1A_1$	$2\,^1A_2$	7.08	7.43	0.35
	$1\,^1A_1$	$1\,^3A_2$	3.79	3.69	-0.10
	$1\,^1A_1$	$1\,^3A_1$	5.64	5.45	-0.19
	$1\,^1A_1$	$1\,^3B_1$	5.68	5.90	0.22
	$1\,^1A_1$	$2\,^3A_2$	7.07	7.25	0.18
Water					
	$1\,^1A_1$	$1\,^1B_1$	7.53	7.95	0.42
	$1\,^1A_1$	$1\,^1A_2$	9.32	9.62	0.30
	$1\,^1A_1$	$2\,^1A_1$	9.94	10.22	0.28
	$1\,^1A_1$	$1\,^3B_1$	7.14	7.51	0.37
	$1\,^1A_1$	$1\,^3A_2$	9.14	9.41	0.27
	$1\,^1A_1$	$1\,^3A_1$	9.49	9.67	0.18
Acrolein					
	$1\,^1A'$	$4\,^1A'$	8.00	7.86	-0.14
Benzene					
	$1\,^1A_{1g}$	$1\,^1E_{2g}$	8.40	8.44	0.04
Butadiene					
	$1\,^1A_g$	$2\,^1A_g$	6.51	6.53	0.02
Ethylene					
	$1\,^1A_g$	$2\,^1A_g$	13.07	13.5	0.43
Formaldehyde					
	$1\,^1A_1$	$3\,^1A_1$	10.45	9.76	-0.69

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Glyoxyl					
	$1\,^1A_g$	$2\,^1A_g$	5.48	5.78	0.30
Nitrosomethane					
	$1\,^1A'$	$2\,^1A'$	4.84	4.77	-0.07
Nitroxyl					
	$1\,^1A'$	$2\,^1A'$	4.40	4.36	-0.04
Acetone					
	$1\,^1A_1$	$1\,^1A_2$	4.60	4.47	-0.13
	$1\,^1A_1$	$1\,^3A_2$	4.18	4.16	-0.02
Acrolein					
	$1\,^1A'$	$1\,^1A''$	3.85	3.74	-0.11
	$1\,^1A'$	$2\,^1A'$	6.59	6.72	0.13
	$1\,^1A'$	$1\,^3A''$	3.60	3.47	-0.13
	$1\,^1A'$	$1\,^3A'$	3.98	3.98	-0.00
Butadiene					
	$1\,^1A_g$	$1\,^1B_u$	6.41	6.35	-0.06
	$1\,^1A_g$	$2\,^1A_g$	6.55	6.61	0.06
	$1\,^1A_g$	$2\,^1A_u$	6.95	6.92	-0.03
	$1\,^1A_g$	$1\,^3B_u$	3.37	3.43	0.06
	$1\,^1A_g$	$1\,^3B_g$	6.40	6.52	0.12
Cyanoacetylene					
	$1\,^1\Sigma^+$	$1\,^1\Sigma^-$	6.02	5.64	-0.38
	$1\,^1\Sigma^+$	$1\,^1\Delta$	6.28	5.84	-0.44
	$1\,^1\Sigma^+$	$1\,^3\Sigma^+$	4.45	4.61	0.16
	$1\,^1\Sigma^+$	$1\,^3\Delta$	5.32	5.12	-0.20
Cyclopropenethione					
	$1\,^1A_1$	$1\,^1A_2$	3.44	3.22	-0.22
	$1\,^1A_1$	$1\,^1B_1$	3.45	3.59	0.14
	$1\,^1A_1$	$1\,^1B_2$	4.59	4.54	-0.05
	$1\,^1A_1$	$1\,^3A_2$	3.29	3.40	0.11
	$1\,^1A_1$	$1\,^3B_2$	4.03	3.94	-0.09
Diacetylene					
	$1\,^1\Sigma_g$	$1\,^1\Sigma_u^-$	5.52	5.18	-0.34
	$1\,^1\Sigma_g$	$1\,^1\Delta_u$	5.84	5.38	-0.46
	$1\,^1\Sigma_g$	$1\,^3\Sigma_u^+$	4.04	4.25	0.21
	$1\,^1\Sigma_g$	$1\,^3\Delta_u$	4.94	4.71	-0.23

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Propynal					
	$1\,^1A'$	$1\,^1A''$	3.84	3.82	-0.02
	$1\,^1A'$	$2\,^1A''$	5.64	5.36	-0.28
	$1\,^1A'$	$1\,^3A''$	3.54	3.59	0.05
	$1\,^1A'$	$1\,^3A'$	4.44	4.42	-0.02
Pyrazine					
	$1\,^1A_g$	$1\,^1B_{3u}$	4.15	4.39	0.24
	$1\,^1A_g$	$1\,^1A_u$	4.98	5.36	0.38
	$1\,^1A_g$	$1\,^1B_{2u}$	5.02	5.17	0.15
	$1\,^1A_g$	$1\,^1B_{2g}$	5.71	5.78	0.07
	$1\,^1A_g$	$2\,^1A_g$	6.65	6.90	0.25
	$1\,^1A_g$	$1\,^1B_{1g}$	6.74	6.91	0.17
	$1\,^1A_g$	$1\,^1B_{1u}$	6.88	7.06	0.18
	$1\,^1A_g$	$2\,^1B_{1g}$	7.21	7.26	0.05
	$1\,^1A_g$	$2\,^1B_{2u}$	7.24	7.48	0.24
	$1\,^1A_g$	$2\,^1B_{1u}$	7.44	7.67	0.23
	$1\,^1A_g$	$1\,^3B_{3u}$	3.59	3.89	0.30
	$1\,^1A_g$	$1\,^3B_{1u}$	4.35	4.58	0.23
	$1\,^1A_g$	$1\,^3B_{2u}$	4.39	4.48	0.09
	$1\,^1A_g$	$1\,^3A_u$	4.93	5.25	0.32
	$1\,^1A_g$	$1\,^3B_{2g}$	5.08	5.27	0.19
	$1\,^1A_g$	$2\,^3B_{1u}$	5.28	5.31	0.03
Thioacetone					
	$1\,^1A_1$	$1\,^1A_2$	2.61	2.76	0.15
	$1\,^1A_1$	$1\,^3A_2$	2.36	2.52	0.16
CCl_2					
	$1\,^1A_1$	$1\,^1B_1$	2.68	2.83	0.15
	$1\,^1A_1$	$1\,^1A_2$	4.46	4.41	-0.05
	$1\,^1A_1$	$1\,^3B_1$	1.22	1.52	0.30
	$1\,^1A_1$	$1\,^3A_2$	4.36	4.27	-0.09
CF_2					
	$1\,^1A_1$	$1\,^1B_1$	5.12	5.22	0.10
	$1\,^1A_1$	$1\,^3B_1$	2.71	2.94	0.23
CH_3					
	$1\,^2A_2''$	$1\,^2A_1'$	5.85	5.54	-0.31
	$1\,^2A_2''$	$1\,^2E'$	6.96	6.79	-0.17

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
CH_3					
	$1\,^2A_2''$	$2\,^2E'$	7.18	7.12	-0.06
	$1\,^2A_2''$	$2\,^2A_2''$	7.65	7.50	-0.15
F_2BO					
	$1\,^2B_2$	$1\,^2B_2$	0.73	0.71	-0.02
	$1\,^2B_2$	$1\,^2B_2$	2.80	2.38	-0.42
H_2BO					
	$1\,^2B_2$	$1\,^2B_1$	2.15	2.34	0.19
	$1\,^2B_2$	$1\,^2A_1$	3.49	3.35	-0.14
HCO					
	$1\,^2A'$	$1\,^2A''$	2.09	1.91	-0.18
	$1\,^2A'$	$2\,^2A'$	5.45	5.42	-0.03
HOC					
	$1\,^2A'$	$1\,^2A''$	0.92	0.74	-0.18
Nitromethyl					
	$1\,^2B_1$	$1\,^2B_2$	2.05	2.42	0.37
	$1\,^2B_1$	$1\,^2A_2$	2.38	2.70	0.32
	$1\,^2B_1$	$1\,^2A_1$	2.56	2.88	0.32
	$1\,^2B_1$	$2\,^2B_1$	5.35	5.47	0.12
Silylidene					
	$1\,^1A_1$	$1\,^1A_2$	2.14	2.19	0.05
	$1\,^1A_1$	$1\,^1B_2$	3.79	3.96	0.17
Aza-Naphthalene					
	$1\,^1A_g$	$1\,^1B_{3g}$	3.14	3.31	0.17
	$1\,^1A_g$	$1\,^1B_{2u}$	4.28	4.37	0.09
	$1\,^1A_g$	$1\,^1B_{1u}$	4.34	4.38	0.04
	$1\,^1A_g$	$1\,^1B_{2g}$	4.55	4.78	0.23
	$1\,^1A_g$	$2\,^1B_{2g}$	4.89	5.06	0.17
	$1\,^1A_g$	$1\,^1A_u$	5.34	5.28	-0.06
Benzoquinone					
	$1\,^1A_g$	$1\,^1B_{1g}$	2.82	2.83	0.01
	$1\,^1A_g$	$1\,^1A_u$	2.96	2.97	0.01
	$1\,^1A_g$	$1\,^3B_{1g}$	2.58	2.59	0.01
	$1\,^1A_g$	$1\,^3A_u$	2.72	2.76	0.04
	$1\,^1A_g$	$1\,^3B_{1u}$	3.12	3.04	-0.08
	$1\,^1A_g$	$1\,^3B_{3g}$	3.46	3.44	-0.02

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Hexatriene					
	$1\,^1A_g$	$1\,^1B_u$	5.37	5.26	-0.11
	$1\,^1A_g$	$1\,^1A_u$	5.79	5.67	-0.12
	$1\,^1A_g$	$1\,^1B_g$	5.94	5.78	-0.16
	$1\,^1A_g$	$1\,^3B_u$	2.73	2.74	0.01
	$1\,^1A_g$	$1\,^3A_g$	4.36	4.25	-0.11
Naphthalene					
	$1\,^1A_g$	$1\,^1B_{3u}$	4.27	4.38	0.11
	$1\,^1A_g$	$1\,^1B_{2u}$	4.90	4.94	0.04
	$1\,^1A_g$	$1\,^1A_u$	5.65	5.60	-0.05
	$1\,^1A_g$	$1\,^1B_{1g}$	5.84	5.78	-0.06
	$1\,^1A_g$	$1\,^1B_{3g}$	6.07	6.00	-0.07
	$1\,^1A_g$	$1\,^1B_{2g}$	6.09	5.95	-0.14
Octatetraene					
	$1\,^1A_g$	$1\,^1B_u$	4.78	4.60	-0.18
Aniline					
	$1\,^1A_1$	$2\,^1A_1$	5.50	5.69	0.19
Azulene					
	$1\,^1A_1$	$2\,^1A_1$	3.85	3.96	0.11
	$1\,^1A_1$	$2\,^1B_2$	4.50	4.61	0.11
DMABN					
	$1\,^1A_1$	$2\,^1A_1$	4.86	4.99	0.13

Table A.3: Validation set data from the QUEST Databases and DFT/MRCI results. Energy in eV.

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Acetaldehyde					
	$1\,^1A'$	$1\,^1A''$	4.34	4.34	0.00
	$1\,^1A'$	$1\,^3A''$	3.98	4.00	0.02
Cyclopropene					
	$1\,^1A_1$	$1\,^1B_1$	6.70	6.76	0.06
	$1\,^1A_1$	$1\,^1B_2$	6.82	6.78	-0.04
	$1\,^1A_1$	$1\,^3B_2$	4.35	4.27	-0.08
	$1\,^1A_1$	$1\,^3B_1$	6.43	6.46	0.03
Diazomethane					
	$1\,^1A_1$	$1\,^1A_2$	3.09	2.95	-0.14
	$1\,^1A_1$	$1\,^1B_1$	5.35	5.67	0.32
	$1\,^1A_1$	$2\,^1A_1$	5.79	5.96	0.17
	$1\,^1A_1$	$1\,^3A_2$	2.81	2.68	-0.13
	$1\,^1A_1$	$1\,^3A_1$	4.03	3.95	-0.08
	$1\,^1A_1$	$1\,^3B_1$	5.18	5.51	0.33
	$1\,^1A_1$	$2\,^3A_1$	6.81	6.70	-0.11
Formamide					
	$1\,^1A'$	$1\,^1A''$	5.70	5.50	-0.20
	$1\,^1A'$	$1\,^3A''$	5.42	5.23	-0.19
	$1\,^1A'$	$1\,^3A'$	5.82	5.73	-0.09
Hydrogen Chloride					
	$1\,^1\Sigma^+$	$1\,^1\Pi$	7.82	7.85	0.03
Methanimine					
	$1\,^1A'$	$1\,^1A''$	5.25	5.11	-0.14
	$1\,^1A'$	$1\,^3A''$	4.63	4.56	-0.07
Nitrosomethane					
	$1\,^1A'$	$1\,^1A''$	1.99	2.08	0.09
	$1\,^1A'$	$3\,^1A'$	6.29	6.45	0.16
	$1\,^1A'$	$1\,^3A''$	1.15	1.33	0.18
	$1\,^1A'$	$1\,^3A'$	5.56	5.26	-0.3
Streptocyanine-C1					
	$1\,^1A_1$	$1\,^1B_2$	7.14	7.44	0.30
	$1\,^1A_1$	$1\,^3B_2$	5.47	5.76	0.29

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Thioformaldehyde					
	$1\,^1A_1$	$1\,^1A_2$	2.26	2.45	0.19
	$1\,^1A_1$	$1\,^1B_2$	5.83	5.89	0.06
	$1\,^1A_1$	$2\,^1A_1$	6.50	6.57	0.07
	$1\,^1A_1$	$1\,^3A_2$	1.97	2.14	0.17
	$1\,^1A_1$	$1\,^3A_1$	3.45	3.60	0.15
	$1\,^1A_1$	$1\,^3B_2$	5.66	5.69	0.03
Cyanogen					
	$1\,^1\Sigma_g^+$	$1\,^1\Sigma_u^-$	6.39	6.10	-0.29
	$1\,^1\Sigma_g^+$	$1\,^1\Delta_u$	6.66	6.31	-0.35
	$1\,^1\Sigma_g^+$	$1\,^3\Sigma_u^+$	4.91	4.98	0.07
Cyclopentadiene					
	$1\,^1A_1$	$1\,^1B_2$	5.54	5.69	0.15
	$1\,^1A_1$	$1\,^1A_2$	5.78	5.72	-0.06
	$1\,^1A_1$	$1\,^1B_1$	6.41	6.29	-0.12
	$1\,^1A_1$	$2\,^1A_2$	6.46	6.35	-0.11
	$1\,^1A_1$	$2\,^1B_2$	6.56	6.42	-0.14
	$1\,^1A_1$	$1\,^3B_2$	3.31	3.39	0.08
	$1\,^1A_1$	$1\,^3A_1$	5.11	5.03	-0.08
	$1\,^1A_1$	$1\,^3A_2$	5.73	5.63	-0.1
	$1\,^1A_1$	$1\,^3B_1$	6.36	6.24	-0.12
Cyclopropenone					
	$1\,^1A_1$	$1\,^1B_1$	4.26	4.23	-0.03
	$1\,^1A_1$	$1\,^1A_2$	5.55	5.46	-0.09
	$1\,^1A_1$	$1\,^3B_1$	3.93	3.91	-0.02
	$1\,^1A_1$	$1\,^3B_2$	4.88	4.77	-0.11
Diacetylene					
	$1\,^1\Sigma_g$	$1\,^1\Sigma_u^-$	5.33	5.04	-0.29
	$1\,^1\Sigma_g$	$1\,^1\Delta_u$	5.61	5.23	-0.38
	$1\,^1\Sigma_g$	$1\,^3\Sigma_u^+$	4.10	4.18	0.08
	$1\,^1\Sigma_g$	$1\,^3\Delta_u$	4.78	4.62	-0.16
Furan					
	$1\,^1A_1$	$1\,^1A_2$	6.09	6.08	-0.01
	$1\,^1A_1$	$1\,^1B_2$	6.37	6.36	-0.01
	$1\,^1A_1$	$2\,^1A_1$	6.56	6.42	-0.14
	$1\,^1A_1$	$1\,^1B_1$	6.64	6.60	-0.04

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Furan					
	$1\,^1A_1$	$2\,^1A_2$	6.81	6.80	-0.01
	$1\,^1A_1$	$2\,^1B_2$	7.24	7.29	0.05
	$1\,^1A_1$	$1\,^3B_2$	4.20	4.10	-0.10
	$1\,^1A_1$	$1\,^3A_1$	5.46	5.30	-0.16
	$1\,^1A_1$	$1\,^3A_2$	6.02	6.01	-0.01
	$1\,^1A_1$	$1\,^3B_1$	6.59	6.55	-0.04
Imidazole					
	$1\,^1A'$	$1\,^1A''$	5.71	5.73	0.02
	$1\,^1A'$	$2\,^1A'$	6.41	6.28	-0.13
	$1\,^1A'$	$2\,^1A''$	6.50	6.50	0.00
	$1\,^1A'$	$1\,^3A'$	4.73	4.61	-0.12
	$1\,^1A'$	$1\,^3A''$	5.66	5.68	0.02
	$1\,^1A'$	$2\,^3A'$	5.74	5.56	-0.18
	$1\,^1A'$	$2\,^3A''$	6.31	6.18	-0.13
Isobutene					
	$1\,^1A_1$	$1\,^1B_1$	6.46	6.30	-0.16
	$1\,^1A_1$	$2\,^1A_1$	7.01	6.83	-0.18
	$1\,^1A_1$	$1\,^3A_1$	4.53	4.33	-0.20
Methylenecyclopropene					
	$1\,^1A_1$	$1\,^1B_2$	4.28	4.15	-0.13
	$1\,^1A_1$	$1\,^1A_2$	5.44	5.32	-0.12
	$1\,^1A_1$	$2\,^1A_1$	5.96	5.82	-0.14
	$1\,^1A_1$	$1\,^3B_2$	3.49	3.37	-0.12
	$1\,^1A_1$	$1\,^3A_1$	4.74	4.65	-0.09
Pyrazine					
	$1\,^1A_g$	$1\,^1B_{3u}$	4.15	4.39	0.24
	$1\,^1A_g$	$1\,^1A_u$	4.98	5.36	0.38
	$1\,^1A_g$	$1\,^1B_{2u}$	5.02	5.17	0.15
	$1\,^1A_g$	$1\,^1B_{2g}$	5.71	5.78	0.07
	$1\,^1A_g$	$2\,^1A_g$	6.65	6.90	0.25
	$1\,^1A_g$	$1\,^1B_{1g}$	6.74	6.91	0.17
	$1\,^1A_g$	$1\,^1B_{1u}$	6.88	7.06	0.18
	$1\,^1A_g$	$2\,^1B_{1g}$	7.21	7.26	0.05
	$1\,^1A_g$	$2\,^1B_{2u}$	7.24	7.48	0.24
	$1\,^1A_g$	$2\,^1B_{1u}$	7.44	7.67	0.23

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Pyrazine					
	$1\,^1A_g$	$1\,^3B_{3u}$	3.59	3.89	0.30
	$1\,^1A_g$	$1\,^3B_{1u}$	4.35	4.58	0.23
	$1\,^1A_g$	$1\,^3B_{2u}$	4.39	4.48	0.09
	$1\,^1A_g$	$1\,^3A_u$	4.93	5.25	0.32
	$1\,^1A_g$	$1\,^3B_{2g}$	5.08	5.27	0.19
	$1\,^1A_g$	$2\,^3B_{1u}$	5.28	5.31	0.03
Pyridazine					
	$1\,^1A_1$	$1\,^1B_1$	3.83	4.02	0.19
	$1\,^1A_1$	$1\,^1A_2$	4.37	4.61	0.24
	$1\,^1A_1$	$2\,^1A_1$	5.26	5.41	0.15
	$1\,^1A_1$	$2\,^1A_2$	5.72	5.87	0.15
	$1\,^1A_1$	$1\,^1B_2$	6.17	6.34	0.17
	$1\,^1A_1$	$2\,^1B_1$	6.37	6.63	0.26
	$1\,^1A_1$	$2\,^1B_2$	6.75	6.9	0.15
	$1\,^1A_1$	$1\,^3B_1$	3.19	3.45	0.26
	$1\,^1A_1$	$1\,^3A_2$	4.11	4.36	0.25
	$1\,^1A_1$	$1\,^3A_1$	4.82	4.86	0.04
Pyridine					
	$1\,^1A_1$	$1\,^1B_1$	4.95	5.13	0.18
	$1\,^1A_1$	$1\,^1B_2$	5.14	5.30	0.16
	$1\,^1A_1$	$1\,^1A_2$	5.40	5.71	0.31
	$1\,^1A_1$	$2\,^1A_1$	6.62	6.80	0.18
	$1\,^1A_1$	$3\,^1A_1$	6.76	6.91	0.15
	$1\,^1A_1$	$2\,^1A_2$	6.82	6.79	-0.03
	$1\,^1A_1$	$2\,^1B_1$	7.38	7.36	-0.02
	$1\,^1A_1$	$4\,^1A_1$	7.39	7.54	0.15
	$1\,^1A_1$	$1\,^3A_1$	4.30	4.51	0.21
	$1\,^1A_1$	$1\,^3B_1$	4.46	4.69	0.23
	$1\,^1A_1$	$1\,^3B_2$	4.79	4.84	0.05
	$1\,^1A_1$	$2\,^3A_1$	5.04	5.09	0.05
	$1\,^1A_1$	$1\,^3A_2$	5.36	5.65	0.29
	$1\,^1A_1$	$2\,^3B_2$	6.24	6.35	0.11
Pyrimidine					
	$1\,^1A_1$	$1\,^1B_1$	4.44	4.72	0.28
	$1\,^1A_1$	$1\,^1A_2$	4.85	5.19	0.34

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Pyrimidine					
	$1\,^1A_1$	$1\,^1B_2$	5.38	5.55	0.17
	$1\,^1A_1$	$2\,^1A_2$	5.92	6.10	0.18
	$1\,^1A_1$	$2\,^1B_1$	6.26	6.53	0.27
	$1\,^1A_1$	$2\,^1B_2$	6.70	6.93	0.23
	$1\,^1A_1$	$2\,^1A_1$	6.88	7.04	0.16
	$1\,^1A_1$	$1\,^3B_1$	4.09	4.40	0.31
	$1\,^1A_1$	$1\,^3A_2$	4.66	4.96	0.3
	$1\,^1A_1$	$1\,^3B_2$	4.96	5.03	0.07
Pyrrole					
	$1\,^1A_1$	$1\,^1A_2$	5.24	5.26	0.02
	$1\,^1A_1$	$1\,^1B_1$	6.00	5.99	-0.01
	$1\,^1A_1$	$2\,^1A_2$	6.00	6.00	-0.00
	$1\,^1A_1$	$1\,^1B_2$	6.26	6.24	-0.02
	$1\,^1A_1$	$2\,^1A_1$	6.30	6.10	-0.20
	$1\,^1A_1$	$2\,^1B_2$	6.83	6.94	0.11
	$1\,^1A_1$	$1\,^3B_2$	4.51	4.39	-0.12
	$1\,^1A_1$	$1\,^3A_2$	5.21	5.21	0.0
	$1\,^1A_1$	$1\,^3A_1$	5.45	5.29	-0.16
	$1\,^1A_1$	$1\,^3B_1$	5.91	5.92	0.01
Tetrazine					
	$1\,^1A_g$	$1\,^1B_{3u}$	2.47	2.78	0.31
	$1\,^1A_g$	$1\,^1A_u$	3.68	4.06	0.38
	$1\,^1A_g$	$1\,^1B_{1g}$	4.93	5.00	0.07
	$1\,^1A_g$	$1\,^1B_{2u}$	5.21	5.34	0.13
	$1\,^1A_g$	$1\,^1B_{2g}$	5.45	5.54	0.09
	$1\,^1A_g$	$2\,^1A_u$	5.53	5.69	0.16
	$1\,^1A_g$	$2\,^1B_{2g}$	6.12	6.24	0.12
	$1\,^1A_g$	$1\,^3B_{3u}$	1.85	2.22	0.37
	$1\,^1A_g$	$1\,^3A_u$	3.45	3.78	0.33
	$1\,^1A_g$	$1\,^3B_{1g}$	4.20	4.40	0.20
	$1\,^1A_g$	$1\,^3B_{2u}$	4.52	4.67	0.15
	$1\,^1A_g$	$1\,^3B_{2g}$	5.04	4.58	-0.46
	$1\,^1A_g$	$2\,^3A_u$	5.11	5.2	0.09
	$1\,^1A_g$	$1\,^3B_{1u}$	5.42	5.36	-0.06

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Thiophene					
	$1\,^1A_1$	$2\,^1A_1$	5.64	5.71	0.07
	$1\,^1A_1$	$1\,^1B_2$	5.98	5.88	-0.1
	$1\,^1A_1$	$1\,^1A_2$	6.14	6.10	-0.04
	$1\,^1A_1$	$1\,^1B_1$	6.14	6.03	-0.11
	$1\,^1A_1$	$2\,^1A_2$	6.21	6.15	-0.06
	$1\,^1A_1$	$2\,^1B_1$	6.49	6.53	0.04
	$1\,^1A_1$	$2\,^1B_2$	7.29	7.36	0.07
	$1\,^1A_1$	$1\,^3B_2$	3.97	4.02	0.05
	$1\,^1A_1$	$1\,^3A_1$	4.76	4.85	0.09
	$1\,^1A_1$	$1\,^3B_1$	5.93	5.71	-0.22
	$1\,^1A_1$	$1\,^3A_2$	6.08	5.93	-0.15
Triazine					
	$1\,^1A_1^{'}$	$1\,^1A_1^{''}$	4.72	5.08	0.36
	$1\,^1A_1^{'}$	$1\,^1A_2^{''}$	4.75	4.99	0.24
	$1\,^1A_1^{'}$	$1\,^1E^{''}$	4.78	5.14	0.36
	$1\,^1A_1^{'}$	$1\,^1A_2^{'}$	5.75	5.92	0.17
	$1\,^1A_1^{'}$	$2\,^1A_1^{'}$	7.24	7.4	0.16
	$1\,^1A_1^{'}$	$1\,^1E^{'}$	7.32	7.61	0.29
	$1\,^1A_1^{'}$	$2\,^1E^{''}$	7.78	7.84	0.06
	$1\,^1A_1^{'}$	$2\,^1E^{'}$	7.94	8.09	0.15
	$1\,^1A_1^{'}$	$1\,^3A_2^{''}$	4.33	4.60	0.27
	$1\,^1A_1^{'}$	$1\,^3E^{''}$	4.51	4.82	0.31
	$1\,^1A_1^{'}$	$1\,^3A_1^{''}$	4.73	5.12	0.39
	$1\,^1A_1^{'}$	$1\,^3A_1^{'}$	4.85	5.01	0.16
	$1\,^1A_1^{'}$	$1\,^3E^{'}$	5.59	5.65	0.06
	$1\,^1A_1^{'}$	$1\,^3A_2^{'}$	6.62	6.79	0.17
Thiopropynal					
	$1\,^1A^{'}$	$1\,^1A^{''}$	2.03	2.24	0.21
BeF					
	$1\,^2\Sigma^+$	$1\,^2\Pi$	4.14	3.89	-0.25
	$1\,^2\Sigma^+$	$2\,^2\Sigma^+$	6.21	6.06	-0.15
BeH					
	$1\,^2\Sigma^+$	$1\,^2\Pi$	2.49	2.31	-0.18
	$1\,^2\Sigma^+$	$2\,^2\Pi$	6.46	6.20	-0.26

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
CClF					
	$1\,^1A'$	$1\,^1A''$	3.57	3.74	0.17
Carbonyl Fluoride					
	$1\,^1A_1$	$1\,^1A_2$	7.30	7.33	0.03
	$1\,^1A_1$	$1\,^3A_2$	7.08	7.06	-0.02
Formyl Fluoride					
	$1\,^1A'$	$1\,^1A''$	6.00	5.87	-0.13
	$1\,^1A'$	$1\,^3A''$	5.65	5.52	-0.13
HCCl					
	$1\,^1A'$	$1\,^1A''$	2.04	2.14	0.10
HCF					
	$1\,^1A'$	$1\,^1A''$	2.54	2.66	0.12
HCP					
	$1\,^2\Sigma^+$	$1\,^1\Sigma^-$	5.04	5.09	0.05
	$1\,^2\Sigma^+$	$1\,^1\Delta$	5.32	5.35	0.03
	$1\,^2\Sigma^+$	$1\,^3\Sigma^+$	3.49	3.80	0.31
	$1\,^2\Sigma^+$	$1\,^3\Delta$	4.34	4.44	0.10
HPO					
	$1\,^1A'$	$1\,^1A''$	2.46	2.44	-0.02
HPS					
	$1\,^1A'$	$1\,^1A''$	1.60	1.74	0.14
HSiF					
	$1\,^1A'$	$1\,^1A''$	3.06	3.12	0.06
OH					
	$1\,^2\Pi$	$1\,^2\Sigma^+$	4.10	4.13	0.03
	$1\,^2\Pi$	$1\,^2\Sigma^-$	8.02	8.24	0.22
$SiCl_2$					
	$1\,^1A_1$	$1\,^1B_1$	3.95	4.01	0.06
	$1\,^1A_1$	$1\,^3B_1$	2.14	2.67	0.53
Vinyl					
	$1\,^2A''$	$2\,^2A''$	3.26	3.30	0.04
	$1\,^2A''$	$3\,^2A''$	4.69	4.56	-0.13
	$1\,^2A''$	$1\,^2A'$	6.20	5.43	-0.77
Cyclopentadienethione					
	$1\,^1A_1$	$1\,^1A_2$	1.70	1.84	0.14
	$1\,^1A_1$	$1\,^1B_2$	2.63	2.70	0.07

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Cyclopentadienethione					
	$1\,^1A_1$	$2\,^1A_1$	4.96	5.36	0.40
	$1\,^1A_1$	$1\,^3A_2$	1.47	1.60	0.13
	$1\,^1A_1$	$1\,^3B_2$	1.88	1.95	0.07
	$1\,^1A_1$	$1\,^3A_1$	2.51	2.64	0.13
Cyclopentadienone					
	$1\,^1A_1$	$1\,^1A_2$	2.94	3.09	0.15
	$1\,^1A_1$	$1\,^1B_2$	3.58	3.67	0.09
	$1\,^1A_1$	$1\,^3B_2$	2.29	2.38	0.09
	$1\,^1A_1$	$1\,^3A_2$	2.65	2.82	0.17
	$1\,^1A_1$	$1\,^3A_1$	4.19	4.14	-0.05
Diazirine					
	$1\,^1A_1$	$1\,^1B_1$	4.09	4.16	0.07
	$1\,^1A_1$	$1\,^1A_2$	7.27	7.41	0.14
	$1\,^1A_1$	$1\,^1B_2$	7.44	7.50	0.06
	$1\,^1A_1$	$2\,^1A_1$	8.03	8.17	0.14
	$1\,^1A_1$	$1\,^3B_1$	3.49	3.56	0.07
	$1\,^1A_1$	$1\,^3B_2$	5.06	4.84	-0.22
	$1\,^1A_1$	$1\,^3A_2$	6.12	6.39	0.27
	$1\,^1A_1$	$1\,^3A_1$	6.81	6.82	0.01
Maleimide					
	$1\,^1A_1$	$1\,^1B_1$	3.80	3.85	0.05
	$1\,^1A_1$	$1\,^1A_2$	4.52	4.49	-0.03
	$1\,^1A_1$	$1\,^1B_2$	4.89	5.13	0.24
	$1\,^1A_1$	$2\,^1B_2$	6.21	6.16	-0.05
	$1\,^1A_1$	$3\,^1B_2$	7.20	7.42	0.22
	$1\,^1A_1$	$1\,^3B_1$	3.57	3.63	0.06
	$1\,^1A_1$	$1\,^3B_2$	3.74	3.76	0.02
	$1\,^1A_1$	$2\,^3B_2$	4.24	4.30	0.06
	$1\,^1A_1$	$1\,^3A_2$	4.32	4.35	0.03
Nitroxyl					
	$1\,^1A'$	$1\,^1A''$	1.74	1.90	0.16
	$1\,^1A'$	$2\,^1A'$	4.30	4.41	0.11
	$1\,^1A'$	$3\,^1A'$	6.27	6.55	0.28
	$1\,^1A'$	$1\,^3A''$	0.88	1.18	0.30
	$1\,^1A'$	$1\,^3A'$	5.61	5.53	-0.08

Molecule	Initial	Final	Reference	DFT/MRCI [QE8]	ΔE
Streptocyanine-c3					
	$1\,^1A_1$	$1\,^1B_2$	4.82	4.98	0.16
	$1\,^1A_1$	$1\,^3B_2$	3.44	3.60	0.16
Streptocyanine-c5					
	$1\,^1A_1$	$1\,^1B_2$	3.64	3.75	0.11
	$1\,^1A_1$	$1\,^3B_2$	2.47	2.58	0.11
Thioacrolein					
	$1\,^1A'$	$1\,^1A''$	2.11	2.29	0.18
	$1\,^1A'$	$1\,^3A''$	1.91	2.05	0.14
Aminobenzonitrile					
	$1\,^1A_1$	$2\,^1A_1$	5.09	5.15	0.06
Aniline					
	$1\,^1A_1$	$2\,^1A_1$	5.50	5.69	0.19
Benzonitrile					
	$1\,^1A_1$	$1\,^1A_2$	7.05	6.81	-0.24
Benzothiadiazole					
	$1\,^1A_1$	$1\,^1B_2$	4.29	4.25	-0.04
Dimethylaniline					
	$1\,^1A_1$	$1\,^1B_2$	4.39	4.49	0.10
	$1\,^1A_1$	$2\,^1A_1$	5.40	5.58	0.18
N-phenylpyrrole					
	$1\,^1A_1$	$2\,^1B_2$	5.32	5.33	0.01
	$1\,^1A_1$	$3\,^1A_1$	5.85	5.59	-0.26
Nitroaniline					
	$1\,^1A_1$	$2\,^1A_1$	4.40	4.22	-0.18
Nitrodimethylaniline					
	$1\,^1A_1$	$2\,^1A_1$	4.14	3.95	-0.19
Phthalazine					
	$1\,^1A_1$	$1\,^1A_2$	3.91	4.03	0.12
	$1\,^1A_1$	$1\,^1B_1$	4.31	4.44	0.13
Quinoxaline					
	$1\,^1A_1$	$1\,^1B_2$	4.64	4.61	-0.03
	$1\,^1A_1$	$3\,^1A_1$	5.66	6.39	0.73
	$1\,^1A_1$	$2\,^1B_1$	6.21	6.54	0.33
Tw-DMABN					
	$1\,^1A_1$	$1\,^1A_2$	4.11	4.42	0.31
	$1\,^1A_1$	$1\,^1B_1$	4.74	4.97	0.23

A.2 SUPPLEMENTARY DETAIL: SELECTION OF XC FUNCTIONAL

The performance of a suite of exchange-correlation (XC) functionals was evaluated by computing excitation energies for a small test set of molecules using time-dependent density functional theory within the Tamm–Dancoff approximation (TDA-TDDFT). The mean absolute errors (MAEs) and mean errors (MEs) for each XC functional were grouped by the type of excitation: either valence or core excitation. The selected reference method used for this evaluation was coupled-cluster singles, doubles and triples (CCSDT). The valence excitation energies are computed at the CCSDT level of theory using the aug-cc-pVTZ basis set. These data and the molecular geometries used for the TDA-TDDFT calculations are pulled from the first QUEST database.[20] Likewise, the core excitation data are computed at the CCSDT level of theory using the aug-cc-pCVTZ basis set. These data are computed using the MRCC code.[136–138] The reference data are summarized in Tables A.4 and A.5, respectively.

The TDA-TDDFT excitation energies were computed in QChem.[133] The analysis of these results is listed in Table A.6 for each functional screened. The subset of functionals screened includes those available in QChem version 5.4, excluding double-hybrid functionals. The time-dependent Hartree–Fock theory within the Tamm–Dancoff approximation (TDA-TDHF) energies (i.e., CIS energies) are included at the top for reference. The QTP17 functional was included in this screening set because it is a more recent addition to the Quantum Theory Project (QTP) family of functionals, which generally performed well, producing low errors.

Table A.4: Valence excitation energies used as reference data for the screening of exchange-correlation (XC) functionals. Energy in eV.

Molecule	State	Symm	Type	Energy
Ammonia				
	2	$A^{'}$	Ryd.	6.57
	3	$A^{''}$	Ryd.	8.14
	4	$A^{'}$	Ryd.	8.14
Carbon Monoxide				
	2	B_2	Val.	8.49
	3	B_1	Val.	8.49
	4	A_2	Val.	9.95
	5	A_2	Val.	10.08
	6	A_1	Val.	10.08
Formaldehyde				
	2	A_2	Val.	3.95
	3	B_1	Ryd.	7.16
	4	B_1	Ryd.	8.07
Acetylene				
	2	A_u	Val.	7.09
	3	B_{1u}	Val.	7.43
	4	A_u	Val.	7.43
Ethylene				
	2	B_{3u}	Ryd.	7.37
	3	B_{1u}	Val.	7.92
Water				
	2	B_1	Ryd.	7.59
	3	A_2	Ryd.	9.37
	4	A_1	Ryd.	9.95
Dinitrogen				
	2	B_{3g}	Val.	9.33
	3	B_{2g}	Val.	9.33
	4	A_u	Val.	9.89
	5	A_u	Val.	10.30
	6	B_{1u}	Val.	10.30

Table A.5: Core excitation energies used as reference data for the screening of exchange-correlation (XC) functionals. Energy in eV.

Molecule	K-Shell	State	Symm	Type	Energy
Ammonia					
	N	1	$A^{'}$	Ryd.	400.76
		2	$A^{''}$	Ryd.	402.44
		3	$A^{'}$	Ryd.	402.44
Carbon Monoxide					
	C	1	B_2	Val.	287.66
		2	B_1	Val.	287.66
	O	1	B_2	Val.	534.21
		2	B_1	Val.	534.21
Formaldehyde					
	C	1	B_1	Val.	286.26
		2	A_1	Ryd.	290.57
	O	1	A_1	Val.	530.93
		2	B_1	Ryd.	535.44
Acetylene					
	C	1	B_{2u}	Val.	286.06
		2	B_{3u}	Val.	286.06
		2	B_{2g}	Val.	286.11
		2	B_{3g}	Val.	286.11
Ethylene					
	C	1	B_{3u}	Val.	285.06
		2	B_{1g}	Val.	285.07
		3	B_{2u}	Ryd.	287.36
		4	A_g	Ryd.	287.40
Water					
	O	1	A_1	Ryd.	534.12
		2	B_2	Ryd.	535.95
Dinitrogen					
	N	1	B_{3u}	Val.	401.22
		2	B_{2u}	Val.	401.22
		3	B_{3g}	Val.	401.27
		4	B_{2g}	Val.	401.27

Although TDDFT generally performs best for describing excitations to low-lying valence orbitals, the work summarized here employed an augmented Dunning basis set. These basis sets are typically used to describe transitions to diffuse Rydberg states, which are known to be poorly described by TDDFT. However, the DFT/MRCI method is often applied with an augmented basis to simulate entire experimental spectra, where one must describe Rydberg states. Therefore, it was essential to ensure that the good performance of a given functional does not degrade when an augmented basis is employed. Moreover, it is necessary to include these Rydberg states in the test set to evaluate the performance of each functional for the intended purpose of simulating complete spectra. For example, the MN12-SX functional appears, unexpectedly, to produce larger errors for the valence excitation energies in Table A.6 than most other functionals; the MAE was determined as 0.98 eV. However, this functional is more accurate for describing valence-valence transitions (e.g. transitions of the type $\pi \rightarrow \pi^*$, or $\pi \rightarrow \sigma^*$, etc.) than it is for valence-Rydberg transitions. Evaluating the MAE of the MN12-SX functional for the two subsets of states within the valence excitation data set, the MAE for the valence transitions is only 0.44 eV, while the MAE for the Rydberg transitions is 1.9 eV. For the QTP17 functional, splitting the data set of valence transitions yields MAEs of 0.40 and 0.43 eV for the valence and Rydberg transitions, respectively. A complete summary of MAEs for the valence and core transitions, partitioned by transition type, has been tabulated in Table A.7.

Table A.6: Mean absolute errors (MAEs) and mean errors (MEs) of TDA-TDDFT excitation energies for test set of exchange-correlation (XC) functionals. Energy in eV.

	Valence		Core	
XC Functional	MAE	ME	MAE	ME
HF	0.78	0.16	12.61	12.61
SPW92	0.55	-0.53	23.07	-23.07
LDA	0.55	-0.52	23.07	-23.07
SVWN5	0.55	-0.52	23.07	-23.07
B97-D3	0.59	-0.59	19.57	-19.57
B97-D	0.59	-0.59	19.57	-19.57
PBE	0.61	-0.61	19.85	-19.85
BLYP	0.74	-0.74	19.20	-19.20
revPBE	0.66	-0.66	19.58	-19.58
BOP	0.70	-0.70	19.25	-19.25
BP86	0.56	-0.56	19.44	-19.44
BP86VWN	0.56	-0.56	19.45	-19.45
BPBE	0.58	-0.58	19.57	-19.57
EDF1	0.67	-0.67	19.43	-19.43
EDF2	0.47	-0.46	13.75	-13.75
GAM	0.51	-0.51	19.02	-19.02
HCTH93	0.63	-0.62	19.67	-19.67
HCTH120	0.56	-0.56	19.65	-19.65
HCTH147	0.56	-0.56	19.76	-19.76
HCTH407	0.54	-0.54	19.87	-19.87
HLE16	0.20	0.04	9.71	-9.71
KT1	0.60	-0.59	13.06	-13.06
KT2	0.55	-0.54	10.58	-10.58
KT3	0.65	-0.63	12.02	-12.02
mPW91	0.64	-0.64	19.50	-19.50
N12	0.92	-0.92	19.20	-19.20
OLYP	0.79	-0.79	19.57	-19.57
PBEOP	0.74	-0.74	19.54	-19.54
PBEsol	0.59	-0.57	21.42	-21.42
PW91	0.63	-0.63	19.56	-19.56
RPBE	0.67	-0.67	19.46	-19.46
rVV10	0.62	-0.62	18.65	-18.65
SOGGA	0.61	-0.59	21.55	-21.55
SOGGA11	1.13	-1.13	20.31	-20.31

	Valence		Core	
XC Functional	MAE	ME	MAE	ME
VV10	0.61	-0.61	18.65	-18.65
B3LYP	0.48	-0.47	12.68	-12.68
PBE0	0.35	-0.32	11.39	-11.39
revPBE0	0.38	-0.35	11.17	-11.17
B97	0.42	-0.40	12.94	-12.94
B1LYP	0.49	-0.46	10.80	-10.80
B1PW91	0.33	-0.29	11.06	-11.06
B3LYP5	0.50	-0.49	12.70	-12.70
B3P86	0.34	-0.32	12.88	-12.88
B3PW91	0.37	-0.35	12.90	-12.90
B5050LYP	0.33	-0.16	2.99	-2.99
B97-1	0.39	-0.37	12.31	-12.31
B97-2	0.38	-0.34	12.35	-12.35
B97-3	0.32	-0.28	10.44	-10.44
B97-K	0.33	-0.26	5.16	-5.16
BHLYP	0.34	-0.18	2.74	-2.74
HFLYP	0.90	0.34	12.76	12.76
MPW1K	0.27	-0.13	5.31	-5.31
MPW1LYP	0.54	-0.52	10.85	-10.85
MPW1PBE	0.37	-0.34	11.22	-11.22
MPW1PW91	0.37	-0.34	11.10	-11.10
O3LYP	0.61	-0.60	15.71	-15.71
PBEh-3c	0.26	-0.10	6.55	-6.55
PBE50	0.32	-0.03	3.25	-3.25
SOGGA11-X	0.33	-0.23	6.02	-6.02
WC04	0.64	-0.51	1.45	-1.03
WP04	0.27	-0.06	9.27	-9.27
X3LYP	0.49	-0.47	12.12	-12.12
wB97X-V	0.28	-0.21	13.10	-13.10
wB97X-D3	0.28	-0.22	12.27	-12.27
wB97X-D	0.33	-0.28	11.90	-11.90
CAM-B3LYP	0.38	-0.34	12.57	-12.57
QTP17	0.42	-0.01	1.05	-0.66
CAM-QTP00	0.43	-0.04	1.48	-1.41
CAM-QTP01	0.31	-0.23	11.29	-11.29
HSE-HJS	0.33	-0.30	11.37	-11.37
LC-rVV10	0.30	0.01	17.80	-17.80

	Valence		Core	
XC Functional	MAE	ME	MAE	ME
LC-VV10	0.30	0.02	17.79	-17.79
LC-wPBE08	0.30	0.02	17.78	-17.78
LRC-BOP	0.30	-0.05	18.09	-18.09
LRC-wPBE	0.35	-0.29	18.97	-18.97
LRC-wPBEh	0.38	-0.33	12.96	-12.96
N12-SX	0.49	-0.48	11.30	-11.30
rCAM-B3LYP	0.31	-0.22	12.23	-12.23
wB97	0.25	-0.06	17.46	-17.46
wB97X	0.23	-0.13	13.41	-13.41
wB97X-rV	0.29	-0.21	12.86	-12.86
B97M-V	0.50	-0.24	10.69	-10.69
B97M-rV	0.54	-0.25	10.70	-10.70
M06-L	0.47	-0.28	14.74	-14.74
TPSS	0.51	-0.49	16.74	-16.74
revTPSS	0.49	-0.43	15.92	-15.92
BLOC	0.50	-0.47	16.78	-16.78
M11-L	1.03	-0.87	6.70	-6.70
mBEEF	0.48	-0.27	15.46	-15.46
MGGA_MS0	0.57	-0.34	15.30	-15.30
MGGA_MS1	0.57	-0.37	15.77	-15.77
MGGA_MS2	0.59	-0.41	16.15	-16.15
MGGA_MVS	0.56	0.43	14.50	-14.50
MN12-L	1.01	-0.62	1.62	-1.16
MN15-L	0.61	-0.07	11.80	-11.80
oTPSS	0.60	-0.53	17.90	-17.90
PKZB	0.59	-0.57	19.47	-19.47
revM06-L	0.51	0.38	11.69	-11.69
SCAN	0.38	-0.12	14.94	-14.94
t-HCTH	0.56	-0.56	19.74	-19.74
TM	0.46	-0.41	16.91	-16.91
VSXC	0.46	-0.35	17.15	-17.15
M06-2X	0.31	-0.30	4.49	-4.49
M08-HX	0.47	-0.44	5.13	-5.13
TPSSh	0.42	-0.38	13.63	-13.63
revTPSSh	0.40	-0.32	12.89	-12.89
B1B95	0.36	-0.35	10.22	-10.22
BB1K	0.26	-0.20	5.69	-5.69

	Valence		Core	
XC Functional	MAE	ME	MAE	ME
BMK	0.24	-0.12	5.29	-5.29
dlDF	0.44	-0.26	1.01	0.50
M05	0.50	-0.50	12.35	-12.35
M05-2X	0.31	-0.03	3.64	-3.64
M06	0.83	-0.83	11.67	-11.67
M06-HF	0.48	-0.37	3.59	3.50
M08-SO	0.64	-0.64	6.72	-6.72
MGGA_MS2h	0.50	-0.32	13.38	-13.38
MGGA_MVSh	0.48	0.48	7.37	-7.37
MN15	0.39	-0.25	19.03	-19.03
MPW1B95	0.38	-0.37	9.28	-9.28
MPWB1K	0.28	-0.22	5.09	-5.09
PW6B95	0.42	-0.41	9.69	-9.69
PWB6K	0.28	-0.19	4.04	-4.04
revM06	0.22	-0.09	6.72	-6.72
SCAN0	0.25	0.01	7.83	-7.83
t-HCTHh	0.39	-0.38	14.61	-14.61
TPSS0	0.30	-0.22	9.06	-9.06
wB97M-V	0.52	-0.46	10.66	-10.66
M06-SX	0.23	0.00	6.88	-6.88
M11	0.57	-0.50	9.67	-9.67
MN12-SX	0.98	-0.90	1.53	-1.44
revM11	0.31	-0.30	15.26	-15.26
wB97M-rV	0.51	-0.45	10.67	-10.67
wM05-D	0.29	-0.18	9.46	-9.46
wM06-D3	0.42	0.12	11.07	-11.07

Table A.7: Mean absolute errors (MAEs) of TDA-TDDFT excitation energies for a test set of exchange-correlation (XC) functionals partitioned by valence or Rydberg-type transition. Energy in eV.

	Valence		Core	
XC Functional	Valence	Rydberg	Valence	Rydberg
HF	0.67	0.94	11.31	14.93
SPW92	0.20	1.14	21.95	25.06
LDA	0.20	1.14	21.95	25.06
SVWN5	0.20	1.14	21.95	25.06
B97-D3	0.22	1.21	18.48	21.50
B97-D	0.22	1.21	18.48	21.50
PBE	0.21	1.28	18.77	21.76
BLYP	0.31	1.46	18.06	21.22
revPBE	0.22	1.40	18.47	21.55
BOP	0.28	1.39	18.13	21.25
BP86	0.21	1.16	18.41	21.28
BP86VWN	0.20	1.16	18.42	21.29
BPBE	0.19	1.22	18.50	21.46
EDF1	0.23	1.42	18.29	21.47
EDF2	0.27	0.80	13.15	14.83
GAM	0.16	1.09	17.97	20.88
HCTH93	0.19	1.35	18.53	21.68
HCTH120	0.20	1.17	18.59	21.54
HCTH147	0.19	1.17	18.69	21.66
HCTH407	0.19	1.12	18.83	21.70
HLE16	0.21	0.18	9.12	10.77
KT1	0.25	1.18	12.18	14.62
KT2	0.25	1.06	9.76	12.03
KT3	0.25	1.30	11.07	13.70
mPW91	0.22	1.34	18.40	21.45
N12	0.37	1.83	17.85	21.58
OLYP	0.25	1.68	18.36	21.73
PBEOP	0.30	1.47	18.40	21.56
PBEsol	0.20	1.25	20.31	23.39
PW91	0.22	1.30	18.48	21.47
RPBE	0.23	1.40	18.36	21.41
rVV10	0.25	1.24	17.63	20.46
SOGGA	0.19	1.31	20.43	23.54
SOGGA11	0.33	2.46	18.87	22.87

	Valence		Core	
XC Functional	Valence	Rydberg	Valence	Rydberg
VV10	0.25	1.20	17.63	20.45
B3LYP	0.30	0.78	12.15	13.62
PBE0	0.24	0.52	11.06	11.97
revPBE0	0.25	0.60	10.83	11.79
B97	0.26	0.69	12.43	13.84
B1LYP	0.34	0.73	10.40	11.52
B1PW91	0.25	0.45	10.76	11.58
B3LYP5	0.31	0.82	12.15	13.66
B3P86	0.23	0.52	12.46	13.64
B3PW91	0.23	0.59	12.45	13.71
B5050LYP	0.44	0.16	3.34	2.37
B97-1	0.27	0.60	11.86	13.09
B97-2	0.25	0.59	11.91	13.14
B97-3	0.27	0.41	10.17	10.91
B97-K	0.39	0.24	5.29	4.92
BHHLYP	0.44	0.17	3.07	2.14
HFLYP	0.64	1.32	11.42	15.14
MPW1K	0.34	0.15	5.52	4.93
MPW1LYP	0.35	0.85	10.40	11.65
MPW1PBE	0.25	0.57	10.88	11.83
MPW1PW91	0.25	0.56	10.76	11.70
O3LYP	0.23	1.23	14.85	17.25
PBEh-3c	0.33	0.13	6.74	6.21
PBE50	0.37	0.25	3.68	2.49
SOGGA11-X	0.32	0.34	6.07	5.94
WC04	0.69	0.55	1.94	0.59
WP04	0.32	0.18	8.96	9.83
X3LYP	0.31	0.78	11.62	12.99
wB97X-V	0.32	0.23	12.81	13.62
wB97X-D3	0.29	0.27	11.95	12.85
wB97X-D	0.28	0.41	11.61	12.43
CAM-B3LYP	0.33	0.47	12.20	13.23
CAM-QTP00	0.48	0.36	1.95	0.64
CAM-QTP01	0.36	0.22	11.10	11.63
HSE-HJS	0.24	0.48	11.06	11.92
LC-rVV10	0.33	0.24	17.40	18.49

	Valence		Core	
XC Functional	Valence	Rydberg	Valence	Rydberg
LC-VV10	0.33	0.25	17.40	18.48
LC-wPBE08	0.34	0.25	17.39	18.47
LRC-BOP	0.37	0.20	17.77	18.66
LRC-wPBE	0.24	0.55	18.23	20.29
LRC-wPBEh	0.25	0.58	12.53	13.72
N12-SX	0.33	0.75	10.87	12.06
rCAM-B3LYP	0.38	0.19	12.00	12.63
wB97	0.31	0.16	17.05	18.19
wB97X	0.30	0.12	13.20	13.78
wB97X-rV	0.32	0.24	12.51	13.48
B97M-V	0.35	0.76	9.81	12.26
B97M-rV	0.40	0.77	9.81	12.27
M06-L	0.26	0.82	13.78	16.45
TPSS	0.15	1.11	15.73	18.54
revTPSS	0.15	1.05	14.93	17.69
BLOC	0.14	1.10	15.75	18.60
M11-L	0.48	1.95	5.64	8.59
mBEEF	0.44	0.54	14.53	17.10
MGGA_MS0	0.50	0.67	14.33	17.03
MGGA_MS1	0.46	0.76	14.78	17.53
MGGA_MS2	0.44	0.83	15.12	17.97
MGGA_MVS	0.72	0.30	13.70	15.92
MN12-L	0.60	1.68	0.89	2.92
MN15-L	0.53	0.75	10.77	13.63
oTPSS	0.26	1.18	16.82	19.81
PKZB	0.14	1.34	18.39	21.39
revM06-L	0.61	0.33	10.92	13.06
SCAN	0.32	0.49	14.08	16.46
t-HCTH	0.23	1.11	18.68	21.60
TM	0.14	1.00	15.94	18.63
VSXC	0.19	0.91	16.21	18.82
M06-2X	0.34	0.26	4.83	3.90
M08-HX	0.29	0.77	5.33	4.77
TPSSh	0.19	0.82	12.91	14.91
revTPSSh	0.19	0.76	12.19	14.14
B1B95	0.26	0.54	10.01	10.59
BB1K	0.32	0.17	5.91	5.31

	Valence		Core	
XC Functional	Valence	Rydberg	Valence	Rydberg
BMK	0.29	0.17	5.52	4.88
dlDF	0.57	0.22	0.67	1.62
M05	0.39	0.68	12.17	12.66
M05-2X	0.31	0.32	4.24	2.57
M06	0.52	1.35	11.08	12.70
M06-HF	0.60	0.27	2.65	5.24
M08-SO	0.57	0.78	7.07	6.11
MGGA_MS2h	0.45	0.59	12.61	14.74
MGGA_MVSh	0.48	0.46	7.32	7.47
MN15	0.51	0.18	18.95	19.16
MPW1B95	0.28	0.56	9.12	9.57
MPWB1K	0.33	0.19	5.33	4.67
PW6B95	0.30	0.62	9.45	10.11
PWB6K	0.35	0.15	4.35	3.49
revM06	0.28	0.13	6.89	6.44
SCAN0	0.31	0.14	7.70	8.07
t-HCTHh	0.24	0.65	14.02	15.65
TPSS0	0.26	0.38	8.80	9.54
wB97M-V	0.38	0.75	10.24	11.42
M06-SX	0.28	0.15	6.95	6.75
M11	0.39	0.86	9.67	9.69
MN12-SX	0.44	1.89	0.93	2.59
revM11	0.35	0.24	15.07	15.61
wB97M-rV	0.37	0.75	10.24	11.43
wM05-D	0.38	0.14	9.63	9.16
wM06-D3	0.30	0.62	11.31	10.65
QTP17	0.43	0.40	1.31	0.57

A.3 MOLECULAR GEOMETRIES: REPRESENTATIVE APPLICATIONS

Table A.8: Adenine, CCSD/cc-pVTZ.

C	1.2231808801	0.6068723167	-0.0393260123
C	-0.1768912390	0.5144928322	-0.0256384084
N	-1.1786411488	1.4749833789	-0.0232947104
C	-2.2759764921	0.7720204144	-0.0022316758
N	-2.0632158756	-0.5862984596	0.0098140182
C	-0.7043697989	-0.7673599809	-0.0026301133
N	-0.0191004917	-1.9157836748	0.0095517137
C	1.2885751393	-1.6873849346	0.0018376163
N	1.9440868111	-0.5154418603	-0.0183649776
H	1.9212021306	-2.5645833818	0.0158363910
H	-2.7564110655	-1.3101996009	0.0269872234
H	-3.2713796728	1.1828172077	0.0057983184
N	1.8642586207	1.8002739832	-0.1083859856
H	2.8427134417	1.7957637106	0.1119435171
H	1.3351838027	2.6230402027	0.1129828409

Table A.9: Cytosine, CCSD/aug-cc-pVTZ.

C	1.0471089794	-1.1826238790	-0.0000961585
C	-0.1966759323	-1.7056707748	0.0000869387
N	-1.2767039413	-0.8875538598	0.0002586550
C	-1.1843717255	0.5178641244	0.0007356591
N	0.0815890193	1.0516798859	0.0001834767
C	1.1227137559	0.2574212391	-0.0000538982
N	2.3436610779	0.8391630082	-0.0021035773
H	2.3926457469	1.8408370817	0.0032239485
H	3.1837925554	0.2972120985	0.0055005470
O	-2.2087835391	1.1668242563	-0.0006103572
H	-2.2116336941	-1.2563812342	0.0001195166
H	-0.3922562547	-2.7681887837	-0.0000934906
H	1.9263959988	-1.8056926715	-0.0007135729

Table A.10: Guanine, CCSD/cc-pVTZ.

C	1.7327873400	-0.4237993145	-0.0635217170
N	1.5015899026	0.9099436459	-0.0433337977
C	0.2506305144	1.3114771624	-0.0176203447
C	-0.7921230183	0.3806962935	-0.0106086695
C	-0.4049143131	-0.9501530908	-0.0224063426
N	-1.5810711737	-1.6508260875	-0.0006158439
C	-2.5972897396	-0.7170288819	0.0215602954
N	-2.1740152251	0.5136244918	0.0177161836
H	-3.6312688588	-1.0163992300	0.0401340959
H	-1.6758448443	-2.6487131827	-0.0066994241
N	0.8429715566	-1.4176673083	-0.0451300793
O	0.0246074501	2.6289014184	0.0061696458
H	-0.9304072858	2.7545363102	0.0252375033
N	3.0517993466	-0.7782988788	-0.1494884206
H	3.7027156448	-0.0685602390	0.1304658579
H	3.2632984240	-1.7194735387	0.1248202994

Table A.11: Thymine, CCSD/aug-cc-pVTZ.

C	2.6321619016	0.7113504160	-0.0001277583
C	1.1515616754	0.4887424137	0.0006425506
C	0.2413299767	1.4755556504	0.0004243853
N	-1.1192593260	1.2381026600	0.0003041386
C	-1.6697340566	-0.0228724677	0.0011070837
N	-0.7247598743	-1.0277327466	0.0003746538
C	0.6638420359	-0.8956047274	0.0016190038
O	1.3754027794	-1.8779067925	-0.0011601938
H	-1.0879167488	-1.9682815600	-0.0008896408
O	-2.8628865651	-0.2285084354	-0.0013004852
H	-1.7774503455	1.9961477362	-0.0001145890
H	0.5231582899	2.5190861120	-0.0001676986
H	3.0897564159	0.2516990221	0.8766529161
H	3.0889556805	0.2516746561	-0.8773238730
H	2.8617860241	1.7768077929	-0.0002661989

Table A.12: Uracil, CCSD(T)/cc-pVTZ.

C	1.2000235037	-1.1099460763	0.0000629297
C	1.2810448863	0.3488875031	0.0001197235
O	2.3013004857	1.0011805611	-0.0000322359
N	0.0340985335	0.9832347067	0.0001192770
C	-1.2139667619	0.3985163542	0.0001205577
N	-1.1740572205	-0.9817384871	0.0001042256
C	-0.0030302753	-1.7040842192	0.0000575674
H	-0.1229449311	-2.7772880419	0.0000034525
H	-2.0682899725	-1.4380153980	-0.0000539544
O	-2.2478313905	1.0263388868	-0.0001137917
H	0.0425784797	1.9913309412	0.0000032674
H	2.1169648305	-1.6739319199	0.0000076669

Table A.13: Chlorophyll a (model).

Mg	0.1096801	0.0774025	-0.5006436
C	-4.0597812	0.4910576	2.8228778
C	-1.0346300	-0.5747823	-5.9560518
C	5.4036377	1.8180171	-1.8069563
C	0.6861348	0.5257261	4.7560697
N	-1.6525660	-0.4503487	0.6255682
H	-4.0863429	1.2568908	2.0298898
C	-1.9444873	-0.4696275	-6.9416619
C	5.4948222	3.3497154	-1.6873721
C	-0.8082952	0.1043590	4.3691719
N	-0.8311478	-0.3377345	-2.2572501
O	1.0788158	0.6644188	5.8914095
C	-1.1968132	-1.1589245	5.1177843
N	1.0731609	0.4948806	1.2262185
C	-1.8276131	-1.9826967	7.2317877
C	-0.8056919	-0.0057564	2.8463758
C	-2.9809731	-1.0933373	-1.3361504
C	1.0125397	0.2385081	-3.7773578
C	3.2877851	1.1590951	0.4337698
C	-3.9102241	-2.9123658	1.1570748
C	-3.7001752	-1.5111365	-4.3607195
C	3.8890321	1.1789393	-4.5930129
C	3.8647726	1.4582160	3.5876028
C	-1.8249674	-0.3816013	1.9879666
C	-2.1025990	-0.8056412	-2.4048396
C	2.0132011	0.6133151	-2.8516017
C	2.3782182	0.9131606	1.4709158
O	-1.1743358	-2.2801667	4.6679456
C	-3.2526417	-0.7402405	2.3760591
C	-2.4115725	-0.9773997	-3.8163782
C	3.3559384	1.0478868	-3.1979854
C	2.6018939	1.0422305	2.9021042
O	-1.5544109	-0.8712398	6.3790413
C	-3.7968061	-1.3807738	1.0734236
C	-1.2780804	-0.5906625	-4.5163182
C	4.0037941	1.3050575	-2.0046361
C	1.3846160	0.6897742	3.4822085

C	-2.7686614	-0.9445733	0.0328945
C	-0.2841425	-0.2023691	-3.5142268
C	3.0436642	1.0292820	-0.9459041
C	0.4924436	0.3667454	2.4216303
N	1.8529345	0.6132377	-1.4973622
H	-3.6302703	0.9551597	3.7238312
H	-5.0991835	0.2069799	3.0557644
H	0.0177260	-0.6500642	-6.2538032
H	6.0353674	1.4848386	-2.6474469
H	5.8451009	1.3568385	-0.9070218
H	-3.0123642	-0.3462925	-6.7492224
H	-1.6317364	-0.4792328	-7.9893032
H	6.5382923	3.6736056	-1.5400681
H	4.9006405	3.7202357	-0.8365607
H	5.1105936	3.8416186	-2.5954353
H	-1.4483946	0.9189871	4.7489870
H	-2.0857529	-1.5603151	8.2109315
H	-0.9439309	-2.6331595	7.3196406
H	-2.6643245	-2.5825682	6.8408634
H	-3.9598243	-1.4873031	-1.6164577
H	1.2954321	0.3019255	-4.8291943
H	4.2876301	1.4876036	0.7260707
H	-4.2282251	-3.3451489	0.1953713
H	-4.6479531	-3.2042743	1.9223615
H	-2.9423875	-3.3628930	1.4306851
H	-3.5332497	-2.0686259	-5.2955422
H	-4.4207734	-0.7048976	-4.5884572
H	-4.1917091	-2.1905454	-3.6478548
H	4.9584847	1.4357838	-4.5949749
H	3.3620299	1.9663906	-5.1594287
H	3.7738297	0.2424778	-5.1647199
H	3.7197921	1.4750429	4.6767777
H	4.1897860	2.4626809	3.2668911
H	4.6939334	0.7654333	3.3631006
H	-3.2411650	-1.4784832	3.1934480
H	-4.7875204	-0.9683363	0.8174060

Table A.14: Chlorophyll a (full molecule)

Mg	6.2834142	-0.1838530	0.3538519
N	4.7957751	1.2260811	-0.3457836
C	3.5621793	0.9432608	-0.8846185
C	3.1126301	-0.3571612	-1.0572384
C	3.8675517	-1.5107782	-0.7283461
N	5.0683052	-1.7070226	-0.1781086
C	5.2543620	-3.0861703	-0.1314528
C	6.4180292	-3.6654729	0.3915400
H	6.4653753	-4.7565932	0.3895151
C	7.5306516	-2.9859154	0.9185566
N	7.6559116	-1.6165591	0.9832860
C	8.8687681	-1.3495044	1.5459953
C	9.3756001	-0.0504573	1.7791354
H	10.3591950	-0.0049565	2.2485876
C	8.7752171	1.1727061	1.4840280
N	7.5346335	1.3228971	0.9042777
C	7.2991865	2.6575846	0.7673995
C	6.1330629	3.2126323	0.1965772
H	6.1016929	4.3027447	0.1538124
C	5.0120560	2.5644116	-0.3185075
C	3.8442896	3.3360532	-0.9284453
C	2.8325342	2.2178803	-1.2942047
H	2.6974981	2.1800040	-2.3896184
C	1.4403379	2.3831967	-0.6593865
H	1.5354799	2.4246431	0.4399623
H	0.8233789	1.5037449	-0.8881632
C	0.6836073	3.6260746	-1.1614380
H	1.1542015	4.5591903	-0.8205875
H	0.6787876	3.6235167	-2.2649952
C	-0.7570275	3.6488398	-0.6889425
O	-1.2410827	4.4603322	0.0617289
O	-1.4455824	2.6153204	-1.2120406
C	-2.8168448	2.4562118	-0.7933839
H	-3.3887957	3.3526411	-1.0783612
H	-2.8338477	2.4125763	0.3110804
C	-3.3367758	1.1950473	-1.4065759
H	-2.6269511	0.3595284	-1.3753375

C	-4.5536235	0.9874991	-1.9458620
C	-5.6086195	2.0645091	-2.0530542
H	-5.1834270	3.0784089	-2.0168873
H	-6.1760646	1.9660311	-2.9926830
H	-6.3445763	1.9959468	-1.2330802
C	-4.9048407	-0.3951077	-2.4681677
H	-3.9649707	-0.9548493	-2.6002537
H	-5.3557294	-0.3004746	-3.4743855
C	-5.8429990	-1.2556348	-1.5876052
H	-5.4963044	-1.2035117	-0.5394247
H	-5.7034410	-2.3043914	-1.8985744
C	-7.3354464	-0.9103291	-1.6595508
H	-7.6730990	-0.9829509	-2.7117207
H	-7.4863968	0.1426862	-1.3704633
C	-8.2610956	-1.7772722	-0.7795629
H	-7.8916316	-1.6994918	0.2625553
C	-8.2268966	-3.2593903	-1.1804403
H	-7.2141612	-3.6851058	-1.1075244
H	-8.8767322	-3.8711621	-0.5360861
H	-8.5707027	-3.3915057	-2.2218756
C	-9.6941001	-1.2070357	-0.8047291
H	-10.1108481	-1.3314062	-1.8231572
H	-9.6390616	-0.1152624	-0.6395414
C	-10.6701784	-1.8012421	0.2175800
H	-10.2498256	-1.6767855	1.2339171
H	-10.7602906	-2.8881474	0.0589116
C	-12.0593381	-1.1525264	0.1619043
H	-12.5061087	-1.3297941	-0.8359521
H	-11.9286727	-0.0595548	0.2394497
C	-13.0590603	-1.6183033	1.2396126
H	-12.5859579	-1.4481727	2.2274869
C	-13.3729989	-3.1174649	1.1257718
H	-13.8282494	-3.3509112	0.1470022
H	-12.4723657	-3.7417005	1.2307607
H	-14.0850426	-3.4341423	1.9054401
C	-14.3690798	-0.8021539	1.2115147
H	-15.0807635	-1.2723177	1.9109312
H	-14.8311063	-0.9024305	0.2103081
C	-14.2399881	0.6898994	1.5526182

H	-13.6030917	1.1755532	0.7952632
H	-13.6957022	0.8029575	2.5089205
C	-15.5800492	1.4553316	1.6295209
H	-16.3257876	0.9606273	0.9780674
H	-15.4402743	2.4623929	1.2001500
C	-16.1883074	1.6430090	3.0353935
H	-15.4355789	2.1836704	3.6420122
C	-16.4967264	0.3248824	3.7575831
H	-17.2207608	-0.2820719	3.1863292
H	-15.5935811	-0.2863029	3.9117059
H	-16.9350640	0.5148238	4.7516512
C	-17.4422422	2.5250472	2.9624544
H	-18.2356439	2.0356552	2.3704366
H	-17.8529266	2.7242968	3.9660291
H	-17.2273542	3.4973631	2.4886856
H	3.4194584	3.9901187	-0.1449456
C	4.2517584	4.2159396	-2.1193731
H	3.3736325	4.7325488	-2.5387982
H	4.9834421	4.9847893	-1.8263287
H	4.7024790	3.6084508	-2.9207665
C	8.4253709	3.4209304	1.2840467
C	9.3492767	2.4942109	1.7431468
C	10.6624048	2.7093749	2.3445671
H	11.3705768	1.8802700	2.2315115
C	11.0885587	3.7922417	3.0198789
H	10.4461061	4.6514021	3.2236848
H	12.1068578	3.8347149	3.4158928
C	8.5402800	4.9133783	1.2639987
H	8.1595018	5.3717512	2.1947234
H	9.5901411	5.2267644	1.1559316
H	7.9709907	5.3531906	0.4311659
C	9.5592693	-2.5881340	1.8637568
C	8.7233921	-3.6144142	1.4671271
C	8.9492425	-5.0948830	1.6058857
H	8.5246954	-5.6190618	0.7327139
H	10.0319042	-5.3020273	1.5749125
C	8.3594299	-5.6971381	2.8934522
H	7.2706135	-5.5365254	2.9471785
H	8.5463109	-6.7824986	2.9444080

H	8.8046159	-5.2331038	3.7882077
C	10.9143974	-2.6843905	2.4978137
H	10.9164965	-2.2738191	3.5225955
H	11.2570618	-3.7274269	2.5639202
H	11.6736494	-2.1232654	1.9268007
C	4.0988775	-3.7664561	-0.6914404
C	3.2322183	-2.7403944	-1.0618173
C	1.9357853	-2.4555803	-1.6713255
O	1.1038604	-3.1993157	-2.1386397
C	1.7760430	-0.8697680	-1.5959614
H	0.9487242	-0.7103781	-0.8837270
C	1.3285501	-0.3140115	-2.9342451
O	2.0547153	0.1442505	-3.7848598
O	-0.0046175	-0.4031312	-3.0642677
C	-0.5541562	0.0185489	-4.3135470
H	-1.6421235	-0.0782080	-4.2155539
H	-0.2845025	1.0639134	-4.5258770
H	-0.1839037	-0.6153344	-5.1337172
C	3.8984878	-5.2414792	-0.8398273
H	4.6816169	-5.6950705	-1.4711307
H	3.9335159	-5.7571641	0.1350203
H	2.9238638	-5.4487479	-1.3033265

B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 REFERENCE DATA

The reference data are computed at the CVS-EOM-CCSDT level of theory with the aug-cc-pCVTZ basis set on non-hydrogen atoms, and the aug-cc-pVTZ basis set on hydrogen atoms. We denote this level of theory as simply aCTZ/CCSDT hereafter. These basis sets are also used in the corresponding DFT/MRCI calculations. The training set consists of 9 molecules, and the validation set consists of 8 molecules. These are detailed in Table B.1. The data sets include 1s $\rightarrow \pi^*$ and σ^* transitions, as well as transitions from the 1s core orbital to Rydberg s- and p-type orbitals. This reference data includes singlet and triplet multiplicity excitation energies and the ionization potential corresponding to the doublet core-ionized states for each K-edge. In Tables B.2 and B.3, the computed results for DFT/MRCI with the CVS-QE12 Hamiltonian are presented alongside the reference values. Table B.2 contains the training set data; likewise, Table B.3 contains the validation set data.

Table B.1: Summary of contents of training (top) and validation (bottom) sets of computed ab initio reference data.

Molecule	Formula	Transitions
Acetylene	C_2H_2	π^*, Ry
Ethylene	C_2H_4	π^*, Ry
Methane	CH_4	Ry
Water	H_2O	Ry
Ammonia	NH_3	Ry
Hydrogen Fluoride	HF	σ^*, Ry
Fluoroamine	NH_2F	σ^*, Ry
Azomethine	CH_2NH	π^*, Ry
Hypofluorous Acid	HOF	σ^*, Ry
Carbon Monoxide	CO	π^*, Ry
Dinitrogen	N_2	π^*, Ry
Formaldehyde	H_2CO	π^*, Ry
Hydrogen Cyanide	HCN	π^*, Ry
Hydrogen Isocyanide	HNC	π^*, Ry
Nitroxyl	HNO	π^*, Ry
Methyl Fluoride	CH_3F	σ^*, Ry
Ketene	C_2H_2O	π^*

Table B.2: Training set reference data computed at the aCTZ/CCSDT level of theory and aCTZ/DFT/MRCI [CVS-QE12] results. Energy in eV.

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
C_2H_2					
	$1\ ^1B_u$	C2 1s $\rightarrow \pi_1$*	286.06	285.28	-0.78
	$2\ ^1B_u$	C2 1s $\rightarrow \pi_2$*	286.06	285.28	-0.78
	$3\ ^1B_g$	C1 1s $\rightarrow \pi_1$*	286.11	285.33	-0.78
	$4\ ^1B_g$	C1 1s $\rightarrow \pi_2$*	286.11	285.33	-0.78
	$1\ ^3B_u$	C2 1s $\rightarrow \pi_1$*	285.52	285.00	-0.52
	$2\ ^3B_u$	C2 1s $\rightarrow \pi_2$*	285.52	285.00	-0.52
	$3\ ^3B_g$	C1 1s $\rightarrow \pi_1$*	285.61	285.07	-0.54
	$4\ ^3B_g$	C1 1s $\rightarrow \pi_2$*	285.61	285.07	-0.54
	$1\ ^2B_u$	C2 1s $\rightarrow \infty$	291.23	290.84	-0.39
	$2\ ^2A_g$	C1 1s $\rightarrow \infty$	291.31	290.91	-0.40
C_2H_4					
	$1\ ^1B_u$	C2 1s $\rightarrow \pi$*	285.06	284.82	-0.24
	$2\ ^1B_g$	C1 1s $\rightarrow \pi$*	285.07	284.83	-0.24
	$3\ ^1B_u$	C2 1s $\rightarrow$ 3s	287.36	286.77	-0.59
	$4\ ^1A_g$	C1 1s $\rightarrow$ 3s	287.40	286.80	-0.60
	$1\ ^3B_u$	C2 1s $\rightarrow \pi$*	284.62	284.58	-0.04
	$2\ ^3B_g$	C1 1s $\rightarrow \pi$*	284.67	284.61	-0.06
	$3\ ^3B_u$	C2 1s $\rightarrow$ 3s	287.12	286.67	-0.45
	$4\ ^3A_g$	C1 1s $\rightarrow$ 3s	287.15	286.70	-0.45
	$1\ ^2B_u$	C2 1s $\rightarrow \infty$	290.79	290.53	-0.26
	$2\ ^2A_g$	C1 1s $\rightarrow \infty$	290.84	290.56	-0.28
CH_4					
	$1\ ^1A_1$	C 1s $\rightarrow$ Ry 3s	286.85	286.64	-0.21
	$2\ ^1A_1$	C 1s $\rightarrow$ Ry 3p	288.18	287.92	-0.26
	$3\ ^1B_2$	C 1s $\rightarrow$ Ry 3p	288.18	287.92	-0.26
	$4\ ^1B_1$	C 1s $\rightarrow$ Ry 3p	288.18	287.92	-0.26
	$1\ ^3A_1$	C 1s $\rightarrow$ Ry 3s	286.47	286.31	-0.16
	$2\ ^3A_1$	C 1s $\rightarrow$ Ry 3p	287.97	287.79	-0.18
	$3\ ^3B_2$	C 1s $\rightarrow$ Ry 3p	287.97	287.79	-0.18
	$4\ ^3B_1$	C 1s $\rightarrow$ Ry 3p	287.97	287.79	-0.18
	$1\ ^2A_1$	C 1s $\rightarrow \infty$	290.84	290.81	-0.03
H_2O					
	$1\ ^1A_1$	O 1s $\rightarrow$ Ry 3s	534.12	535.27	1.15
	$2\ ^1B_1$	O 1s $\rightarrow$ Ry 3p	535.95	536.85	0.90

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
H_2O					
	$3\ ^1A_1$	O 1s → Ry 3p	537.65	538.66	1.01
	$1\ ^3A_1$	O 1s → Ry 3s	533.65	534.79	1.14
	$2\ ^3B_1$	O 1s → Ry 3p	535.57	536.48	0.91
	$3\ ^3A_1$	O 1s → Ry 3p	537.48	538.41	0.93
	$1\ ^2A_1$	O 1s → ∞	539.78	539.96	0.18
NH_3					
	$1\ ^1A'$	N 1s → Ry 3s	400.76	401.27	0.51
	$2\ ^1A'$	N 1s → Ry 3p	402.44	402.83	0.39
	$3\ ^1A''$	N 1s → Ry 3p	402.44	402.83	0.39
	$4\ ^1A'$	N 1s → Ry 3p	403.57	404.04	0.47
	$1\ ^3A'$	N 1s → Ry 3s	400.34	400.89	0.55
	$2\ ^3A'$	N 1s → Ry 3p	402.15	402.63	0.48
	$3\ ^3A''$	N 1s → Ry 3p	402.15	402.63	0.48
	$4\ ^3A'$	N 1s → Ry 3p	403.43	403.92	0.49
	$1\ ^2A'$	N 1s → ∞	405.58	405.63	0.05
HF					
	$1\ ^1A_1$	F 1s → σ*	687.50	689.11	1.60
	$2\ ^1A_1$	F 1s → Ry 3s	691.19	692.69	1.50
	$3\ ^1B_1$	F 1s → Ry 3p	692.08	693.86	1.78
	$4\ ^1B_2$	F 1s → Ry 3p	692.08	693.86	1.78
	$1\ ^3A_1$	F 1s → σ*	686.99	688.61	1.62
	$2\ ^3A_1$	F 1s → Ry 3s	691.03	692.44	1.41
	$3\ ^3B_1$	F 1s → Ry 3p	691.92	693.61	1.69
	$4\ ^3B_2$	F 1s → Ry 3p	691.92	693.61	1.69
	$1\ ^2A_1$	F 1s → ∞	694.00	694.28	0.28
NH_2F					
	$1\ ^1A'$	N 1s → σ*	402.15	401.96	-0.19
	$2\ ^1A'$	N 1s → Ry 3s	403.51	403.81	0.30
	$3\ ^1A''$	N 1s → Ry 3p	404.81	405.00	0.19
	$1\ ^3A'$	N 1s → σ*	400.78	401.16	0.38
	$2\ ^3A'$	N 1s → Ry 3s	403.06	403.47	0.41
	$3\ ^3A''$	N 1s → Ry 3p	404.69	404.77	0.08
	$1\ ^2A'$	N 1s → ∞	408.24	408.44	0.20
	$1\ ^1A'$	F 1s → σ*	686.64	687.26	0.62

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
NH_2F					
	$2\,^1A'$	F 1s → Ry 3s	688.92	689.99	1.07
	$3\,^1A''$	F 1s → Ry 3p	690.24	691.10	0.86
	$1\,^3A'$	F 1s → σ*	685.68	686.53	0.85
	$2\,^3A'$	F 1s → Ry 3s	688.89	689.79	0.90
	$3\,^3A''$	F 1s → Ry 3p	690.24	690.97	0.73
	$1\,^2A'$	F 1s → ∞	693.22	693.69	0.47
CH_2NH					
	$1\,^1A''$	C 1s → π*	285.65	285.33	-0.32
	$2\,^1A'$	C 1s → Ry 3s	288.75	288.29	-0.46
	$3\,^1A'$	C 1s → Ry $3p_z$	289.51	289.08	-0.43
	$1\,^3A''$	C 1s → π*	285.04	285.00	-0.04
	$2\,^3A'$	C 1s → Ry 3s	288.46	288.09	-0.37
	$3\,^3A'$	C 1s → Ry $3p_z$	289.33	288.94	-0.39
	$1\,^2A'$	C 1s → ∞	292.75	292.66	-0.09
	$1\,^1A''$	N 1s → π*	398.57	398.66	0.09
	$2\,^1A'$	N 1s → Ry 3s	401.50	401.32	-0.18
	$3\,^1A'$	N 1s → Ry $3p_z$	402.48	402.22	-0.26
	$1\,^3A''$	N 1s → π*	398.12	398.37	0.25
	$2\,^3A'$	N 1s → Ry 3s	401.23	401.12	-0.11
	$3\,^3A'$	N 1s → Ry $3p_z$	402.40	402.13	-0.27
	$1\,^2A'$	N 1s → ∞	405.51	405.43	-0.08
HOF					
	$1\,^1A'$	O 1s → σ*	532.50	531.77	-0.73
	$2\,^1A'$	O 1s → Ry 3s	537.20	537.37	0.17
	$1\,^3A'$	O 1s → σ*	531.80	531.05	-0.75
	$2\,^3A'$	O 1s → Ry 3s	536.71	537.00	0.29
	$1\,^2A'$	O 1s → ∞	542.51	542.71	0.20
	$1\,^1A'$	F 1s → σ*	683.88	683.48	-0.40
	$2\,^1A'$	F 1s → Ry 3s	689.31	689.22	-0.09
	$1\,^3A'$	F 1s → σ*	683.03	682.86	-0.17
	$2\,^3A'$	F 1s → Ry 3s	689.29	688.87	-0.42
	$1\,^2A'$	F 1s → ∞	694.26	693.82	-0.44

Table B.3: Validation set reference data computed at the aCTZ/CCSDT level of theory and aCTZ/DFT/MRCI [CVS-QE12] results. Energy in eV.

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
CO					
	$1\ ^1B_2$	C 1s $\rightarrow \pi_1$*	287.66	286.58	-1.08
	$2\ ^1B_1$	C 1s $\rightarrow \pi_2$*	287.66	286.58	-1.08
	$3\ ^1A_1$	C 1s $\rightarrow$ Ry 3s	292.87	292.85	-0.02
	$4\ ^1B_2$	C 1s $\rightarrow$ Ry 3p	293.92	292.96	-0.96
	$5\ ^1B_1$	C 1s $\rightarrow$ Ry 3p	293.92	292.96	-0.96
	$1\ ^3B_2$	C 1s $\rightarrow \pi_1$*	286.25	285.98	-0.27
	$2\ ^3B_1$	C 1s $\rightarrow \pi_2$*	286.25	285.98	-0.27
	$3\ ^3A_1$	C 1s $\rightarrow$ Ry 3s	292.77	292.73	-0.04
	$4\ ^3B_2$	C 1s $\rightarrow$ Ry 3p	293.79	292.79	-1.00
	$5\ ^3B_1$	C 1s $\rightarrow$ Ry 3p	293.79	292.79	-1.00
	$1\ ^2A_1$	C 1s $\rightarrow \infty$	296.32	296.46	0.14
	$1\ ^1B_2$	O 1s $\rightarrow \pi_1$*	534.21	534.24	0.03
	$2\ ^1B_1$	O 1s $\rightarrow \pi_2$*	534.21	534.24	0.03
	$3\ ^1A_1$	O 1s $\rightarrow$ Ry 3s	538.78	539.60	0.82
	$4\ ^1B_2$	O 1s $\rightarrow$ Ry 3p	539.98	539.75	-0.23
	$5\ ^1B_1$	O 1s $\rightarrow$ Ry 3p	539.98	539.75	-0.23
	$1\ ^3B_2$	O 1s $\rightarrow \pi_1$*	533.78	533.91	0.13
	$2\ ^3B_1$	O 1s $\rightarrow \pi_2$*	533.78	533.91	0.13
	$3\ ^3A_1$	O 1s $\rightarrow$ Ry 3s	538.68	539.51	0.83
	$4\ ^3B_2$	O 1s $\rightarrow$ Ry 3p	539.90	539.58	-0.32
	$5\ ^3B_1$	O 1s $\rightarrow$ Ry 3p	539.90	539.67	-0.23
	$1\ ^2A_1$	O 1s $\rightarrow \infty$	542.33	541.18	-1.15
N_2					
	$1\ ^1B_{3u}$	N2 1s $\rightarrow \pi_1$*	401.22	400.74	-0.48
	$2\ ^1B_{2u}$	N2 1s $\rightarrow \pi_2$*	401.22	400.74	-0.48
	$3\ ^1B_{3g}$	N1 1s $\rightarrow \pi_1$*	401.27	400.78	-0.49
	$4\ ^1B_{2g}$	N1 1s $\rightarrow \pi_2$*	401.27	400.78	-0.49
	$5\ ^1A_g$	N2 1s $\rightarrow$ Ry 3s	406.79	405.98	-0.81
	$6\ ^1A_g$	N1 1s $\rightarrow$ Ry 3s	406.87	406.04	-0.83
	$1\ ^3B_{3u}$	N2 1s $\rightarrow \pi_1$*	400.34	400.20	-0.14
	$2\ ^3B_{2u}$	N2 1s $\rightarrow \pi_2$*	400.34	400.21	-0.13
	$3\ ^3B_{3g}$	N1 1s $\rightarrow \pi_1$*	400.43	400.26	-0.17

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
N_2					
	$4\,^3B_{2g}$	N1 1s $\rightarrow \pi_2$*	400.43	400.27	-0.16
	$5\,^3A_g$	N2 1s $\rightarrow$ Ry 3s	406.69	405.83	-0.86
	$6\,^3A_g$	N1 1s $\rightarrow$ Ry 3s	406.77	405.88	-0.89
	$1\,^2B_{1u}$	N2 1s $\rightarrow \infty$	410.01	409.75	-0.26
	$2\,^2A_g$	N1 1s $\rightarrow \infty$	409.90	409.69	-0.21
H_2CO					
	$1\,^1B_2$	C 1s $\rightarrow \pi$*	286.26	285.89	-0.37
	$2\,^1A_1$	C 1s $\rightarrow$ Ry 3s	290.57	290.09	-0.48
	$3\,^1B_1$	C 1s $\rightarrow$ Ry 3p	291.57	291.00	-0.57
	$1\,^3B_2$	C 1s $\rightarrow \pi$*	285.40	285.47	0.07
	$2\,^3A_1$	C 1s $\rightarrow$ Ry 3s	290.10	289.78	-0.32
	$3\,^3B_1$	C 1s $\rightarrow$ Ry 3p	291.32	290.86	-0.46
	$1\,^2A_1$	C 1s $\rightarrow \infty$	294.67	294.46	-0.21
	$1\,^1B_2$	O 1s $\rightarrow \pi$*	530.93	531.29	0.36
	$2\,^1A_1$	O 1s $\rightarrow$ Ry 3s	535.44	535.12	-0.32
	$3\,^1A_1$	O 1s $\rightarrow$ Ry 3p	536.30	536.13	-0.17
	$1\,^3B_2$	O 1s $\rightarrow \pi$*	530.52	530.98	0.46
	$2\,^3A_1$	O 1s $\rightarrow$ Ry 3s	535.44	535.03	-0.41
	$3\,^3A_1$	O 1s $\rightarrow$ Ry 3p	536.09	535.87	-0.22
	$1\,^2A_1$	O 1s $\rightarrow \infty$	539.21	538.96	-0.25
HCN					
	$1\,^1A_1$	C 1s $\rightarrow \pi_1$*	286.82	285.84	-0.98
	$2\,^1A_1$	C 1s $\rightarrow \pi_2$*	286.82	285.84	-0.98
	$3\,^1A_1$	C 1s $\rightarrow$ Ry 3s	289.53	288.67	-0.86
	$4\,^1A_1$	C 1s $\rightarrow$ Ry 3p	290.08	289.20	-0.88
	$1\,^3A_1$	C 1s $\rightarrow \pi_1$*	286.17	285.52	-0.65
	$2\,^3A_1$	C 1s $\rightarrow \pi_2$*	286.17	285.52	-0.65
	$3\,^3A_1$	C 1s $\rightarrow$ Ry 3s	289.00	288.34	-0.66
	$4\,^3A_1$	C 1s $\rightarrow$ Ry 3p	290.15	289.12	-1.03
	$1\,^2A_1$	C 1s $\rightarrow \infty$	293.42	293.37	-0.05
	$1\,^1B_1$	N 1s $\rightarrow \pi_1$*	400.02	398.95	-1.07
	$2\,^1B_2$	N 1s $\rightarrow \pi_2$*	400.02	398.99	-1.03
	$3\,^1A_1$	N 1s $\rightarrow$ Ry 3s	402.74	401.69	-1.05
	$4\,^1A_1$	N 1s $\rightarrow$ Ry 3p	403.22	402.33	-0.89

Molecule	State	Assignment	CCSDT	DFT/MRCI	ΔE
HCN					
	$1\,^3B_1$	N 1s $\rightarrow \pi_1$*	399.35	398.65	-0.70
	$2\,^3B_2$	N 1s $\rightarrow \pi_2$*	399.35	398.72	-0.63
	$3\,^3A_1$	N 1s $\rightarrow$ Ry 3s	402.63	401.64	-0.99
	$4\,^3A_1$	N 1s $\rightarrow$ Ry 3p	403.09	402.14	-0.95
	$1\,^2A_1$	N 1s $\rightarrow \infty$	406.66	406.20	-0.46
HNC					
	$1\,^1B_1$	C 1s $\rightarrow \pi_1$*	287.00	285.70	-1.30
	$2\,^1B_2$	C 1s $\rightarrow \pi_2$*	287.00	285.70	-1.30
	$3\,^1A_1$	C 1s $\rightarrow$ Ry 3s	289.33	288.11	-1.22
	$4\,^1A_1$	C 1s $\rightarrow$ Ry 3p	290.92	289.70	-1.22
	$1\,^3B_1$	C 1s $\rightarrow \pi_1$*	285.77	285.20	-0.57
	$2\,^3B_2$	C 1s $\rightarrow \pi_2$*	285.77	285.20	-0.57
	$3\,^3A_1$	C 1s $\rightarrow$ Ry 3s	289.30	288.07	-1.23
	$4\,^3A_1$	C 1s $\rightarrow$ Ry 3p	290.65	289.51	-1.14
	$1\,^2A_1$	C 1s $\rightarrow \infty$	293.64	293.35	-0.29
	$1\,^1B_1$	N 1s $\rightarrow \pi_1$*	400.93	400.76	-0.17
	$2\,^1B_2$	N 1s $\rightarrow \pi_2$*	400.93	400.76	-0.17
	$3\,^1A_1$	N 1s $\rightarrow$ Ry 3s	402.85	402.49	-0.36
	$4\,^1A_1$	N 1s $\rightarrow$ Ry 3p	404.33	404.20	-0.13
	$1\,^3B_1$	N 1s $\rightarrow \pi_1$*	400.58	400.47	-0.11
	$2\,^3B_2$	N 1s $\rightarrow \pi_2$*	400.58	400.47	-0.11
	$3\,^3A_1$	N 1s $\rightarrow$ Ry 3s	402.38	402.13	-0.25
	$4\,^3A_1$	N 1s $\rightarrow$ Ry 3p	404.36	404.09	-0.27
	$1\,^2A_1$	N 1s $\rightarrow \infty$	407.17	406.68	-0.49
HNO					
	$1\,^1A''$	O 1s $\rightarrow \pi$*	530.66	531.43	0.77
	$2\,^1A'$	O 1s $\rightarrow$ Ry 3s	537.01	537.01	-0.00
	$3\,^1A'$	O 1s $\rightarrow$ Ry 3p	538.68	538.60	-0.08
	$1\,^3A''$	O 1s $\rightarrow \pi$*	530.15	531.06	0.91
	$2\,^3A'$	O 1s $\rightarrow$ Ry 3s	537.01	536.89	-0.12
	$3\,^3A'$	O 1s $\rightarrow$ Ry 3p	538.30	538.16	-0.14
	$1\,^2A'$	O 1s $\rightarrow \infty$	541.47	541.68	0.21
	$1\,^1A''$	N 1s $\rightarrow \pi$*	399.27	399.62	0.35
	$2\,^1A'$	N 1s $\rightarrow$ Ry 3s	405.17	405.11	-0.06

Molecule	State	Assignment	CCSDT	DFT/MRCI	Δ E
HNO					
	$3\ ^1A'$	N 1s → Ry 3p	407.14	407.63	0.49
	$1\ ^3A''$	N 1s → π*	398.45	399.16	0.71
	$2\ ^3A'$	N 1s → Ry 3s	404.62	404.77	0.15
	$3\ ^3A'$	N 1s → Ry 3p	407.01	407.06	0.05
	$1\ ^2A'$	N 1s → ∞	410.09	410.52	0.43
CH_3F					
	$1\ ^1A'$	C 1s → σ*	289.42	289.15	-0.27
	$2\ ^1A'$	C 1s → Ry 3s	289.66	289.44	-0.22
	$3\ ^1A'$	C 1s → Ry 3p	290.68	290.40	-0.28
	$4\ ^1A''$	C 1s → Ry 3p	290.68	290.40	-0.28
	$1\ ^3A'$	C 1s → σ*	288.82	288.85	0.03
	$2\ ^3A'$	C 1s → Ry 3s	289.14	289.07	-0.07
	$3\ ^3A'$	C 1s → Ry 3p	290.37	290.24	-0.13
	$4\ ^3A''$	C 1s → Ry 3p	290.37	290.24	-0.13
	$1\ ^2A'$	C 1s → ∞	293.70	293.94	0.24
	$1\ ^1A'$	F 1s → σ*	688.22	689.55	1.33
	$2\ ^1A'$	F 1s → Ry 3s	688.88	690.30	1.42
	$3\ ^1A'$	F 1s → Ry 3p	689.90	691.32	1.42
	$4\ ^1A''$	F 1s → Ry 3p	689.90	691.32	1.42
	$1\ ^3A'$	F 1s → σ*	687.63	689.01	1.38
	$2\ ^3A'$	F 1s → Ry 3s	688.79	690.18	1.39
	$3\ ^3A'$	F 1s → Ry 3p	689.90	691.25	1.35
	$4\ ^3A''$	F 1s → Ry 3p	689.90	691.25	1.35
	$1\ ^2A'$	F 1s → ∞	692.48	693.09	0.61
C_2H_2O					
	$1\ ^1B_2$	C2 1s → π*	285.26	284.96	-0.30
	$2\ ^1B_2$	C1 1s → π*	287.47	287.98	0.51
	$1\ ^3B_2$	C2 1s → π*	285.46	284.92	-0.54
	$2\ ^3B_2$	C1 1s → π*	286.23	286.52	0.29
	$1\ ^2A_1$	C2 1s → ∞	295.02	294.94	-0.08
	$2\ ^2A_1$	C1 1s → ∞	291.14	290.82	-0.32
	$1\ ^1B_2$	O 1s → π*	532.92	533.24	0.32
	$1\ ^2A_1$	O 1s → ∞	540.26	532.97	0.45

B.2 Parameter Sets

The damping function applied to the off-diagonal matrix elements of the Hamiltonian has three parameters: d_1, d_2, and d_3. The form of this function is given in Eqn. 2.2 of the main text. The first parameter, d_1, corresponds to the default damping of off-diagonal matrix elements at a small value of $\Delta E_{\mathbf{ww'}}$. The latter two parameters, d_2 and d_3, affect how steeply the damping function approaches zero. Whereas the n in QEn corresponds to d_3, the choice of the integer exponent term in the DFT/MRCI Hamiltonian damping function, d_2 is an explicitly optimized parameter which is not otherwise strictly defined in the name of the Hamiltonian. As a part of the parameter optimization procedure, constraints were placed on d_2 to reduce the N_{root} dependence of the resultant parameter set for the DFT/MRCI CVS-QEn Hamiltonian. The full set of parameters yielded by applying each constraint is summarized in Table B.4, corresponding to $E_{sel} = 1\ E_h$. The constraints $d_2 \geq 2.303$ and $d_2 \geq 4.605$ correspond to respective damping function values of 0.10 and 0.01 at $\Delta E_{\mathbf{ww'}} = 1\ E_h$. The constraint $d_2 \geq 0.200$ corresponds to a damping function value close to 0.8 at this value of $\Delta E_{\mathbf{ww'}}$. Unconstrained optimizations were observed to push this value to zero to fit to outliers in the training data set, namely H_2O and HF. This, in turn, was observed to result in a strong N_{root} dependence, wherein the computed excitation energies are shifted up to 1.0 eV when an increasing number of roots (100 roots or more) were requested in a calculation. The computational cost of such calculations increases accordingly. Thus, a minimal constraint was set such that $d_2 \geq 0.200$.

The mean absolute errors (MAEs) for the parameter sets listed in Table B.4 are tabulated in Tables B.5 and B.6 for the training set and validation set data, respectively. These data are

Table B.4: Complete set of DFT/MRCI CVS-QE*n* Hamiltonian parameters for $E_{sel} = 1.0E_h$.

Constraint	$p_C^{(v,v)}$	$p_C^{(c,v)}$	p_X	d_1	d_2	d_3	E_{sel}
$d_2 \geq 0.200$	0.425815	0.565179	0.254799	0.494512	0.200010	6.0	1.0
	0.425815	0.566093	0.254799	0.492991	0.200001	8.0	1.0
	0.425815	0.565268	0.254799	0.490723	0.200921	10.0	1.0
	0.425815	0.565711	0.254799	0.490446	0.200036	12.0	1.0
	0.425815	0.565546	0.254799	0.491028	0.200028	14.0	1.0
	0.425815	0.565868	0.254799	0.490210	0.200038	16.0	1.0
$d_2 \geq 2.303$	0.425815	0.539416	0.254799	0.553897	2.302585	6.0	1.0
	0.425815	0.548974	0.254799	0.540495	2.302619	8.0	1.0
	0.425815	0.558624	0.254799	0.534700	2.303182	10.0	1.0
	0.425815	0.564040	0.254799	0.529467	2.302585	12.0	1.0
	0.425815	0.567684	0.254799	0.522341	2.302903	14.0	1.0
	0.425815	0.569995	0.254799	0.518889	2.302585	16.0	1.0
$d_2 \geq 4.605$	0.425815	0.524422	0.254799	0.565402	4.605213	6.0	1.0
	0.425815	0.537032	0.254799	0.554308	4.605170	8.0	1.0
	0.425815	0.542700	0.254799	0.539983	4.605170	10.0	1.0
	0.425815	0.548460	0.254799	0.533383	4.613243	12.0	1.0
	0.425815	0.557102	0.254799	0.531182	4.605334	14.0	1.0
	0.425815	0.560389	0.254799	0.526050	4.609609	16.0	1.0

presented graphically in Figure B.1. To compare the N_{root} dependence of these parameter sets, a representative example was selected. The carbon NEXAFS of thymine was computed for each parameter set. These calculations employ the aug-cc-pCVDZ basis set. The energetic shift of the first root in the spectrum was determined by separately computing both $N = 1$ and $N = 40$ core excited-state roots. This value of ΔE, for the first carbon core-excited state root, is reported in Table B.7, and depicted graphically in Figure B.2, for each parameter set listed in Table B.7. The spectra simulated using each of the parameter sets (using the latter computation of 40 core-excited state roots) are depicted across Figures B.3, B.4, and B.5, with comparison to the experimental NEXAFS spectrum.[257] These spectra are simulated by Lorentzian convolution

using a FWHM of 0.40 eV.

Table B.5: Mean absolute errors of CVS-QE*n* Hamiltonian parameter sets across the training data set. Errors in eV.

	d_3					
d_2	6	8	10	12	14	16
≥ 0.200	0.30	0.31	0.32	0.33	0.33	0.34
≥ 2.303	0.65	0.58	0.55	0.53	0.51	0.49
≥ 4.605	0.77	0.69	0.63	0.58	0.55	0.54

Table B.6: Mean absolute errors of CVS-QE*n* Hamiltonian parameter sets across the validation data set. Errors in eV.

	d_3					
d_2	6	8	10	12	14	16
≥ 0.200	0.46	0.47	0.47	0.48	0.49	0.49
≥ 2.303	0.59	0.55	0.54	0.54	0.51	0.51
≥ 4.605	0.66	0.62	0.57	0.54	0.54	0.53

Table B.7: ΔE for the first excited-state root in carbon K-shell excitations of thymine from $N = 1$ to $N = 40$ roots. The ΔE values are tabulated corresponding to each Hamiltonian parameter set in Table B.4. Energies in eV

	d_3					
d_2	6	8	10	12	14	16
≥ 0.200	-0.53	-0.54	-0.54	-0.56	-0.56	-0.57
≥ 2.303	-0.24	-0.30	-0.36	-0.40	-0.43	-0.46
≥ 4.605	-0.18	-0.20	-0.26	-0.30	-0.35	-0.38

The final parameter set selected in this work employs the constraint $d_2 \geq 2.303$, with an exponent value of d_3 = 12. This parameter set was determined to be balanced in its MAE across the training and validation data sets and in the observed N_{root} dependence for the selected example. The data points corresponding to this Hamiltonian, which we denote CVS-QE12, are indicated

and labeled in the plots of Figures B.1 and B.2. The parameters for CVS-QE12 are listed to the full six decimal places in Table B.8, as determined by the optimization routine employed in this work. The DFT/MRCI Hamiltonian has canonically been presented in terms of two parameter sets, respectively corresponding to an E_{sel} cutoff value of either 1.0 or 0.8 E_h.[29] The smaller E_{sel} value of 0.8 E_h is intended to afford further computational savings. Here, the selected CVS-QE12 Hamiltonian corresponding to the standard cutoff of E_{sel} = 1 E_h was then additionally optimized in terms of a cutoff of E_{sel} = 0.8 E_h. This parameter set employs the appropriate $p_C^{(v,v)}$ and p_X parameters from the corresponding valence Hamiltonian. In terms of computational savings, this approach may nevertheless be superseded by the approximate methods *p*-DFT/MRCI and DFT/MRCI(2).[45, 46]

Table B.8: Final set of DFT/MRCI parameters for the CVS-QE12 Hamiltonian.

$p_C^{(v,v)}$	$p_C^{(c,v)}$	p_X	d_1	d_2	d_3	E_{sel}
0.425815	0.564040	0.254799	0.529467	2.302585	12.0	1.0
0.419210	0.560186	0.257655	0.545822	2.302585	12.0	0.8

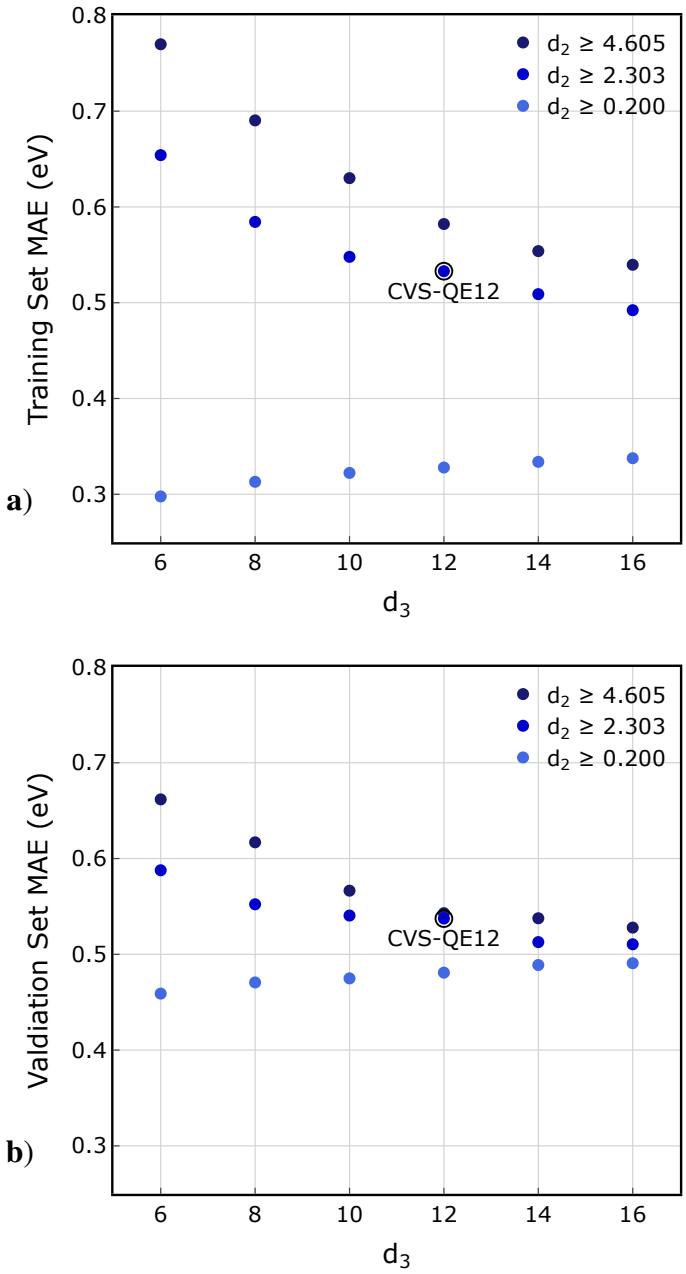


Figure B.1: Training (a) and validation (b) set MAE of CVS-QE*n* Hamiltonian parameter sets. The series in the plots are grouped by the d_2 constraint applied, and errors are plotted as a function of the exponent d_3.

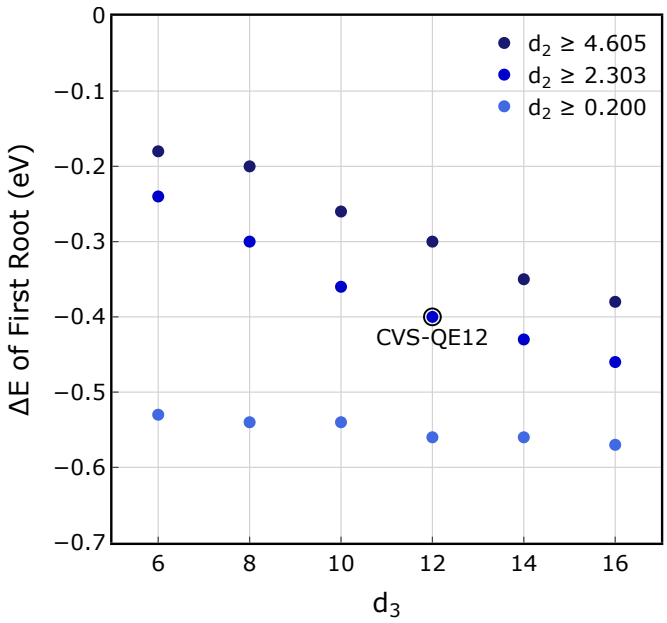


Figure B.2: The ΔE for the first excited-state root in carbon K-shell excitations of thymine from $N = 1$ to $N = 40$ roots. The ΔE values are tabulated corresponding to each CVS-QEn Hamiltonian parameter set in Table B.4. The series in the plot are grouped by the d_2 constraint applied, and energy differences are plotted as a function of the exponent d_3.

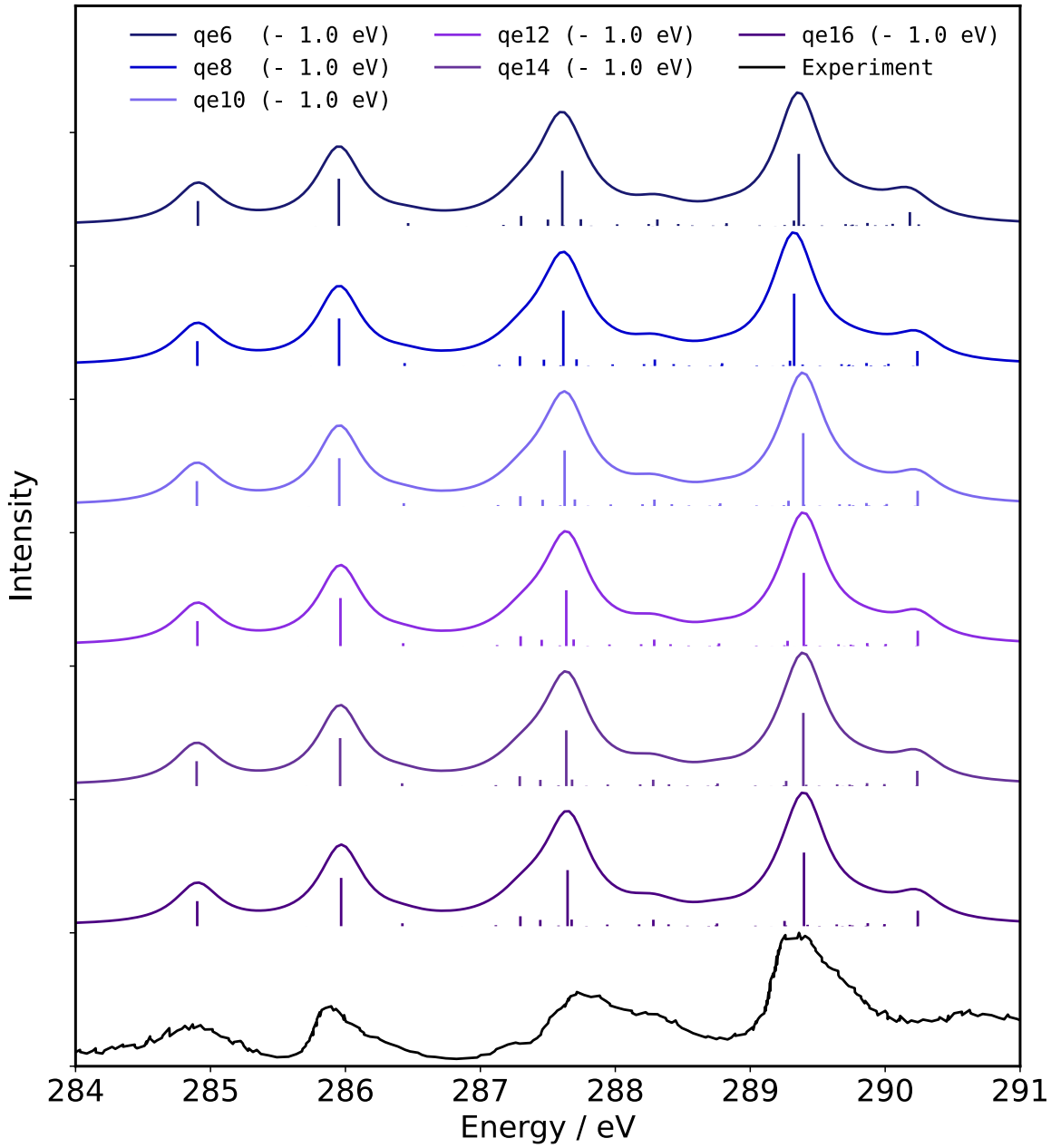


Figure B.3: Carbon K-shell NEXAFS spectrum of thymine simulated with CVS-QE*n* Hamiltonian parameter sets for the $d_2 \geq 0.200$ constraint. Experimental spectrum (black) digitized from source.[257]

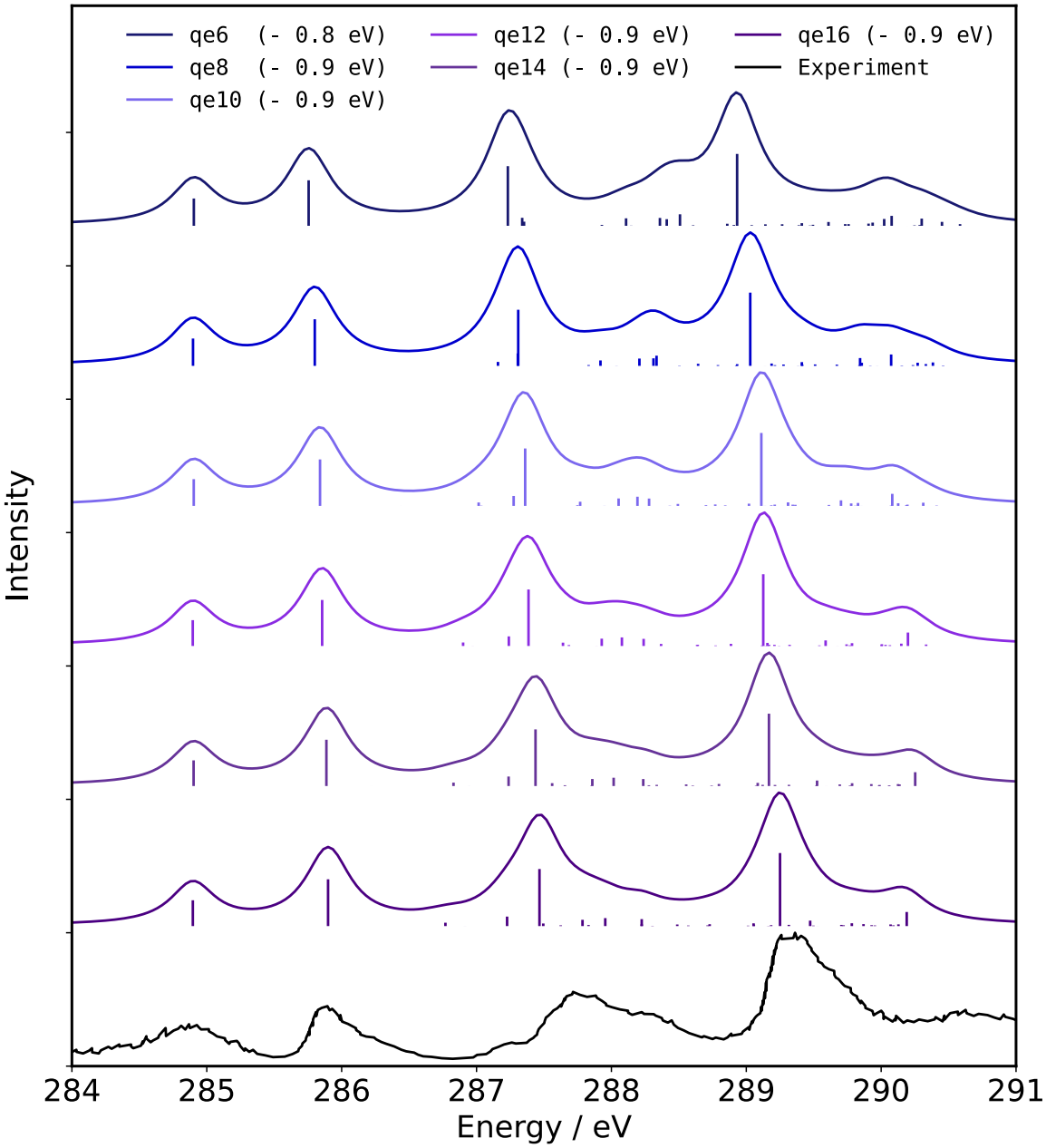


Figure B.4: Carbon K-shell NEXAFS spectrum of thymine simulated with CVS-QE*n* Hamiltonian parameter sets for the $d_2 \geq 2.303$ constraint. Experimental spectrum (black) digitized from source.[257]

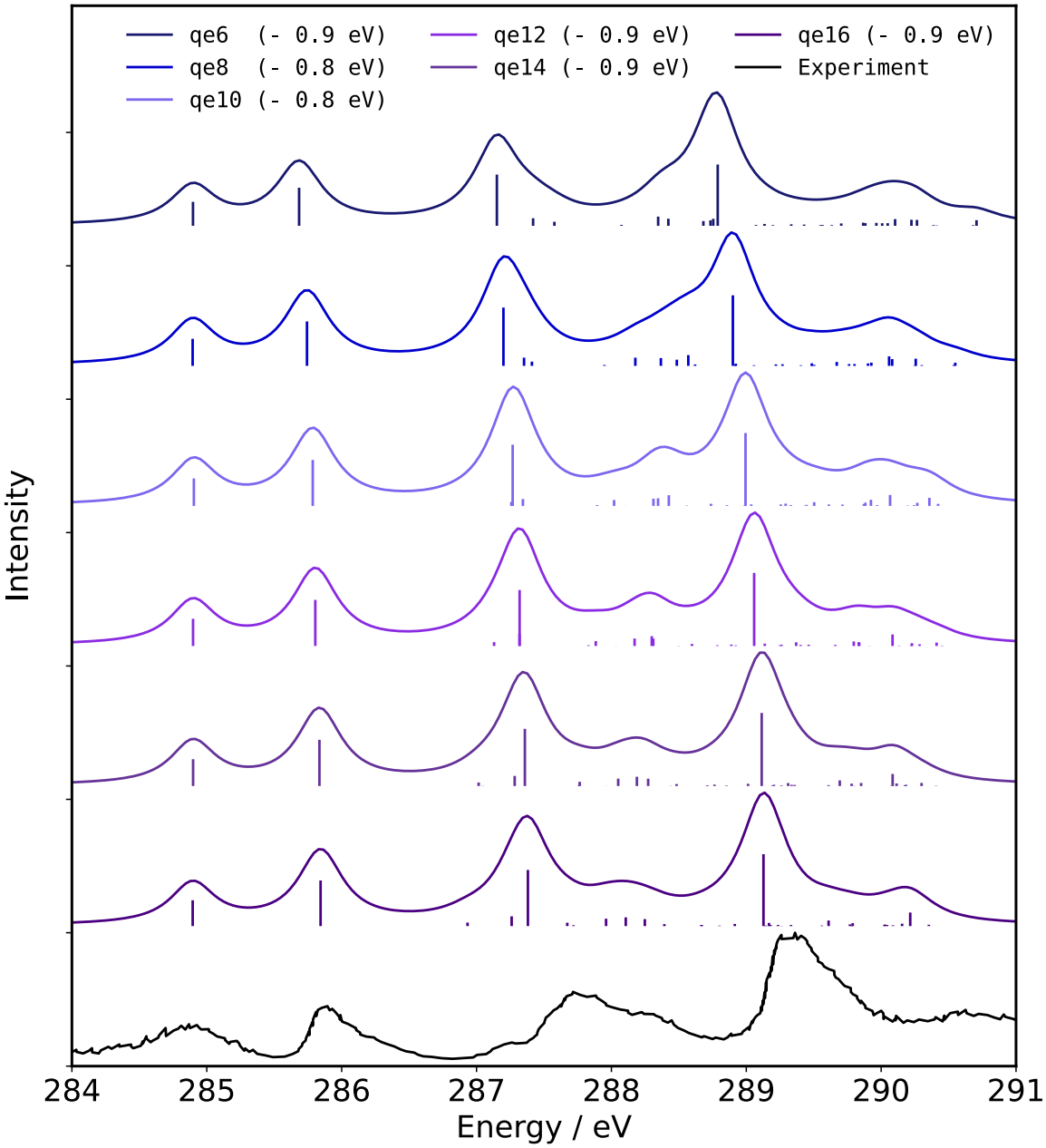


Figure B.5: Carbon K-shell NEXAFS spectrum of thymine simulated with CVS-QE*n* Hamiltonian parameter sets for the $d_2 \geq 4.605$ constraint. Experimental spectrum (black) digitized from source.[257]

B.3 SIMULATION OF NEXAFS AND X-RAY PHOTOELECTRON SPECTRA

In this section, additional data are provided for each of the representative applications presented in the main text. In these applications, the carbon XPS of ethyl trifluoroacetate molecule, and the NEXAFS spectra for two nucleobases, cytosine and thymine, are computed at the DFT/MRCI level of theory and compared to experimental spectra. To compare DFT/MRCI to its comparative methods, *p*-DFT/MRCI and DFT/MRCI(2), these spectra are recomputed using the CVS-QE12 Hamiltonian. Additional details of the calculations required to simulate the spectra of cytosine are also provided.

For thymine, the ground and excited-state oxygen K-shell NEXAFS spectra of thymine are computed using the three DFT/MRCI methods in the main text (using the aug-cc-pCVTZ/aug-cc-pVTZ basis sets). Here, a similar example is displayed for the set of ground-state NEXAFS spectra, for the carbon, nitrogen, and oxygen atoms. These are included here as an additional illustrative example. The C, N, and O XAS are simulated for each nucleobase using the aug-cc-pCVDZ basis set for all non-hydrogen atoms, and the aug-cc-pVDZ basis set for hydrogen atoms.

In each figure, the vertical bars of the computed spectra represent the oscillator strengths, which are normalized to unity, and the lines represent the Lorentzian convolution, with a FWHM of 0.60 eV. The simulated spectra are compared against the corresponding experimental spectrum, in black, which are digitized from their reference publications.[228, 229, 257] The computed spectra are appropriately red- or blue-shifted, as indicated in the legend of each plot. The

peaks are typically shifted to align the first 1s $\rightarrow \pi^*$ transition to the first experimental peak.

Similarly, the carbon K-edge XPS of ethyl trifluoroacetate is simulated using the DFT/MRCI methods with CVS-QE12 and the aug-cc-pCVTZ basis set on the carbon, oxygen, and fluorine atoms, and the aug-cc-pVTZ basis set on the hydrogen atoms. Using the ionization probabilities associated with each state, the spectrum is simulated using a Lorentzian convolution with a FWHM of 0.60 eV. The experimental XPS is once again displayed for reference.[226]

B.3.1 ETHYL TRIFLUOROACETATE XPS

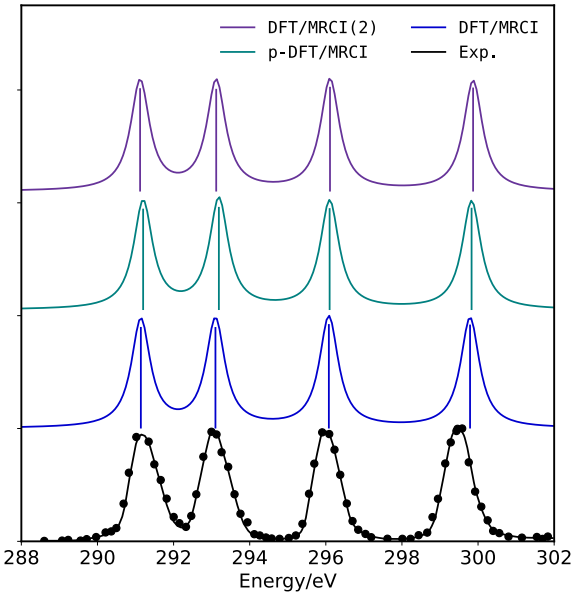


Figure B.6: Carbon XPS of ethyl trifluoroacetate simulated with DFT/MRCI(2) (purple), *p*-DFT/MRCI (teal), and DFT/MRCI (blue), using the CVS-QE12 Hamiltonian, with comparison to experimental gas-phase X-Ray photoelectron spectrum (black). The simulated spectra consider only the anti-anti-conformer. experimental data, shown with a line of best fit, are digitized from their source publication.[226]

In the main text, the CVS-QE12 and CVS-R2017 Hamiltonians are directly compared by

simulating the carbon XPS of ethyl trifluoroacetate. In Figure B.6, this spectrum is additionally simulated with the approximate methods *p*-DFT/MRCI and DFT/MRCI(2), for comparison. These results all employ the CVS-QE12 Hamiltonian, demonstrating that the approximative variants to the parent DFT/MRCI formulation do not degrade its good performance.

B.3.2 CYTOSINE NEXAFS

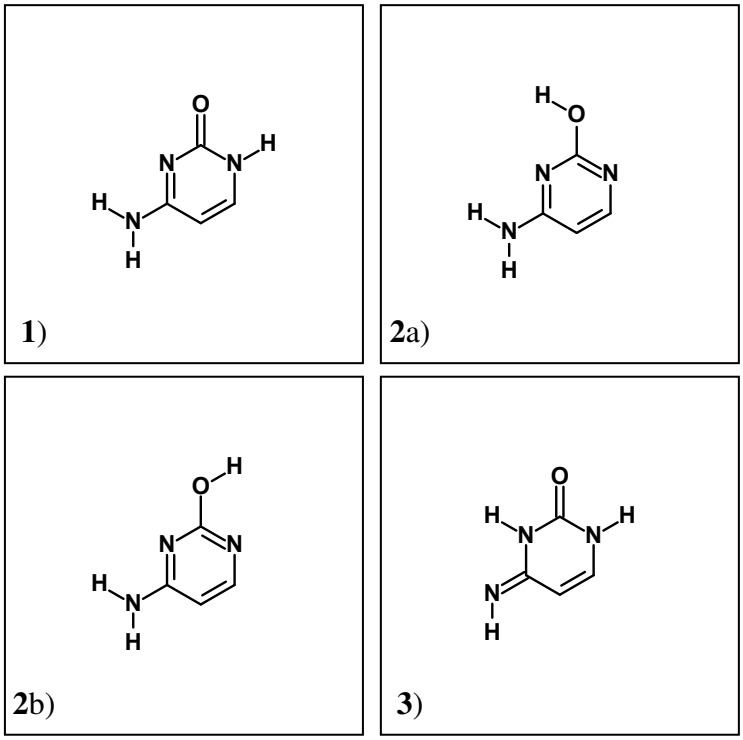


Figure B.7: Four tautomers of cytosine used to simulate the NEXAFS spectra. The structures and the naming convention used here are those employed by Feyer et al.[228, 229]

The NEXAFS spectra for the carbon, nitrogen, and oxygen atoms of cytosine are simulated with CVS-QE12. In the main text, these spectra are compared to those simulated by the CVS-

ADC(2)-x and fc-CVS-EOM-CCSD methods. For each spectrum, four structures of cytosine must be considered; the method for obtaining the experimental spectra required an evaporation temperature of 450K to obtain the gas-phase spectral data, and so the four lowest energy structures of cytosine were considered by Feyer et al. in their theoretical analysis.[228, 229] The same four structures are considered here. Their spectra are combined and weighted relative to one another using the BPRs computed for a temperature of 450K by Trygubenko et al.,[242] as cited by Feyer et al.[228] The four structures of cytosine, following the naming convention employed by Feyer et al., are illustrated in Figure B.7. This procedure is also used to simulate these spectra with the approximate DFT/MRCI methods. A comparison between these methods for the resultant spectra is shown in Figure B.8.

B.3.3 THYMINE NEXAFS

In the main text, the ground and excited-state oxygen K-shell NEXAFS spectra of thymine are simulated using the ground and excited-state geometries. These results include a comparison between DFT/MRCI and its approximative variants, *p*-DFT/MRCI, and DFT/MRCI(2); however, simulation of the complete NEXAFS spectra may better illustrate the agreement between each of the DFT/MRCI methods. Below, the ground-state NEXAFS spectra are computed for each of the carbon, nitrogen, and oxygen atoms using each of the DFT/MRCI methods with the CVS-QE12 Hamiltonian (see Fig. B.9). As was the case for the previous applications, these examples show remarkable agreement across the three methods.

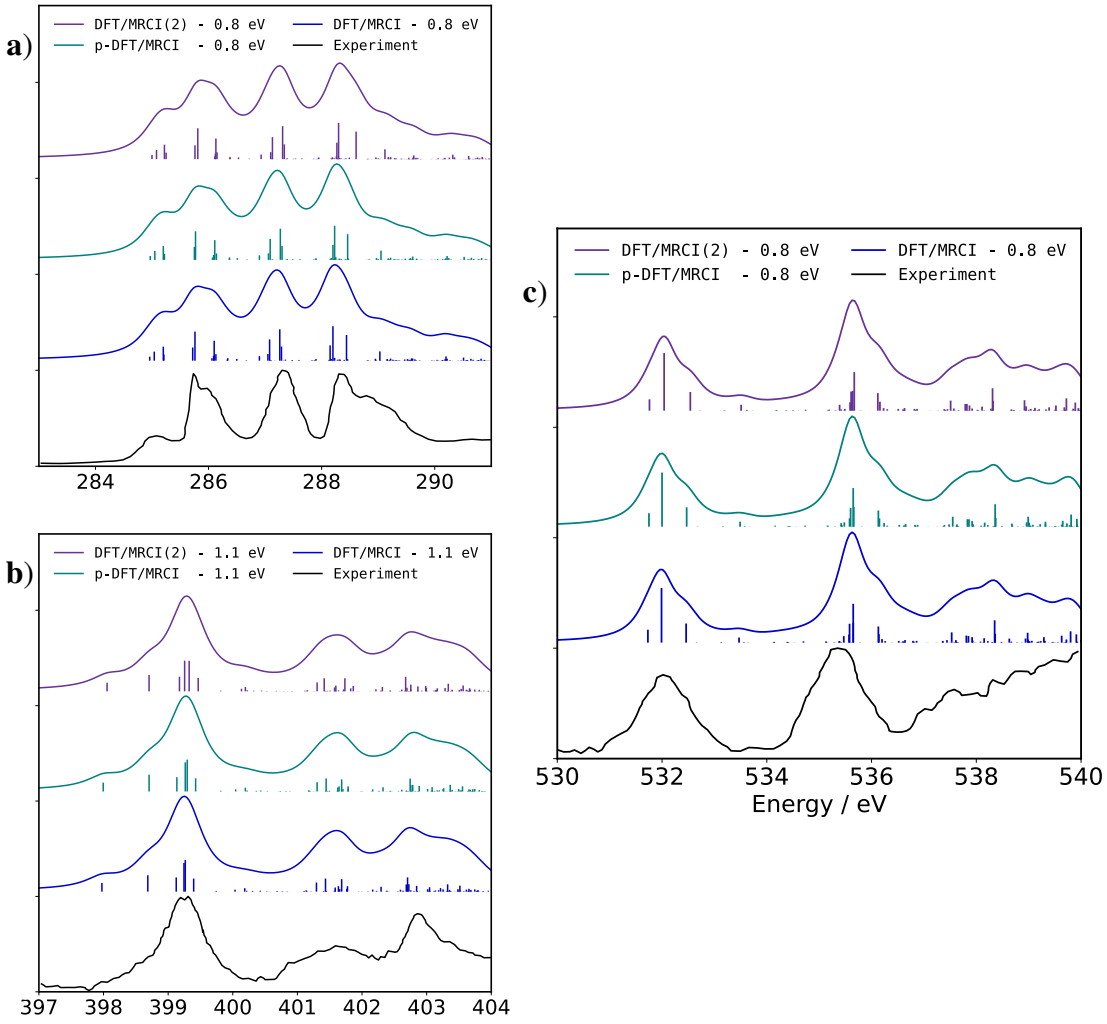


Figure B.8: NEXAFS spectrum of carbon (**a**), nitrogen (**b**), and oxygen (**c**) K-shells of cytosine simulated with DFT/MRCI(2) (purple), *p*-DFT/MRCI (teal), and DFT/MRCI (blue), with the CVS-QE12 Hamiltonian. Experimental spectra (black) digitized from source.[228, 229]

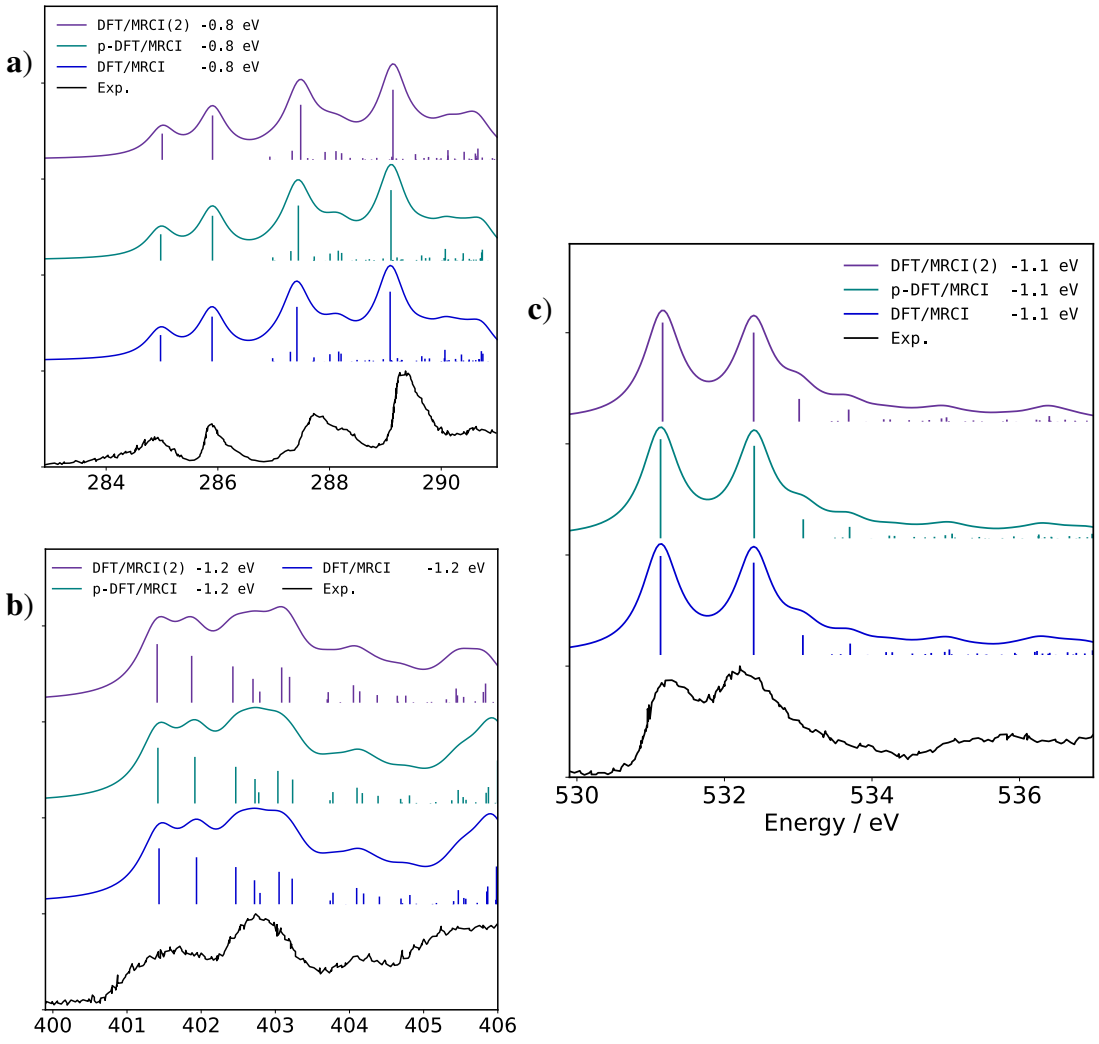


Figure B.9: NEXAFS spectrum of carbon (**a**), nitrogen (**b**), and oxygen (**c**) K-shells of thymine Simulated with DFT/MRCI(2) (purple), *p*-DFT/MRCI (teal), and DFT/MRCI (blue), with the CVS-QE12 Hamiltonian. Experimental spectra (black) digitized from source.[257]

B.4 MOLECULAR GEOMETRIES

B.4.1 TRAINING & VALIDATION DATA SETS

Table B.9: Training set molecules.

Acetylene (Å) from Ref. [20]			
C	0.0000000	0.0000000	0.6035179
C	0.0000000	0.0000000	-0.6035179
H	0.0000000	0.0000000	1.6616643
H	0.0000000	0.0000000	-1.6616643
Ethylene (Å) from Ref. [20]			
C	0.0000000	0.6669040	0.0000000
C	0.0000000	-0.6669040	0.0000000
H	0.0000000	1.2295220	0.9222906
H	0.0000000	-1.2295220	0.9222906
H	0.0000000	1.2295220	-0.9222906
H	0.0000000	-1.2295220	-0.9222906
Methane (a.u.) from Ref. [202]			
C	0.0000000	0.0000000	0.0000000
H	0.0000000	-1.6802672	1.1881283
H	0.0000000	1.6802672	1.1881283
H	1.6802672	0.0000000	-1.1881283
H	-1.6802672	0.0000000	-1.1881283
Water (a.u.) from Ref. [202]			
O	0.0000000	0.0000000	-0.1239094
H	0.0000000	1.4299373	0.9832658
H	0.0000000	-1.4299373	0.9832658
Ammonia (a.u.) from Ref. [202]			
N	0.0000000	0.0000000	-0.1277339
H	0.0000000	-1.7699235	0.5915931
H	1.5327987	0.8849617	0.5915931
H	-1.5327987	0.8849617	0.5915931

Hydrogen Fluoride (a.u.) from Ref. [202]			
F	0.0000000	0.0000000	0.0873178
H	0.0000000	0.0000000	-1.6460179
Fluoroamine (a.u.) from Ref. [202]			
N	-0.1327913	-1.3902255	0.0000000
H	0.7312334	-2.0796879	-1.5632639
F	0.0202951	1.2453338	0.0000000
H	0.7312334	-2.0796879	1.5632639
Azomethine (a.u.) from Ref. [202]			
H	-1.9941130	1.6102925	0.0000000
N	-1.2098856	-0.1539692	0.0000000
C	1.1968660	0.0335860	0.0000000
H	2.3196128	-1.6891573	0.0000000
H	2.2341950	1.8182637	0.0000000
Hypofluorous Acid (a.u.) from Ref. [202]			
O	-1.4228157	-0.1186463	0.0000000
F	1.2907140	0.0107481	0.0000000
H	-1.7499948	1.6803922	0.0000000

Table B.10: Validation set molecules.

Carbon Monoxide (a.u.) from Ref. [202]			
O	0.0000000	0.0000000	0.9139730
C	0.0000000	0.0000000	-1.2182430
Dinitrogen (Å) from Ref. [202]			
N	0.0000000	0.0000000	0.5503900
N	0.0000000	0.0000000	-0.5503900
Formaldehyde (a.u.) from Ref. [202]			
C	0.0000000	0.0000000	1.1442483
O	0.0000000	0.0000000	-1.1415137
H	0.0000000	1.7698527	2.2461407
H	0.0000000	-1.7698527	2.2461407
Hydrogen Cyanide (Å) from Ref. [202]			
H	0.0000000	0.0000000	1.0680000
C	0.0000000	0.0000000	0.0000000
N	0.0000000	0.0000000	-1.1560000
Hydrogen Isocyanide (a.u.) from Ref. [224]			
N	0.0000000	0.0000000	0.431366
C	0.0000000	0.0000000	-0.740682
H	0.0000000	0.0000000	1.424529
Nitroxyl (a.u.) from Ref. [202]			
N	-1.1543258	-0.1614981	0.0000000
O	1.1319819	0.0357051	0.0000000
H	-1.9267677	1.6772444	0.0000000
Fluoromethane (a.u) from Ref. [202]			
F	1.2127043	0.0000000	0.0000000
C	-1.3975536	0.0000000	0.0000000
H	-2.0733764	0.9730934	-1.6854471
H	-2.0733764	-1.9461867	0.0000000
H	-2.0733764	0.9730934	1.6854471

Ketene (Å) from Ref. [20]			
C	0.0000000	0.0000000	-1.2954795
C	0.0000000	0.0000000	0.0185135
O	0.0000000	0.0000000	1.1835785
H	0.0000000	0.9389301	-1.8188138
H	0.0000000	-0.9389301	-1.8188138

B.4.2 REPRESENTATIVE APPLICATIONS

Table B.11: Ethyl trifluoroacetate (Å), aug-cc-pVTZ/MP2.

anti-anti Conformer			
C	1.357858	-0.218256	0.000000
F	2.400882	0.558853	0.000000
C	0.058629	0.610200	0.000000
O	0.060562	1.784722	0.000000
O	-0.976264	-0.176763	0.000000
C	-2.275883	0.425209	0.000000
C	-3.296073	-0.686593	0.000000
H	-4.291642	-0.259911	0.000000
H	-3.188619	-1.308775	-0.878808
H	-3.188619	-1.308775	0.878808
H	-2.360919	1.051580	0.875722
H	-2.360919	1.051580	-0.875722
F	1.409583	-0.984896	-1.062925
F	1.409583	-0.984896	1.062925

Table B.12: Cytosine (Å), aug-cc-pVTZ/CCSD.

Tautomer 1			
C	1.047109	-1.182624	-0.000096
C	-0.196676	-1.705671	0.000087
N	-1.276704	-0.887554	0.000259
C	-1.184372	0.517864	0.000736
N	0.081589	1.051680	0.000183
C	1.122714	0.257421	-0.000054
N	2.343661	0.839163	-0.002104
H	2.392646	1.840837	0.003224
H	3.183793	0.297212	0.005501
O	-2.208784	1.166824	-0.000610
H	-2.211634	-1.256381	0.000120
H	-0.392256	-2.768189	-0.000093
H	1.926396	-1.805693	-0.000714
Tautomer 2a			
C	1.133587	-0.260295	0.000000
C	1.108261	1.146011	0.000000
C	-0.138836	1.727919	0.000000
N	-1.289777	1.038757	0.000000
C	-1.130229	-0.273068	0.000000
N	0.005380	-0.972181	0.000000
O	-2.261946	-0.993106	0.000000
H	-1.994993	-1.918045	0.000000
H	-0.236827	2.806628	0.000000
H	2.013202	1.734006	0.000000
N	2.293962	-0.960228	0.000000
H	2.255186	-1.961049	0.000000
H	3.179851	-0.497138	0.000000

Tautomer 2b			
C	1.149196	-0.248218	0.000000
C	1.088399	1.159870	0.000000
C	-0.171036	1.709026	0.000000
N	-1.297868	0.979208	0.000000
C	-1.101116	-0.333460	0.000000
N	0.046917	-0.998498	0.000000
O	-2.196184	-1.107557	0.000000
H	-2.946519	-0.503487	0.000000
H	-0.302682	2.783849	0.000000
H	1.977788	1.771128	0.000000
N	2.328949	-0.913094	0.000000
H	2.314216	-1.914961	0.000000
H	3.202145	-0.426571	0.000000
Tautomer 3			
C	1.266040	-0.330491	0.000000
C	1.165048	1.126942	0.000000
C	-0.042971	1.707812	0.000000
N	-1.205514	0.967313	0.000000
C	-1.219279	-0.415212	0.000000
N	0.029648	-0.982883	0.000000
H	0.051052	-1.990106	0.000000
O	-2.247168	-1.055814	0.000000
H	-2.107202	1.407382	0.000000
H	-0.174573	2.779580	0.000000
H	2.066156	1.718204	0.000000
N	2.315358	-1.056510	0.000000
H	3.144112	-0.471187	0.000000

Table B.13: Thymine (Å), aug-cc-pVDZ/CCSD(T) from Ref. [4].

Franck–Condon point minimum			
C	-1.6332	0.0426	-0.0001
C	0.7570	0.8384	-0.0001
C	1.1979	-0.5713	0.0000
C	0.2400	-1.5432	0.0000
C	2.6817	-0.8510	0.0002
N	-0.6415	1.0215	-0.0007
N	-1.1219	-1.2509	-0.0002
O	-2.8353	0.2919	0.0005
O	1.5118	1.8105	0.0003
H	-0.9734	1.9835	-0.0002
H	0.4869	-2.6087	0.0000
H	-1.8153	-1.9901	0.0005
H	3.1589	-0.4028	-0.8884
H	3.1587	-0.4025	0.8887
H	2.8721	-1.9376	0.0005

Table B.14: Thymine (Å), aug-cc-pVDZ/CCSD from Ref. [4].

$n\pi^*$ minimum			
C	1.6756	0.0108	0.0230
C	-0.6393	0.8270	-0.0428
C	-1.1709	-0.4819	-0.0168
C	-0.2637	-1.5185	0.0106
C	-2.6680	-0.6951	-0.0181
N	0.7453	1.0515	0.1200
N	1.1227	-1.2492	0.0249
O	2.8776	0.2283	-0.0350
O	-1.3720	1.9491	0.1211
H	1.1236	1.9641	-0.1066
H	-0.5521	-2.5678	-0.0154
H	1.7941	-2.0025	-0.0250
H	-3.1289	-0.2542	0.8817
H	-3.1341	-0.2281	-0.9015
H	-2.8981	-1.7710	-0.0329
$\pi\pi^*$ saddle-point			
N	0.6451	-1.0681	-0.0013
H	0.9407	-2.0373	-0.0009
C	1.6033	-0.1081	0.0001
O	2.8197	-0.2668	0.0006
N	1.1090	1.2319	0.0005
H	1.8635	1.9158	0.0010
C	-0.1892	1.6224	-0.0010
H	-0.4105	2.6884	-0.0014
C	-0.7824	-0.8153	-0.0001
O	-1.5291	-1.8220	0.0005
C	-1.1881	0.5501	-0.0002
C	-2.6474	0.8865	0.0006
H	-2.8068	1.9771	-0.0001
H	-3.1465	0.4513	0.8852
H	-3.1475	0.4502	-0.8830

REFERENCES

1. Cheng, Y.-C. & Fleming, G. R. Dynamics of light harvesting in photosynthesis. *Annu. Rev. Phys. Chem.* **60,** 241–262 (2009).
2. Polli, D. *et al.* Conical intersection dynamics of the primary photoisomerization event in vision. *Nat.* **467,** 440–443 (2010).
3. McFarland, B. *et al.* Ultrafast X-ray Auger probing of photoexcited molecular dynamics. *Nat. Comm.* **5,** 4235 (2014).
4. Wolf, T. *et al.* Probing ultrafast $\pi\pi^*/n\pi^*$ internal conversion in organic chromophores via K-edge resonant absorption. *Nat. Comm.* **8,** 29 (2017).
5. Wolf, T. J. *et al.* Observation of ultrafast intersystem crossing in thymine by extreme ultraviolet time-resolved photoelectron spectroscopy. *J. Phys. Chem. A* **123,** 6897–6903 (2019).
6. Wolf, T. J. *et al.* Transient resonant Auger–Meitner spectra of photoexcited thymine. *Faraday Discuss.* **228,** 555–570 (2021).
7. Pham, T. C., Nguyen, V.-N., Choi, Y., Lee, S. & Yoon, J. Recent strategies to develop innovative photosensitizers for enhanced photodynamic therapy. *Chem. Rev.* **121,** 13454–13619 (2021).
8. Gao, Q. *et al.* Single-molecule characterization from the perspective of optics, photonics, and optoelectronics: a review. *Adv. Photonics* **6,** 064002–064002 (2024).
9. Irie, M., Fukaminato, T., Matsuda, K. & Kobatake, S. Photochromism of diarylethene molecules and crystals: memories, switches, and actuators. *Chem. Rev.* **114,** 12174–12277 (2014).
10. Brinks, D. *et al.* Ultrafast dynamics of single molecules. *Chem. Soc. Rev.* **43,** 2476–2491 (2014).
11. Berera, R., van Grondelle, R. & Kennis, J. T. Ultrafast transient absorption spectroscopy: principles and application to photosynthetic systems. *Photosynth. Res.* **101,** 105–118 (2009).
12. Schoenlein, R. *et al.* Recent advances in ultrafast X-ray sources. *Philos. Trans. R. Soc., A* **377,** 20180384 (2019).
13. Maiuri, M., Garavelli, M. & Cerullo, G. Ultrafast spectroscopy: State of the art and open challenges. *J. Am. Chem. Soc.* **142,** 3–15 (2019).
14. Chen, L. X., Zhang, X & Shelby, M. Recent advances on ultrafast X-ray spectroscopy in the chemical sciences. *Chem. Sci.* **5,** 4136–4152 (2014).
15. Pellegrini, C. The development of XFELs. *Nat. Rev. Phys.* **2,** 330–331 (2020).
16. Sirohiwal, A., Berraud-Pache, R., Neese, F., Izsák, R. & Pantazis, D. A. Accurate computation of the absorption spectrum of chlorophyll a with pair natural orbital coupled cluster methods. *J. Phys. Chem. B* **124,** 8761–8771 (2020).

17. Korona, T. & Werner, H.-J. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.* **118,** 3006–3019 (2003).

18. Dahlbom, M. G. & Reimers, J. R. Successes and failures of time-dependent density functional theory for the low-lying excited states of chlorophylls. *Mol. Phys.* **103,** 1057–1065 (2005).

19. Ghosh, S., Verma, P., Cramer, C. J., Gagliardi, L. & Truhlar, D. G. Combining wave function methods with density functional theory for excited states. *Chem. Rev.* **118,** 7249–7292 (2018).

20. Loos, P.-F. *et al.* A mountaineering strategy to excited states: Highly accurate reference energies and benchmarks. *J. Chem. Theory Comput.* **14,** 4360–4379 (2018).

21. Loos, P.-F., Boggio-Pasqua, M., Scemama, A., Caffarel, M. & Jacquemin, D. Reference energies for double excitations. *J. Chem. Theory Comput.* **15,** 1939–1956 (2019).

22. Loos, P.-F., Lipparini, F., Boggio-Pasqua, M., Scemama, A. & Jacquemin, D. A mountaineering strategy to excited states: Highly accurate energies and benchmarks for medium sized molecules. *J. Chem. Theory Comput.* **16,** 1711–1741 (2020).

23. Loos, P.-F., Scemama, A., Boggio-Pasqua, M. & Jacquemin, D. Mountaineering strategy to excited states: highly accurate energies and benchmarks for exotic molecules and radicals. *J. Chem. Theory Comput.* **16,** 3720–3736 (2020).

24. Véril, M. *et al.* QUESTDB: A database of highly accurate excitation energies for the electronic structure community. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **11,** e1517 (2021).

25. Loos, P.-F., Comin, M., Blase, X. & Jacquemin, D. Reference energies for intramolecular charge-transfer excitations. *J. Chem. Theory Comput.* **17,** 3666–3686 (2021).

26. Loos, P.-F. & Jacquemin, D. A mountaineering strategy to excited states: Highly accurate energies and benchmarks for bicyclic systems. *J. Phys. Chem. A* **125,** 10174–10188 (2021).

27. Chrayteh, A., Blondel, A., Loos, P.-F. & Jacquemin, D. Mountaineering strategy to excited states: highly accurate oscillator strengths and dipole moments of small molecules. *J. Chem. Theory Comput.* **17,** 416–438 (2020).

28. Loos, P.-F., Scemama, A. & Jacquemin, D. The quest for highly accurate excitation energies: A computational perspective. *J. Phys. Chem. Lett.* **11,** 2374–2383 (2020).

29. Grimme, S. & Waletzke, M. A combination of Kohn–Sham density functional theory and multi-reference configuration interaction methods. *J. Chem. Phys.* **111,** 5645–5655 (1999).

30. Marian, C. M., Heil, A. & Kleinschmidt, M. The DFT/MRCI method. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **9,** e1394 (2019).

31. Parusel, A. B. & Grimme, S. A theoretical study of the excited states of chlorophyll a and pheophytin a. *J. Phys. Chem. B* **104,** 5395–5398 (2000).

32. Parusel, A. B. & Grimme, S. DFT/MRCI calculations on the excited states of porphyrin, hydroporphyrins, tetrazaporphyrins and metalloporphyrins. *J. Porphyrins Phthalocyanines* **5,** 225–232 (2001).

33. Kleinschmidt, M., Tatchen, J. & Marian, C. M. Spin-orbit coupling of DFT/MRCI wavefunctions: Method, test calculations, and application to thiophene. *J. Comput. Chem.* **23,** 824–833 (2002).

34. Marian, C. M. & Gilka, N. Performance of the density functional theory/multireference configuration interaction method on electronic excitation of extended π-systems. *J. Chem. Theory Comput.* **4,** 1501–1515 (2008).

35. Perun, S., Tatchen, J. & Marian, C. M. Singlet and triplet excited states and intersystem crossing in free-base porphyrin: TDDFT and DFT/MRCI study. *ChemPhysChem* **9,** 282–292 (2008).

36. Kleinschmidt, M., Marian, C. M., Waletzke, M. & Grimme, S. Parallel multireference configuration interaction calculations on mini-β-carotenes and β-carotene. *J. Chem. Phys.* **130** (2009).

37. Knecht, S., Marian, C. M., Kongsted, J. & Mennucci, B. On the photophysics of carotenoids: a multireference DFT study of peridinin. *J. Phys. Chem. B* **117,** 13808–13815 (2013).

38. Andreussi, O., Knecht, S., Marian, C. M., Kongsted, J. & Mennucci, B. Carotenoids and light-harvesting: from DFT/MRCI to the Tamm–Dancoff approximation. *J. Chem. Theory Comput.* **11,** 655–666 (2015).

39. Bettinger, H. F., Tönshoff, C., Doerr, M. & Sanchez-Garcia, E. Electronically excited states of higher acenes up to nonacene: A density functional theory/multireference configuration interaction study. *J. Chem. Theory Comput.* **12,** 305–312 (2016).

40. Jovanović, V., Lyskov, I., Kleinschmidt, M. & Marian, C. M. On the performance of DFT/MRCI-R and MR-MP2 in spin–orbit coupling calculations on diatomics and polyatomic organic molecules. *Mol. Phys.* **115,** 109–137 (2017).

41. Heil, A., Kleinschmidt, M. & Marian, C. M. On the performance of DFT/MRCI Hamiltonians for electronic excitations in transition metal complexes: The role of the damping function. *J. Chem. Phys.* **149,** 164106 (2018).

42. Heil, A. & Marian, C. M. DFT/MRCI-R2018 study of the photophysics of the zinc (ii) tripyrrindione radical: non-Kasha emission? *Phys. Chem. Chem. Phys.* **21,** 19857–19867 (2019).

43. Lyskov, I., Kleinschmidt, M. & Marian, C. M. Redesign of the DFT/MRCI Hamiltonian. *J. Chem. Phys.* **144,** 034104 (2016).

44. Heil, A. & Marian, C. M. DFT/MRCI Hamiltonian for odd and even numbers of electrons. *J. Chem. Phys.* **147,** 194104 (2017).

45. Neville, S. P. & Schuurman, M. S. Removing the deadwood from DFT/MRCI wave functions: The p-DFT/MRCI method. *J. Chem. Theory Comput.* **17,** 7657–7665 (2021).

46. Neville, S. P. & Schuurman, M. S. A perturbative approximation to DFT/MRCI: DFT/MRCI (2). *J. Chem. Phys.* **157,** 164103 (2022).

47. Dombrowski, D. R., Schulz, T., Kleinschmidt, M. & Marian, C. M. R2022: A DFT/MRCI Ansatz with Improved Performance for Double Excitations. *J. Phys. Chem. A* **127.** PMID: 36799533, 2011–2025 (2023).

48. Seidu, I. *et al.* The simulation of X-ray absorption spectra from ground and excited electronic states using core-valence separated DFT/MRCI. *J. Chem. Phys.* **151,** 144104 (2019).

49. Koch, W. & Holthausen, M. C. *A Chemist's Guide to Density Functional Theory* (John Wiley & Sons, 2015).

50. Helgaker, T., Jørgensen, P. & Olsen, J. *Density-Functional Theory: A Convex Treatment* (Wiley Blackwell, 2016).

51. Teale, A. M. *et al.* DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science. *Phys. Chem. Chem. Phys.* **24,** 28700–28781 (2022).

52. Slater, J. C. The theory of complex spectra. *Phys. Rev.* **34,** 1293 (1929).

53. Ritz, W. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *Journal für die Reine und Angewandte Mathematik* **135,** 1–61 (1909).

54. Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. U.S.A.* **76,** 6062–6065 (1979).

55. Lieb, E. H. Density functionals for Coulomb systems. *Int. J. Quantum Chem.* **24,** 243–277 (1983).

56. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136,** B864 (1964).

57. Kohn, W & Sham, L. Quantum density oscillations in an inhomogeneous electron gas. *Phys. Rev.* **137,** A1697 (1965).

58. Harris, J. Adiabatic-connection approach to Kohn-Sham theory. *Phys. Rev. A* **29,** 1648 (1984).

59. Levy, M. & Perdew, J. P. Hellmann-Feynman, virial, and scaling requisites for the exact universal density functionals. Shape of the correlation potential and diamagnetic susceptibility for atoms. *Phys. Rev. A* **32,** 2010 (1985).

60. Levy, M. Asymptotic coordinate scaling bound for exchange-correlation energy in density-functional theory. *Int. J. Quantum Chem.* **36,** 617–619 (1989).

61. Levy, M. Density-functional exchange correlation through coordinate scaling in adiabatic connection and correlation hole. *Phys. Rev. A* **43,** 4637 (1991).

62. Wang, Y. A. Coordinate scaling and adiabatic-connection formulation in density-functional theory. *Phys. Rev. A* **56,** 1646 (1997).

63. Levy, M., Yang, W. & Parr, R. G. A new functional with homogeneous coordinate scaling in density functional theory: F[ρ,λ]. *J. Chem. Phys.* **83,** 2334–2336. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/83/5/2334/18954795/2334\_1\_online.pdf` (Sept. 1985).

64. Friesecke, G., Gerolin, A. & Gori-Giorgi, P. The strong-interaction limit of density functional theory, 183–266 (2022).

65. Burke, K. in *Electronic Density Functional Theory: Recent Progress and New Directions* (eds Dobson, J. F., Vignale, G. & Das, M. P.) 19–29 (Springer US, Boston, MA, 1998). ISBN: 978-1-4899-0316-7.

66. Burke, K., Perdew, J. P. & Ernzerhof, M. Why semilocal functionals work: Accuracy of the on-top pair density and importance of system averaging. *J. Chem. Phys.* **109,** 3760–3771 (1998).

67. Von Barth, U. & Hedin, L. A local exchange-correlation potential for the spin polarized case. I. *J. Phys. C: Solid State Phys.* **5,** 1629 (1972).

68. Wigner, E. On the interaction of electrons in metals. *Phys. Rev.* **46,** 1002 (1934).

69. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38,** 3098 (1988).

70. Slater, J. C. A simplification of the Hartree-Fock method. *Phys. Rev.* **81,** 385 (1951).

71. Janesko, B. G., Proynov, E., Kong, J., Scalmani, G. & Frisch, M. J. Practical density functionals beyond the overdelocalization–underbinding zero-sum game. *J. Phys. Chem. Lett.* **8,** 4314–4318 (2017).

72. Janesko, B. G. Replacing hybrid density functional theory: motivation and recent advances. *Chem. Soc. Rev.* **50,** 8470–8495 (2021).

73. Ruzsinszky, A., Perdew, J. P., Csonka, G. I., Vydrov, O. A. & Scuseria, G. E. Spurious fractional charge on dissociated atoms: Pervasive and resilient self-interaction error of common density functionals. *J. Chem. Phys.* **125** (2006).

74. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23,** 5048 (1981).

75. Perdew, J. P., Ruzsinszky, A., Sun, J. & Pederson, M. R. in *Advances In Atomic, Molecular, and Optical Physics* 1–14 (Elsevier, 2015).

76. Cremer, D. Density functional theory: coverage of dynamic and non-dynamic electron correlation effects. *Mol. Phys.* **99,** 1899–1940 (2001).

77. Yabana, K. & Bertsch, G. Time-dependent local-density approximation in real time. *Phys. Rev. B* **54,** 4484 (1996).

78. Runge, E. & Gross, E. K. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **52,** 997 (1984).

79. Casida, M. E. in *Recent Advances In Density Functional Methods: (Part I)* 155–192 (World Scientific, 1995).

80. Bauernschmitt, R. & Ahlrichs, R. Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem. Phys. Lett.* **256,** 454–464 (1996).

81. Zhu, H., Wang, J., Wang, F., Feng, E. & Sheng, X. Linear and quadratic response TDDFT methods for the excited-state absorption in oligofluorenes. *Chem. Phys. Lett.* **785,** 139150 (2021).

82. Lopata, K., Van Kuiken, B. E., Khalil, M. & Govind, N. Linear-response and real-time time-dependent density functional theory studies of core-level near-edge x-ray absorption. *J. Chem. Theory Comput.* **8,** 3284–3292 (2012).

83. Tussupbayev, S., Govind, N., Lopata, K. & Cramer, C. J. Comparison of real-time and linear-response time-dependent density functional theories for molecular chromophores ranging from sparse to high densities of states. *J. Chem. Theory Comput.* **11,** 1102–1109 (2015).

84. Li, X., Govind, N., Isborn, C., DePrince III, A. E. & Lopata, K. Real-time time-dependent electronic structure theory. *Chem. Rev.* **120,** 9951–9993 (2020).

85. Casida, M. E. Time-dependent density-functional theory for molecules and molecular solids. *J. Mol. Struct.: THEOCHEM* **914,** 3–18 (2009).

86. Hirata, S. & Head-Gordon, M. Time-dependent density functional theory within the Tamm–Dancoff approximation. *Chem. Phys. Lett.* **314,** 291–299 (1999).

87. Dreuw, A., Weisman, J. L. & Head-Gordon, M. Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange. *J. Chem. Phys.* **119,** 2943–2946 (2003).

88. Dreuw, A. & Head-Gordon, M. Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorin- bacteriochlorin and bacteriochlorophyll- spheroidene complexes. *J. Am. Chem. Soc.* **126,** 4007–4016 (2004).

89. Norman, P. & Dreuw, A. Simulating X-ray spectroscopies and calculating core-excited states of molecules. *Chem. Rev.* **118,** 7208–7248 (2018).

90. Carter-Fenk, K., Cunha, L. A., Arias-Martinez, J. E. & Head-Gordon, M. Electron-affinity time-dependent density functional theory: Formalism and applications to core-excited states. *J. Phys. Chem. Lett.* **13,** 9664–9672 (2022).

91. Imamura, Y. & Nakai, H. Analysis of self-interaction correction for describing core excited states. *Int. J. Quantum Chem.* **107,** 23–29 (2007).

92. Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. *Math. Proc. Cambridge Philos. Soc.* **24,** 89–110 (1928).

93. Hartree, D. R. The wave mechanics of an atom with a non-coulomb central field. Part II. Some results and discussion. *Math. Proc. Cambridge Philos. Soc.* **24,** 111–132 (1928).

94. Fock, V. A. in *Selected Works - V.A. Fock: Quantum Mechanics and Quantum Field Theory* (eds Faddeev, L. D., Khalfin, L. A. & Komarov, I. V.) 137–163 (Chapman & Hall/CRC Press, 2004).

95. Slater, J. A generalized self-consistent field method. *Phys. Rev.* **91,** 528 (1953).

96. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* ISBN: 9780486691862 (Dover Publications, 1996).

97. Wetmore, R. W. & Segal, G. A. Efficient generation of configuration interaction matrix elements. *Chem. Phys. Lett.* **36,** 478–483 (1975).

98. Segal, G. A. & Wetmore, R. W. An efficient method for large scale configuration interaction calculations. *Chem. Phys. Lett.* **32,** 556–560 (1975).

99. Segal, G. A., Wetmore, R. W. & Wolf, K. Efficient methods for configuration interaction calculations. *Chem. Phys.* **30,** 269–297. ISSN: 0301-0104 (1978).

100. Buenker, R. J. & Peyerimhoff, S. D. Individualized configuration selection in CI calculations with subsequent energy extrapolation. *Theor. Chim. Acta* **35,** 33–58 (1974).

101. Buenker, R. J., Peyerimhoff, S. D. & Butscher, W. Applicability of the multi-reference double-excitation CI (MRD-CI) method to the calculation of electronic wavefunctions and comparison with related techniques. *Mol. Phys.* **35,** 771–791 (1978).

102. Werner, H.-J. & Knowles, P. J. An efficient internally contracted multiconfiguration–reference configuration interaction method. *J. Chem. Phys.* **89,** 5803–5814 (1988).

103. Plasser, F. & Lischka, H. in *Quantum Chemistry and Dynamics of Excited States: Methods and Applications* (eds González, L. & Lindh, R.) 277–297 (Wiley Online Library, 2020).

104. Szalay, P. G., Muller, T., Gidofalvi, G., Lischka, H. & Shepard, R. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.* **112,** 108–181 (2012).

105. Lischka, H. *et al.* Multireference approaches for excited states of molecules. *Chem. Rev.* **118,** 7293–7361 (2018).

106. Grimme, S. Density functional calculations with configuration interaction for the excited states of molecules. *Chem. Phys. Lett.* **259,** 128–137 (1996).

107. Costain, T. S., Ogden, V., Neville, S. P. & Schuurman, M. S. A DFT/MRCI Hamiltonian parameterized using only ab initio data: I. valence excited states. *J. Chem. Phys.* **160,** 224106. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/5.0210897/19985974/224106\_1\_5.0210897.pdf` (June 2024).

108. Costain, T. S., Rolston, J. B., Neville, S. P. & Schuurman, M. S. A DFT/MRCI Hamiltonian parameterized using only ab initio data. II. Core-excited states. *J. Chem. Phys.* **161** (2024).

109. Zinchenko, K. S. *et al.* Sub-7-femtosecond conical-intersection dynamics probed at the carbon K-edge. *Sci.* **371,** 489–494. eprint: `https://www.science.org/doi/pdf/10.1126/science.abf1656` (2021).

110. Avila Ferrer, F. J., Cerezo, J., Stendardo, E., Improta, R. & Santoro, F. Insights for an Accurate Comparison of Computational Data to Experimental Absorption and Emission Spectra: Beyond the Vertical Transition Approximation. *J. Chem. Theory Comput.* **9.** PMID: 26583553, 2072–2082 (2013).

111. Lasorne, B. *et al.* Vertical transition energies vs. absorption maxima: Illustration with the UV absorption spectrum of ethylene. *Spectrochim. Acta, Part A* **119.** Frontiers in molecular vibrational calculations and computational spectroscopy, 52–58. ISSN: 1386-1425 (2014).

112. Bai, S., Mansour, R., Stojanović, L., Toldo, J. M. & Barbatti, M. On the origin of the shift between vertical excitation and band maximum in molecular photoabsorption. *J. Mol. Model.* **26,** 107 (2020).

113. Neville, S. P. & Worth, G. A. A reinterpretation of the electronic spectrum of pyrrole: A quantum dynamics study. *J. Chem. Phys.* **140,** 034317. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4861223/15472151/034317\_1\_online.pdf` (Jan. 2014).

114. Neville, S. P., Stolow, A. & Schuurman, M. S. Vacuum ultraviolet excited state dynamics of the smallest ring, cyclopropane. I. A reinterpretation of the electronic spectrum and the effect of intensity borrowing. *J. Chem. Phys.* **149,** 144310. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5044392/15549648/144310\_1\_online.pdf` (Oct. 2018).

115. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98,** 1372–1377 (Jan. 1993).

116. Lyskov, I. *Redesign and Reparameterization of the DFT/MRCI Hamiltonian and its Application to Electronically Excited Linear Polyenes* PhD thesis (Heinrich-Heine-Universität Dusselfdorf, 2016).

117. Olsen, J., Roos, B. O., Jørgensen, P. & Jensen, H. J. A. Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces. *J. Chem. Phys.* **89,** 2185–2192 (1988).

118. Koch, H., Christiansen, O., Jørgensen, P., Sanchez de Merás, A. M. & Helgaker, T. The CC3 model: An iterative coupled cluster approach including connected triples. *J. Chem. Phys.* **106,** 1808–1818. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/106/5/1808/19128978/1808\_1\_online.pdf` (Feb. 1997).

119. Huron, B., Malrieu, J. P. & Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.* **58,** 5745–5759. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/58/12/5745/18885418/5745\_1\_online.pdf` (June 1973).

120. Giner, E., Scemama, A. & Caffarel, M. Using perturbatively selected configuration interaction in quantum Monte Carlo calculations. *Can. J. Chem.* **91,** 879–885 (2013).

121. Giner, E., Scemama, A. & Caffarel, M. Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F2 using selected configuration interaction trial wavefunctions. *J. Chem. Phys.* **142,** 044115. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4905528/15490864/044115\_1\_online.pdf` (Jan. 2015).

122. Garniron, Y. *et al.* Quantum Package 2.0: An Open-Source Determinant-Driven Suite of Programs. *J. Chem. Theory Comput.* **15.** PMID: 31082265, 3591–3609 (2019).

123. Kendall, R. A., Dunning Thom H., J. & Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96,** 6796–6806. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/96/9/6796/18998924/6796\_1\_online.pdf` (May 1992).

124. Woon, D. E. & Dunning Thom H., J. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **100,** 2975–2988. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/100/4/2975/19303178/2975\_1\_online.pdf` (Feb. 1994).

125. Neville, S. & Schuurman, M. *GRaCI: General Reference Configuration Interaction* version 0.1. Oct. 2021.

126. Sun, Q. Libcint: An efficient general integral library for g aussian basis functions. *J. Comput. Chem.* **36,** 1664–1671 (2015).

127. Sun, Q. *et al.* PySCF: the Python-based simulations of chemistry framework. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8,** e1340 (2018).

128. Sun, Q. *et al.* Recent developments in the PySCF program package. *J. Chem. Phys.* **153** (2020).

129. Marques, M. A., Oliveira, M. J. & Burnus, T. LIBXC: A library of exchange and correlation functionals for density functional theory. *Comput. Phys. Commun.* **183,** 2272–2281 (2012).

130. Lehtola, S., Steigemann, C., Oliveira, M. J. & Marques, M. A. Recent developments in LIBXC — A comprehensive library of functionals for density functional theory. *SoftwareX* **7,** 1–5 (2018).

131. Koch, H. & Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **93,** 3333–3344. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/93/5/3333/18987229/3333\_1\_online.pdf` (Sept. 1990).

132. Stanton, J. F. & Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **98,** 7029–7039. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/98/9/7029/19054846/7029\_1\_online.pdf` (May 1993).

133. Epifanovsky, E. *et al.* Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **155** (2021).

134. Noga, J. & Bartlett, R. J. The full CCSDT model for molecular electronic structure. *J. Chem. Phys.* **86,** 7041–7050. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/86/12/7041/18964408/7041\_1\_online.pdf` (June 1987).

135. Scuseria, G. E. & Schaefer, H. F. A new implementation of the full CCSDT model for molecular electronic structure. *Chem. Phys. Lett.* **152,** 382–386. ISSN: 0009-2614 (1988).

136. Kállay, M. *et al.* The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **152,** 074107 (2020).

137. Kállay, M. *et al.* MRCC, a quantum chemical program suite written by M. Kallay et. al. *J. Chem. Phys.* **139,** 094105 (2013).

138. Kállay, M. & Gauss, J. Calculation of excited-state properties using general coupled-cluster and configuration-interaction models. *J. Chem. Phys.* **121,** 9257–9269 (2004).

139. Miralles, J., Castell, O., Caballol, R. & Malrieu, J.-P. Specific CI calculation of energy differences: Transition energies and bond energies. *Chem. Phys.* **172,** 33–43. ISSN: 0301-0104 (1993).

140. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **7,** 308–313. ISSN: 0010-4620. eprint: `https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf` (Jan. 1965).

141. Gao, F. & Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* **51,** 259–277. ISSN: 1573-2894 (Jan. 2012).

142. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17,** 261–272 (2020).

143. Besley, N. A., Peach, M. J. & Tozer, D. J. Time-dependent density functional theory calculations of near-edge X-ray absorption fine structure with short-range corrected functionals. *Phys. Chem. Chem. Phys.* **11,** 10350–10358 (2009).

144. Jin, Y. & Bartlett, R. J. Accurate computation of X-ray absorption spectra with ionization potential optimized global hybrid functional. *J. Chem. Phys.* **149,** 064111 (2018).

145. Verma, P. & Bartlett, R. J. Increasing the applicability of density functional theory. IV. Consequences of ionization-potential improved exchange-correlation potentials. *J. Chem. Phys.* **140** (2014).

146. Pernal, K., Podeszwa, R., Patkowski, K. & Szalewicz, K. Dispersionless Density Functional Theory. *Phys. Rev. Lett.* **103,** 263201 (26 Dec. 2009).

147. Wiitala, K. W., Hoye, T. R. & Cramer, C. J. Hybrid Density Functional Methods Empirically Optimized for the Computation of 13C and 1H Chemical Shifts in Chloroform Solution. *J. Chem. Theory Comput.* **2.** PMID: 26633067, 1085–1092 (2006).

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

149. Smith, D. W. & Day, O. W. Extension of Koopmans' theorem. I. Derivation. *J. Chem. Phys.* **62,** 113–114 (1975).

150. Day, O. W., Smith, D. W. & Morrison, R. C. Extension of Koopmans' theorem. II. Accurate ionization energies from correlated wavefunctions for closed-shell atoms. *J. Chem. Phys.* **62,** 115–119 (1975).

151. Morrison, R. C., Day, O. W. & Smith, D. W. An extension of koopmans' theorem III. ionization energies of the open-shell atoms Li and B. *Int. J. Quantum Chem.* **9,** 229–235 (1975).

152. Ellenbogen, J. C., Day, O. W., Smith, D. W. & Morrison, R. C. Extension of Koopmans' theorem. IV. Ionization potentials from correlated wavefunctions for molecular fluorine. *J. Chem. Phys.* **66,** 4795–4801 (1977).

153. Janak, J. F. Proof that $\frac{\partial E}{\partial n_i} = \varepsilon$ in density-functional theory. *Phys. Rev. B* **18,** 7165 (1978).

154. Harbola, M. K. Relationship between the highest occupied Kohn-Sham orbital eigenvalue and ionization energy. *Phys. Rev. B* **60,** 4545 (1999).

155. Verma, P. & Bartlett, R. J. Increasing the applicability of density functional theory. V. X-ray absorption spectra with ionization potential corrected exchange and correlation potentials. *J. Chem. Phys.* **145** (2016).

156. Jin, Y. & Bartlett, R. J. The QTP family of consistent functionals and potentials in Kohn-Sham density functional theory. *J. Chem. Phys.* **145** (2016).

157. Margraf, J. T., Verma, P. & Bartlett, R. J. Ionization potential optimized double-hybrid density functional approximations. *J. Chem. Phys.* **145** (2016).

158. Kirtman, B. Simultaneous calculation of several interacting electronic states by generalized Van Vleck perturbation theory. *J. Chem. Phys.* **75,** 798–808. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/75/2/798/18929991/798\_1\_online.pdf` (July 1981).

159. Neville, S. P., Seidu, I. & Schuurman, M. S. Propagative block diagonalization diabatization of DFT/MRCI electronic states. *J. Chem. Phys.* **152** (2020).

160. Forbes, R. *et al.* Vacuum Ultraviolet Excited State Dynamics of the Smallest Ketone: Acetone. *J. Phys. Chem. Lett.* **12.** PMID: 34464141, 8541–8547 (2021).

161. Chrayteh, A., Blondel, A., Loos, P.-F. & Jacquemin, D. Mountaineering Strategy to Excited States: Highly Accurate Oscillator Strengths and Dipole Moments of Small Molecules. *J. Chem. Theory Comput.* **17.** PMID: 33256412, 416–438 (2021).

162. Gouterman, M. Spectra of porphyrins. *J. Mol. Spectrosc.* **6,** 138–163. ISSN: 0022-2852 (1961).

163. Ceulemans, A., Oldenhof, W., Gorller-Walrand, C. & Vanquickenborne, L. G. Gouterman's "four-orbital" model and the MCD spectra of high-symmetry metalloporphyrins. *J. Am. Chem. Soc.* **108,** 1155–1163 (1986).

164. Sarkar, R., Boggio-Pasqua, M., Loos, P.-F. & Jacquemin, D. Benchmarking TD-DFT and Wave Function Methods for Oscillator Strengths and Excited-State Dipole Moments. *J. Chem. Theory Comput.* **17.** PMID: 33492950, 1117–1132 (2021).

165. Wang, T. Y., Neville, S. P. & Schuurman, M. S. Machine learning seams of conical intersection: A characteristic polynomial approach. *J. Phys. Chem. Lett.* **14,** 7780–7786 (2023).

166. Wang, T. Y., Neville, S. P. & Schuurman, M. S. Minimum Energy Conical Intersection Optimization Using DFT/MRCI (2). *J. Chem. Theory Comput.* (2024).

167. Rankine, C. D. & Penfold, T. J. Progress in the theory of x-ray spectroscopy: From quantum chemistry to machine learning and ultrafast dynamics. *J. Phys. Chem. A* **125,** 4276–4293 (2021).

168. Cederbaum, L. S., Domcke, W & Schirmer, J. Many-body theory of core holes. *Phys. Rev. A* **22,** 206 (1980).

169. Besley, N. A. Density functional theory based methods for the calculation of x-ray spectroscopy. *Acc. Chem. Res.* **53,** 1306–1315 (2020).

170. Stener, M., Fronzoni, G. & de Simone, M. d. Time dependent density functional theory of core electrons excitations. *Chem. Phys. Lett.* **373,** 115–123 (2003).

171. Rehr, J. & Ankudinov, A. Time-dependent density functional theory calculations of X-ray absorption. *Int. J. Quantum Chem.* **95,** 487–492 (2003).

172. Imamura, Y. & Nakai, H. 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. *Chem. Phys. Lett.* **419,** 297–303 (2006).

173. Besley, N. A. & Asmuruf, F. A. Time-dependent density functional theory calculations of the spectroscopy of core electrons. *Phys. Chem. Chem. Phys.* **12,** 12024–12039 (2010).

174. Fernando, R. G., Balhoff, M. C. & Lopata, K. X-ray absorption in insulators with non-Hermitian real-time time-dependent density functional theory. *J. Chem. Theory Comput.* **11,** 646–654 (2015).

175. Kadek, M., Konecny, L., Gao, B., Repisky, M. & Ruud, K. X-ray absorption resonances near L 2, 3-edges from real-time propagation of the Dirac–Kohn–Sham density matrix. *Phys. Chem. Chem. Phys.* **17,** 22566–22570 (2015).

176. Kasper, J. M., Lestrange, P. J., Stetina, T. F. & Li, X. Modeling L2, 3-edge x-ray absorption spectroscopy with real-time exact two-component relativistic time-dependent density functional theory. *J. Chem. Theory Comput.* **14,** 1998–2006 (2018).

177. Evangelista, F. A., Shushkov, P. & Tully, J. C. Orthogonality constrained density functional theory for electronic excited states. *J. Phys. Chem. A* **117,** 7378–7392 (2013).

178. Derricotte, W. D. & Evangelista, F. A. Simulation of X-ray absorption spectra with orthogonality constrained density functional theory. *Phys. Chem. Chem. Phys.* **17,** 14360–14374 (2015).

179. Hait, D. & Head-Gordon, M. 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. *J. Phys. Chem. Lett.* **11,** 775–786 (2020).

180. Hait, D. & Head-Gordon, M. Orbital optimized density functional theory for electronic excited states. *J. Phys. Chem. Lett.* **12,** 4517–4529 (2021).

181. Hait, D., Oosterbaan, K. J., Carter-Fenk, K. & Head-Gordon, M. Computing x-ray absorption spectra from linear-response particles atop optimized holes. *J. Chem. Phys.* **156** (2022).

182. Arias-Martinez, J. E., Cunha, L. A., Oosterbaan, K. J., Lee, J. & Head-Gordon, M. Accurate core excitation and ionization energies from a state-specific coupled-cluster singles and doubles approach. *Phys. Chem. Chem. Phys.* **24,** 20728–20741 (2022).

183. Cunha, L. A., Hait, D., Kang, R., Mao, Y. & Head-Gordon, M. Relativistic orbital-optimized density functional theory for accurate core-level spectroscopy. *J. Phys. Chem. Lett.* **13,** 3438–3449 (2022).

184. Hait, D. & Head-Gordon, M. 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. *J. Chem. Theory Comput.* **16,** 1699–1710 (2020).

185. Cullen, J., Krykunov, M. & Ziegler, T. The formulation of a self-consistent constricted variational density functional theory for the description of excited states. *Chem. Phys.* **391,** 11–18 (2011).

186. Onida, G., Reining, L. & Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **74,** 601 (2002).

187. Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* **139,** A796 (1965).

188. Yao, Y., Golze, D., Rinke, P., Blum, V. & Kanai, Y. All-electron BSE@ GW method for K-edge core electron excitation energies. *J. Chem. Theory Comput.* **18,** 1569–1583 (2022).

189. Liu, C. *et al.* All-electron ab initio Bethe-Salpeter equation approach to neutral excitations in molecules with numeric atom-centered orbitals. *J. Chem. Phys.* **152** (2020).

190. Golze, D., Keller, L. & Rinke, P. Accurate absolute and relative core-level binding energies from GW. *J. Phys. Chem. Lett.* **11,** 1840–1847 (2020).

191. Keller, L., Blum, V., Rinke, P. & Golze, D. Relativistic correction scheme for core-level binding energies from GW. *J. Chem. Phys.* **153** (2020).

192. Li, J., Jin, Y., Rinke, P., Yang, W. & Golze, D. Benchmark of GW methods for core-level binding energies. *J. Chem. Theory Comput.* **18,** 7570–7585 (2022).

193. Gilmore, K. *et al.* Efficient implementation of core-excitation Bethe–Salpeter equation calculations. *Comput. Phys. Commun.* **197,** 109–117 (2015).

194. Barth, A & Schirmer, J. Theoretical core-level excitation spectra of N2 and CO by a new polarisation propagator method. *J. Phys. B: Atom. Mol. Phys.* **18,** 867 (1985).

195. Wenzel, J., Wormit, M. & Dreuw, A. 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. *J. Comput. Chem.* **35,** 1900–1915 (2014).

196. Wenzel, J., Wormit, M. & Dreuw, A. Calculating X-ray absorption spectra of open-shell molecules with the unrestricted algebraic-diagrammatic construction scheme for the polarization propagator. *J. Chem. Theory Comput.* **10,** 4583–4598 (2014).

197. Wenzel, J., Holzer, A., Wormit, M. & Dreuw, A. Analysis and comparison of CVS-ADC approaches up to third order for the calculation of core-excited states. *J. Chem. Phys.* **142** (2015).

198. Peng, B., Lestrange, P. J., Goings, J. J., Caricato, M. & Li, X. Energy-specific equation-of-motion coupled-cluster methods for high-energy excited states: Application to K-edge X-ray absorption spectroscopy. *J. Chem. Theory Comput.* **11,** 4146–4153 (2015).

199. Coriani, S. & Koch, H. Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework. *J. Chem. Phys.* **143** (2015).

200. Carbone, J. P. *et al.* in *Advances in Quantum Chemistry* 241–261 (Elsevier, 2019).

201. Matthews, D. A. EOM-CC methods with approximate triple excitations applied to core excitation and ionisation energies. *Mol. Phys.* **118,** e1771448 (2020).

202. Simons, M. & Matthews, D. A. Transition-potential coupled cluster. *J. Chem. Phys.* **154,** 014106 (2021).

203. Simons, M. & Matthews, D. A. Accurate Core-Excited States via Inclusion of Core Triple Excitations in Similarity-Transformed Equation-of-Motion Theory. *J. Chem. Theory Comput.* **18,** 3759–3765 (2022).

204. Simons, M. & Matthews, D. A. Transition-potential coupled cluster II: Optimisation of the core orbital occupation number. *Mol. Phys.* **121,** e2088421 (2023).

205. Sorensen, S. *et al.* From synchrotrons for XFELs: the soft x-ray near-edge spectrum of the ESCA molecule. *J. Phys. B: At. Mol. Opt. Phys.* **53,** 244011 (2020).

206. Vidal, M. L., Krylov, A. I. & Coriani, S. Dyson orbitals within the fc-CVS-EOM-CCSD framework: theory and application to X-ray photoelectron spectroscopy of ground and excited states. *Phys. Chem. Chem. Phys.* **22,** 2693–2703 (2020).

207. Vidal, M. L., Pokhilko, P., Krylov, A. I. & Coriani, S. Equation-of-motion coupled-cluster theory to model L-edge x-ray absorption and photoelectron spectra. *J. Phys. Chem. Lett.* **11,** 8314–8321 (2020).

208. Liu, J., Matthews, D., Coriani, S. & Cheng, L. Benchmark calculations of K-edge ionization energies for first-row elements using scalar-relativistic core–valence-separated equation-of-motion coupled-cluster methods. *J. Chem. Theory Comput.* **15,** 1642–1651 (2019).

209. Kucharski, S. A., Włoch, M., Musiał, M. & Bartlett, R. J. Coupled-cluster theory for excited electronic states: The full equation-of-motion coupled-cluster single, double, and triple excitation method. *J. Chem. Phys.* **115,** 8263–8266 (2001).

210. Myhre, R. H., Coriani, S. & Koch, H. Near-edge X-ray absorption fine structure within multilevel coupled cluster theory. *J. Chem. Theory Comput.* **12,** 2633–2643 (2016).

211. Roos, B. O., Taylor, P. R. & Sigbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **48,** 157–173 (1980).

212. Malmqvist, P. Å., Rendell, A. & Roos, B. O. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *J. Phys. Chem.* **94,** 5477–5482 (1990).

213. Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J. & Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **94,** 5483–5488 (1990).

214. Andersson, K., Malmqvist, P.-Å. & Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **96,** 1218–1226 (1992).

215. Malmqvist, P. Å., Pierloot, K., Shahi, A. R. M., Cramer, C. J. & Gagliardi, L. The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO2 and Cu2O2 systems. *J. Chem. Phys.* **128** (2008).

216. Josefsson, I. *et al.* 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. *J. Phys. Chem. Lett.* **3,** 3565–3570 (2012).

217. Pinjari, R. V., Delcey, M. G., Guo, M., Odelius, M. & Lundberg, M. Cost and sensitivity of restricted active-space calculations of metal L-edge X-ray absorption spectra. *J. Comput. Chem.* **37,** 477–486 (2016).

218. De Moura, C. E. & Sokolov, A. Y. Simulating X-ray photoelectron spectra with strong electron correlation using multireference algebraic diagrammatic construction theory. *Phys. Chem. Chem. Phys.* **24,** 4769–4784 (2022).

219. De Moura, C. E. & Sokolov, A. Y. Correction: Simulating X-ray photoelectron spectra with strong electron correlation using multireference algebraic diagrammatic construction theory. *Phys. Chem. Chem. Phys.* **24,** 8041–8046 (2022).

220. Maganas, D., Kowalska, J. K., Nooijen, M., DeBeer, S. & Neese, F. Comparison of multireference ab initio wavefunction methodologies for X-ray absorption edges: A case study on [Fe (II/III) Cl4] 2–/1–molecules. *J. Chem. Phys.* **150** (2019).

221. Wadey, J. D. & Besley, N. A. Quantum chemical calculations of X-ray emission spectroscopy. *J. Chem. Theory Comput.* **10,** 4557–4564 (2014).

222. Nakata, A., Imamura, Y., Otsuka, T. & Nakai, H. Time-dependent density functional theory calculations for core-excited states: Assessment of standard exchange-correlation functionals and development of a novel hybrid functional. *J. Chem. Phys.* **124,** 094105. ISSN: 0021-9606. eprint: `https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.2173987/15379376/094105\_1\_online.pdf` (Mar. 2006).

223. Besley, N. A. & Noble, A. Time-Dependent Density Functional Theory Study of the X-ray Absorption Spectroscopy of Acetylene, Ethylene, and Benzene on Si(100). *J. Phys. Chem. C* **111,** 3333–3340 (2007).

224. Fransson, T. *et al.* XABOOM: An X-ray absorption benchmark of organic molecules based on carbon, nitrogen, and oxygen 1s$\rightarrow \pi^*$ transitions. *J. Chem. Theory Comput.* **17,** 1618–1637 (2021).

225. Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **4,** 4285–4291 (18 2002).

226. Gelius, U, Basilier, E, Svensson, S, Bergmark, T & Siegbahn, K. A high resolution ESCA instrument with X-ray monochromator for gases and solids. *J. Electron Spectrosc. Relat. Phenom.* **2,** 405–434 (1973).

227. Travnikova, O. *et al.* The ESCA molecule—Historical remarks and new results. *J. Electron Spectrosc. Relat. Phenom.* **185,** 191–197 (2012).

228. Feyer, V. *et al.* Tautomerism in cytosine and uracil: An experimental and theoretical core level spectroscopic study. *J. Phys. Chem. A* **113,** 5736–5742 (2009).

229. Feyer, V. *et al.* Tautomerism in cytosine and uracil: A theoretical and experimental X-ray absorption and resonant Auger study. *J. Phys. Chem. A* **114,** 10270–10276 (2010).

230. Hait, D. *et al.* Accurate prediction of core-level spectra of radicals at density functional theory cost via square gradient minimization and recoupling of mixed configurations. *J. Chem. Phys.* **153** (2020).

231. Alagia, M. *et al.* The soft X-ray absorption spectrum of the allyl free radical. *Phys. Chem. Chem. Phys.* **15,** 1310–1318 (2013).

232. Herbst, M. F., Scheurer, M., Fransson, T., Rehn, D. R. & Dreuw, A. adcc: A versatile toolkit for rapid development of algebraic-diagrammatic construction methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **10,** e1462 (2020).

233. Pickup, B. T. On the theory of fast photoionization processes. *Chem. Phys.* **19,** 193–208. ISSN: 0301-0104 (1977).

234. öhrn, Y. & Born, G. in *Advances in Quantum Chemistry* (ed Löwdin, P.-O.) 1–88 (Academic Press, 1981).

235. Spanner, M. *et al.* Dyson norms in XUV and strong-field ionization of polyatomics: Cytosine and uracil. *Phys. Rev. A* **86,** 053406 (5 Nov. 2012).

236. Plasser, F. *et al.* Efficient and Flexible Computation of Many-Electron Wave Function Overlaps. *J. Chem. Theory Comput.* **12.** PMID: 26854874, 1207–1219 (2016).

237. Nakata, A., Imamura, Y. & Nakai, H. Hybrid exchange-correlation functional for core, valence, and Rydberg excitations: Core-valence-Rydberg B3LYP. *J. Chem. Phys.* **125** (2006).

238. Nakata, A., Imamura, Y. & Nakai, H. Extension of the Core-Valence-Rydberg B3LYP Functional to Core-Excited-State Calculations of Third-Row Atoms. *J. Chem. Theory Comput.* **3.** PMID: 26633203, 1295–1305 (2007).

239. Maier, T. M., Bahmann, H., Arbuznikov, A. V. & Kaupp, M. Validation of local hybrid functionals for TDDFT calculations of electronic excitation energies. *J. Chem. Phys.* **144** (2016).

240. Fouda, A. A. & Besley, N. A. Improving the predictive quality of time-dependent density functional theory calculations of the X-ray emission spectroscopy of organic molecules. *J. Comput. Chem.* **41,** 1081–1090 (2020).

241. Siegbahn, K., Nordling, C. & Fahlman, A. *ESCA: Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy* (Uppsala: Almqvist & Wiksell, 1967).

242. Trygubenko, S. A. *et al.* 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. *Phys. Chem. Chem. Phys.* **4,** 4192–4203 (2002).

243. Vidal, M. L. *et al.* Interplay of open-shell spin-coupling and Jahn–Teller distortion in benzene radical cation probed by x-ray spectroscopy. *J. Phys. Chem. A* **124,** 9532–9541 (2020).

244. Ruckman, M. *et al.* Interpreting the near edges of O 2 and O 2- in alkali-metal superoxides. *Phys. Rev. Lett.* **67,** 2533 (1991).

245. Couto, R. C. *et al.* The carbon and oxygen K-edge NEXAFS spectra of CO+. *Phys. Chem. Chem. Phys.* **22,** 16215–16223 (2020).

246. Seidu, I., Neville, S. P., MacDonell, R. J. & Schuurman, M. S. Resolving competing conical intersection pathways: time-resolved X-ray absorption spectroscopy of trans-1, 3-butadiene. *Phys. Chem. Chem. Phys.* **24,** 1345–1354 (2022).

247. Janesko, B. G. Projected Hybrid Density Functionals: Method and Application to Core Electron Ionization. *J. Chem. Theory Comput.* **19,** 837–847 (2023).

248. Janesko, B. G. Core-Projected Hybrids Fix Systematic Errors in Time-Dependent Density Functional Theory Predicted Core-Electron Excitations. *J. Chem. Theory Comput.* **19,** 5112–5121 (2023).

249. Mandal, A., Berquist, E. J. & Herbert, J. M. A new parameterization of the DFT/CIS method with applications to core-level spectroscopy. *J. Chem. Phys.* **161** (2024).

250. Kossoski, F., Boggio-Pasqua, M., Loos, P.-F. & Jacquemin, D. Reference energies for double excitations: Improvement and extension. *J. Chem. Theory Comput.* **20,** 5655–5678 (2024).

251. Pacher, T, Cederbaum, L. & Köppel, H. Approximately diabatic states from block diagonalization of the electronic Hamiltonian. *J. Chem. Phys.* **89,** 7367–7381 (1988).

252. Pacher, T., Mead, C. A., Cederbaum, L. S. & Köppel, H. Gauge theory and quasidiabatic states in molecular physics. *J. Chem. Phys.* **91,** 7057–7062 (1989).

253. Pacher, T, Cederbaum, L. & Köppel, H. Adiabatic and quasidiabatic states in a gauge theoretical framework. *J. Chem. Phys.* **84,** 293–391 (1993).

254. Neville, S. P. & Schuurman, M. S. Calculation of quasi-diabatic states within the DFT/MRCI (2) framework: The QD-DFT/MRCI (2) method. *J. Chem. Phys.* **160** (2024).

255. Zewail, A. H. Femtochemistry: Atomic-scale dynamics of the chemical bond. *J. Phys. Chem. A* **104,** 5660–5694 (2000).

256. Nisoli, M., Decleva, P., Calegari, F., Palacios, A. & Martín, F. Attosecond electron dynamics in molecules. *Chem. Rev.* **117,** 10760–10825 (2017).

257. Plekan, O *et al.* A theoretical and experimental study of the near edge X-ray absorption fine structure (NEXAFS) and X-ray photoelectron spectra (XPS) of nucleobases: Thymine and adenine. *Chem. Phys.* **347,** 360–375 (2008).

GLOSSARY

***eh*-SIE** electron-hole self-interaction error. 22, 23, 38, 52, 96

***p*-DFT/MRCI** configuration pruned variant of the density functional theory and multireference configuration interaction method. 40, 46, 49, 64, 66, 67, 72, 73, 99, 101, 102, 108, 110, 114, 117, 120, 127, 178, 184, 186

ADC algebraic-diagrammatic construction. 80, 82

ADC(2)-x extended algebraic-diagrammatic construction through second order. 80, 90, 105, 107

ADC(3) algebraic-diagrammatic construction through third order. 80

AO atomic orbital. 24

BDD block diagonalization diabatization. 122

BO Born–Oppenheimer. 4, 8, 122

BPR Boltzmann population ratio. 105, 187

BSE Bethe–Salpeter equation. 79, 80

CAS complete active-space. 82

CASPT2 complete active-space multiconfigurational second-order perturbation theory. 81

CASSCF complete active-space self-consistent field. 81, 82

CC coupled-cluster. 4, 5, 30, 80, 82

CC2 second-order approximate coupled-cluster singles and doubles. 81

CC3 third-order approximate coupled-cluster singles and doubles. 61, 89

CCSD coupled-cluster singles and doubles. 81

CCSDT coupled-cluster singles, doubles and triples. 49, 61, 88, 89, 143

CI configuration interaction. 6, 22, 25–27, 29–33, 35, 38, 46, 56, 82, 93, 116, 122

CIPSI configuration interaction using a perturbative selection made iteratively. 49

CIS configuration interaction singles. 22, 23, 27, 33, 34, 52, 143

CISD configuration interaction singles and doubles. 27

CSF configuration state function. 27–29, 32, 36, 37, 44, 46, 49, 56, 60, 61, 116

CT charge-transfer. 4, 5, 22, 23, 40, 77, 79

CVS core-valence separation. 51, 78, 80, 83, 84

CVS-ADC core-valence separated extended algebraic-diagrammatic construction. 80

CVS-ADC(2) core-valence separated algebraic-diagrammatic construction through second order. 80

CVS-ADC(2)-x core-valence separated extended algebraic-diagrammatic construction through second order. 80, 90, 105, 186

CVS-ADC(3) core-valence separated algebraic-diagrammatic construction through third order. 80

CVS-DFT/MRCI core-valence separated density functional theory and multireference configuration interaction. 40, 83, 85

CVS-EOM-CCSD core-valence separated equation of motion coupled cluster singles and doubles. 81, 105

CVS-EOM-CCSDT core-valence separated equation of motion coupled cluster singles, doubles and triples. 52, 81, 88, 89, 117, 166

CVS-EOMIP-CCSDT core-valence separated equation of motion ionization potential coupled cluster singles, doubles and triples. 81, 88, 117

DDCI difference dedicated configuration interaction. 49

DFA density functional approximation. 12, 16

DFT density functional theory. 5–8, 18–21, 24, 27, 29, 40, 51, 55, 77, 78, 82, 83, 114, 119, 120

DFT/CIS density functional theory and configuration interaction singles. 32–34, 47, 121

DFT/MRCI density functional theory and multireference configuration interaction. 5, 6, 20, 32, 35–38, 40–46, 48–52, 54–57, 60, 61, 64, 66, 67, 69, 70, 72–75, 82–84, 87, 89, 92, 93, 96, 99, 101, 102, 105, 107, 108, 110, 111, 114, 116–122, 127, 146, 166, 175, 178, 184–187

DFT/MRCI(2) perturbative approximation variant of the density functional theory and multireference configuration interaction method. 40, 46, 64, 66, 70–74, 99, 101, 102, 108, 110, 114, 117, 120, 122, 123, 127, 178, 184, 186

DLPNO-STEOM-CCSD domain local pair natural orbital similarity transformed EOM-CCSD. 71

EA-TDA electron-affinity time-dependent density functional theory within the Tamm–Dancoff approximation. 79

EEX exact exchange. 18, 19, 21–23, 42, 55, 59, 85, 96, 97

EKT extended Koopmans' theorem. 55

EOM equation-of-motion. 80

EOM-CCSD equation of motion coupled cluster singles and doubles. 49, 69, 90, 105, 107

EOM-CCSDT equation of motion coupled cluster singles, doubles and triples. 52

ESCA electron spectroscopy for chemical analysis (a.k.a. X-ray photoelectron spectroscopy). 103

ET-SIE electron-transfer self-interaction error. 22

fc-CVS-EOM-CCSD *frozen-core* core-valence separated equation of motion coupled cluster singles and doubles. 90, 113, 187

FCI full configuration interaction. 29, 30, 49, 60, 126, 127

FOIS first-order interacting space. 32, 35–37, 40, 44–46, 49, 60, 61, 64, 116

FWHM full width at half maximum. 69, 103, 113, 177, 184, 185

GGA generalized gradient approximation. 17, 54

GVVPT2 second-order generalized van Vleck perturbation theory. 64

HF Hartree–Fock. 18, 19, 24, 25, 27, 30, 33–36, 52, 55

HHG high-order harmonic generation. 77

HK Hohenberg–Kohn. 9, 10

HK-DFT Hohenberg–Kohn density functional theory. 9

HKS Hohenberg–Kohn–Sham. 10

HOMO highest occupied molecular orbital. 55, 70

IMOM initial maximum overlap method. 79

IP ionization potential. 55

KS Kohn–Sham. 9–11, 16, 18, 21, 24, 33–35, 41, 43, 44, 51, 52, 55, 60, 79, 83, 92, 96, 97, 110, 117, 118, 120

KS-DFT Kohn–Sham density functional theory. 10, 20, 24, 33, 36, 49, 55, 72, 89, 114

LDA local density approximation. 16–18

LR linear response. 21, 80, 113

LR-TDDFT linear-response time-dependent density functional theory. 21, 78

LSDA local spin-density approximation. 16

LUMO lowest unoccupied molecular orbital. 70, 114

MAE mean absolute error. 52, 54, 55, 60–62, 64, 66, 69, 74, 80, 97–99, 101, 102, 112, 117, 119, 143, 146, 175, 177

ME mean error. 54, 55, 66, 143

MLCCSD multilevel coupled-cluster singles and doubles. 81

MO molecular orbital. 19, 25, 27, 29, 33, 47

MOM maximum overlap method. 79

MR multireference. 4, 6, 23, 25, 30, 40, 77, 78, 81, 111, 114

MRCI multireference configuration interaction. 6, 30–32, 34–36, 38, 40, 41, 44, 46

MRCISD multireference configuration interaction singles and doubles. 31

NEXAFS near-edge X-ray absorption fine structure. 89, 90, 105, 108, 113, 115, 116, 176, 184, 186, 187

OCDFT orthagonality-constrained density functional theory. 78, 79

OLED organic light-emitting diode. 2

OO orbital optimized. 79, 112

P-BDD propagative block diagonalization diabatization. 122

PZ-SIC Perdew–Zunger self-interaction correction. 20

QD-DFT/MRCI(2) perturbative approximation variant of the density functional theory and multireference configuration interaction method for quasi-diabatic states. 122

QTP Quantum Theory Project. 55, 143

RAS restricted active-space. 46, 47, 82

RASCI restricted active-space configuration interaction. 46

RASPT2 restricted active-space multiconfigurational second-order perturbation theory. 81, 82

RASSCF restricted active-space self-consistent field. 81, 82

RKS restricted Kohn–Sham. 91, 110

RMSD root-mean squared deviation. 79, 87, 112

ROKS restricted open-shell Kohn–Sham. 79, 91, 110

RT real time. 21, 78

RT-TDDFT real-time time-dependent density functional theory. 78

SCF self-consistent field. 24, 27

SIE self-interaction error. 16, 19, 20, 22, 23, 38, 97, 120

SOMO singly occupied molecular orbital. 90, 112, 113

SR single reference. 24, 30, 81, 82

STEX static exchange approximation. 79

TA transient absorption. 3, 107

TBE theoretical best estimate. 49, 60, 61, 69, 71, 72, 83, 89, 119, 126, 127

TDA Tamm–Dancoff approximation. 22, 34, 52, 79

TDA-TDDFT time-dependent density functional theory within the Tamm–Dancoff approximation. 33, 34, 51, 52, 92, 96, 143

TDA-TDHF time-dependent Hartree–Fock theory within the Tamm–Dancoff approximation. 143

TDDFT time-dependent density functional theory. 5, 6, 21–23, 38, 59, 78–80, 85, 95, 96, 113, 121, 146

UV/Vis ultraviolet/visible. 41–43, 57, 62, 71, 72, 107

VEE vertical excitation energy. 40–43, 48, 51, 57, 60–62, 64, 67–72, 74, 81, 87, 92, 101, 119, 122

WFT wave function theory. 4, 5, 24, 77, 78, 80, 82

XAS X-ray absorption spectra. 3, 90, 93, 107, 112, 113, 115

XC exchange-correlation. 5, 6, 9–12, 14–18, 21–24, 34, 38, 42, 43, 51, 52, 54, 55, 59, 60, 74, 78, 83, 85, 92, 96, 117, 118, 120, 143

XFEL X-ray free-electron laser. 3, 77

XPS X-ray photoelectron spectra. 82, 90, 93, 103, 105, 115, 184–186