

# Predicting Carbonyl Excitation Energies Efficiently Using EOM-CC Trends

Keiran N. Rowell,<sup>†</sup> Scott H. Kable,<sup>†</sup> and Meredith J. T. Jordan<sup>\*,‡</sup>

<sup>†</sup>*School of Chemistry, University of New South Wales, Sydney*

<sup>‡</sup>*School of Chemistry, University of Sydney, Sydney*

E-mail: meredith.jordan@sydney.edu.au

Phone: +61 2 9351 4420. Fax: +61 2 9351 3329

## Abstract

We approach the problem of predicting excitation energies of diverse, larger (5–6 carbons) carbonyl species central to earth’s tropospheric chemistry. Triples contributions are needed for the vertical excitation energy ( $E^{\text{vert}}$ ), while EOM-CCSD//TD-DFT calculations provide acceptable estimates for the  $S_1$  relaxation energy ( $E^{\text{relax}}$ ), and (TD-)DFT suffices for the  $S_0 \rightarrow S_1$  zero-point vibration energy correction ( $\Delta E^{\text{ZPVE}}$ ).

Perturbative triples corrections deliver  $E^{\text{vert}}$  values close in accuracy to full iterative triples EOM-CC calculations. The error between EOM-CCSD and triples-corrected  $E^{\text{vert}}$  values appears to be systematic and can be accounted for with scaling factors. However, saturated and  $\alpha, \beta$ -unsaturated carbonyls must be treated separately. Double-hybrid  $S_0$  minima can be used to calculate  $E^{\text{vert}}$  with negligible loss in accuracy, relegating the  $\mathcal{O}(N^5)$  expense of CCSD to only single-point energy and excitation calculations. This affordable protocol can be applied to all volatile carbonyl species.

$E^{0-0}$  predictions do overestimate measured values by  $\sim 8$  kJ/mol due to a lack of triples contribution in  $E^{\text{relax}}$ , but this overestimation is systematic and the mean unsigned error is within 4 kJ/mol once this is accounted for.

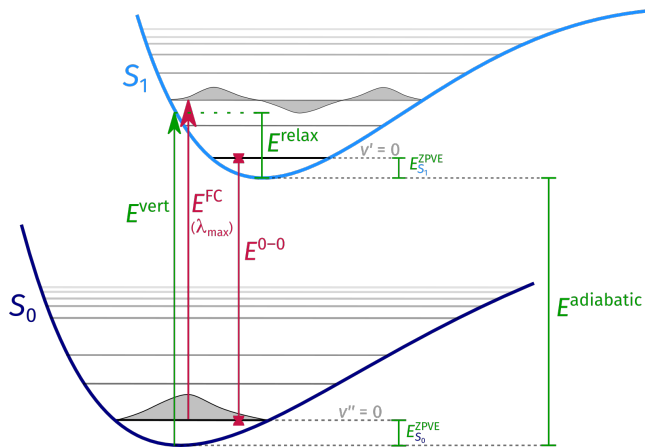
## Introduction

**Excitation energies:**  $\lambda_{\text{max}}$ ,  $E^{\text{vert}}$ , and  $E^{0-0}$ .

In order to predict photochemical reaction quantities (*e.g.* photolysis threshold energies), the excitation energy offset between the ground and excited electronic states ( $E^{0-0}$ ) can be computed ‘once-off’ for each molecule. We have previously identified 20 small carbonyl species that are relevant to tropospheric chemistry and for which experimental data is available.<sup>1</sup> Here, we add an additional 18 larger carbonyl species that require an affordable protocol for theoretical calculation of their excitation energies to be viable. The states of interest here for carbonyls are the ground and excited singlet states ( $S_0$  and  $S_1$  respectively), since direct transition from the ground state to the triplet state ( $S_0 \rightarrow T_1$ ) is spin-forbidden in the  $n \rightarrow \pi^*$  excitation. This  $S_0 \rightarrow S_1$   $E^{0-0}$  energy can then be added to quantities computed on the excited state potential energy surface, such as intrinsic reaction barriers relative to the excited-state minimum, to give photochemical energies relative to the ground state energetic minimum. The prediction of molecular excitation energies is generally one of the largest sources of errors in *in silico* photochemistry prediction,<sup>2</sup> and subsequently requires use of some of the most computationally demanding *ab initio* methods.

One of the most common approaches to bench-

mark computational excitation energy methods is to approximate the experimental excitation energy at the absorbance maximum ( $\lambda_{\max}$ ) with a single point vertical excitation energy computed on an  $S_0$  minimum energy geometry<sup>3-8</sup> ( $E^{\text{vert}}$ ), as illustrated in Figure 1.



**Figure 1:** Schematic of the excitation energies in carbonyls (see text). Quantities corresponding to experimental observables are shown in red, while calculated quantities are shown in green. Corrections for zero-point vibrational energy (ZPVE) are indicated. Excitation between  $S_0$  and  $S_1$  is illustrated, analogous quantities for  $S_0 \rightarrow T_1$  excitations can also be obtained.

However,  $E^{\text{vert}}$  cannot be directly compared to wavelength of maximum absorbance ( $\lambda_{\max}$ ), as the latter contains several factors not captured by  $E^{\text{vert}}$  such as: zero-point energy; vibrational wavefunction overlap; and complications arising from selection rules (such as the  $\nu_4$  out-of-plane bending promoting Herzberg–Teller coupling in formaldehyde).<sup>9</sup> Nevertheless, direct comparison of  $\lambda_{\max}$  and theoretical  $E^{\text{vert}}$  values persists in the literature because, for many molecules, the difference between these values is small — often significantly less than typical errors of many commonly used excited state theoretical methods.<sup>6,10</sup>

While corrections to  $\lambda_{\max}$  ( $E^{\text{FC}}$ ) accounting for the above factors can be calculated to obtain ‘quasi-experimental’ ‘ $E^{\text{vert}}$ ’ estimates with uncertainties of  $\pm 10$  kJ/mol,<sup>11</sup> here we wish to benchmark computational prediction against a directly corresponding experimental observable — the  $E^{0-0}$  energy illustrated in Figure 1.

Fortunately, Godunov and Yakovlev have compiled a series of experimental carbonyl  $E^{0-0}$

excitation energies<sup>12</sup> (reproduced in Table S1) against which our calculations can be validated. However, experimental data are only available for species containing four carbon atoms or less. This highlights the need for accurate theoretical  $E^{0-0}$  protocols which can be validated for “smaller” carbonyls ( $\leq 4$  carbons) and then applied to larger species (5–6 carbons) at modest cost. It is expected that larger carbonyls should follow the same trends in excitation energy as smaller species, since carbon chains and distant moieties have little effect on the C=O chromophore environment. This is shown in the absorbance maximum data in Table S2 and Figure S1, where there is negligible change of  $\lambda_{\max}$  with increasing sidechain length, within a given carbonyl class.

Calculating a corresponding theoretical  $E^{0-0}$  excitation energy involves several corrections to  $E^{\text{vert}}$ , as illustrated in Figure 1. After vertical photoexcitation from the  $S_0$  minimum to  $S_1$ , the nuclei will relax to the minimum energy configuration on  $S_1$ . This relaxation energy,  $E^{\text{relax}}$  is calculated using an excited state geometry optimisation. The ‘adiabatic’ excitation energy is:  $E^{\text{adiabatic}} = E^{\text{vert}} - E^{\text{relax}}$ . A theoretical  $E^{0-0}$  prediction can be obtained after computing the zero-point vibrational energy (ZPVE) at both the ground and excited-state minimum energy geometries.  $\Delta E^{\text{ZPVE}} = E_{S_1}^{\text{ZPVE}} - E_{S_0}^{\text{ZPVE}}$ . Then, the 0–0 excitation energy is the ZPVE-corrected adiabatic energy:  $E^{0-0} = E^{\text{adiabatic}} + \Delta E^{\text{ZPVE}}$ . In this work only harmonic frequency calculations are performed, and so the effect of anharmonicity on ZPVEs is not explicitly accounted for, rather, it is included implicitly through the use of ZPVE scaling factors.

The sensitivity to the level of theory used to calculate each component of a theoretical  $E^{0-0}$  energy will be reported and discussed in separate subsections below. For calculated values which have no experimental counterpart, comparison with respect to high-level theory<sup>13</sup> is given. Where ‘benchmark quality’ results are unavailable, the variation between different methods is studied, with reference literature precedents. Finally, protocols to obtain a theoretical  $S_0 \rightarrow S_1$   $E^{0-0}$  estimate are proposed, and their relative accuracy assessed against the

available carbonyl 0–0 excitation measurements.

## Theoretical considerations for carbonyl excitation energies

The excited state formalisms used for calculating excited electronic states in this work are: equation-of-motion coupled cluster (EOM-CC), in which an equation-of-motion excitation operator is used in the CC ansatz;<sup>14</sup> and time-dependent density functional theory (TD-DFT), in which excitation energies are found from the poles of the density response in solutions of the time-dependent DFT equations.<sup>11</sup> Both methods inherit the strengths and shortcomings of their ground state counterparts. TD-DFT is relatively inexpensive compared to *ab initio* methods, but the accuracy is dependent on the specific functional used and is not systematically improvable. EOM-CC energies can be systematically improved by including higher order terms in the cluster operator, but cost scaling with system size is very high and prohibitive for all but very small molecules. Both methods are used in a ‘linear response’ formalism, which is applicable to electronic excited states of ‘single excitation’ character originating from closed-shell configurations. The excitations studied in this paper are of such single excitation character, and therefore the use of the linear response (LR) formalism is taken as implicit in the methods below.

Both EOM-CC and TD-DFT are also ‘single reference’ methods and do not require definition an active space. Multiconfigurational methods, such as the complete active space self-consistent field (CASSCF) method,<sup>15</sup> can also be applied to electronic excited states by using a pre-defined active space of ‘relevant’ molecular orbitals. Defining this active space can be difficult, and even contentious,<sup>16</sup> and these methods are also limited by their treatment, or lack thereof, of dynamic electronic correlation. This can lead to poor quality electronic excitation energies, even when second order perturbation corrections (CASPT2)<sup>17</sup> are used. CASPT2 excitation energies for small organic molecules are typically only accurate to 20 kJ/mol with reference to benchmark CC3 theoretical results.<sup>6,18,19</sup>

Despite LR response formalisms being appro-

priate for carbonyl  $n \rightarrow \pi^*$  electronic states, there are still significant issues in obtaining reliable energies. The common goal in ground state theoretical chemistry of ‘chemical accuracy’ of 4 kJ/mol deviation to experiment is not achievable, in general, for excited state energies.<sup>20,21</sup> This is especially true when the computational cost has to be low enough for routine calculations on several dozen molecules, which is desired in this work to further the project of obtaining structure–activity relationships (SARs) of all possible photochemical channels across entire carbonyl classes.<sup>22</sup> An aim of this work is the development of an affordable computational protocol which can consistently achieve a 10 kJ/mol accuracy for carbonyl photochemistry.

Here, composite approaches are adopted that use (TD-)DFT optimisations and EOM-CC single point energies for calculating  $S_0 \rightarrow S_1$  excitation energies, and are benchmarked against available experimental data. The sections below outline key literature findings on the treatment of  $n \rightarrow \pi^*$  excited states using EOM-CC and TD-DFT methods.

Previous work on computational carbonyl photochemistry has shown high level *ab initio* approaches to be highly dependent on the accuracy of geometries<sup>2,23</sup> and the level of theory used.<sup>13</sup> Indeed, the lack of experimental structural data is a driving motivator for this current work. In the calculation of  $E^{0-0}$  energies, the major source of error is the value of  $E^{\text{vert}}$ . High level *ab initio* calculations are prohibitively expensive for many molecules in the “large” carbonyl dataset and even the extrapolated W1-X1 composite scheme<sup>24</sup> is not sufficiently accurate for carbonyl excited state energies.<sup>23</sup>

## Time-dependent Density Functional Theory (TD-DFT)

**Choice of density functional** Time-dependent density functional theory (TD-DFT) is the most widely used excited state method<sup>3,25</sup> due to its computational efficiency and ease of use while delivering ‘good’ energetics<sup>26</sup> — provided an appropriate exchange-correlation functional has been used. For singlet excited states of organic molecules, various TD-DFT

functionals have been benchmarked across  $\sim 500$  molecules extending Thiel’s dataset.<sup>4</sup> From this benchmark work, global hybrid (GH) functionals tend to be more accurate than generalised gradient approximation (GGA) and meta-GGA functionals. The inclusion of exact Fock exchange energy is therefore important for obtaining accurate excited state energies. Both GHs and range-separate hybrids (RSHs) are consistently the most accurate classes of functionals for excited states, delivering errors of the same magnitude as some of the most affordable EOM-CC methods (specifically, CC2).<sup>27,28</sup> CAM-B3LYP<sup>29</sup> was one of the best RSHs for singlet excited states in the extended dataset, though the GH B3LYP<sup>30,31</sup> yielded slightly smaller average errors. While B3LYP should perform well for valence single excitations (such as those studied here), RSH functionals are important if the excited state has some Rydberg character.<sup>32–34</sup>

**Geometric accuracy of B3LYP and CAM-B3LYP** From a study of the excited state geometry optimisation of 41 small organic molecules using a triple- $\zeta$  basis set, B3LYP underestimates C=O bond lengths by 0.034 Å, while CAM-B3LYP underestimates them by 0.046 Å, with reference to benchmark CC3 geometries.<sup>35</sup> For small unsaturated molecules, B3LYP and CAM-B3LYP were found to give  $n \rightarrow \pi^*$  excited state geometries similar to those from CASPT2 calculations.<sup>36</sup> CAM-B3LYP is found to slightly outperform B3LYP in calculating excited state geometries, but underperforms B3LYP for ground state geometries.<sup>37</sup>

**The Tamm-Dancoff approximation** The Tamm-Dancoff approximation<sup>38</sup> (TDA) neglects the occupied–virtual interactions in the full TD-DFT equations, leading to a formalism similar to configuration interaction singles (CIS).<sup>38</sup> The TDA greatly reduces the cost of TD-DFT calculations, particularly with hybrid functionals, while delivering excitation energies of comparable quality to full TD-DFT calculations.<sup>39</sup> In the context of  $n \rightarrow \pi^*$  excitation energies in small organic molecules, application of the TDA was found to slightly increase calculated

$S_0 \rightarrow S_1$  transition energies, decreasing the mean unsigned error (MUE) of CAM-B3LYP calculations from 11 to 8 kJ/mol. This is not an isolated observation, in many applications the use of TDA TD-DFT results in better agreement with experimental quantities, such as  $E^{0-0}$  energies.<sup>40,41</sup> The TDA is used in all TD-DFT calculations herein, and is taken as implicit.

**Effect of exact exchange** The Minnesota functionals (M06-L, M06, M06-2X and M06-HF) differ in percentage contribution of Fock exchange, and provide a case study in effect of exact exchange in TD-DFT. Greater amounts of Fock exchange is shown to blue-shift excitation energies.<sup>3,11,42</sup> That the M06 family of functionals do not provide sufficiently accurate excitation energies,<sup>19</sup> regardless of exact exchange —0% (M06-L) or 100% (M06-HF)— means that agreement with experiment cannot be achieved by changing the amount of exact exchange alone. The treatment of electron correlation must also be considered, through either the use of different correlation functionals or the use of correlated *ab initio* methods.

**Effect of correlation energy in double-hybrid DFT** Double-hybrid density functionals (DHDFs) improve the treatment of electron correlation by incorporating an *ab initio* second order perturbative correction to the correlation energy in DFT.<sup>43</sup> Excited state TD-DHDFs include a perturbative doubles-correction similar to the approach used in CIS(D).<sup>44</sup> Most of the expense of this correlation correction can be alleviated with resolution of the identity (RI) methods.<sup>39,45,46</sup> This correlation correction causes DHDFs to suffer less from self-interaction error and excited ‘ghost’ states than conventional functionals.<sup>47</sup> TD-DHDFs perform better than conventional DFT for excited state properties<sup>44</sup> and excitation energies,<sup>11</sup> with the DHDF B2GP-PLYP performing particularly well.<sup>27</sup>

Though TD-DHDF is more accurate than conventional TD-DFT, deviations in vertical excitation energies still remain around 10 kJ/mol.<sup>11</sup> The accuracy of TD-DHDF is also dependent on the quality of the excited state geometry used. TD-B2GP-PLYP//B3LYP calcu-

lations delivered a mean unsigned error (MUE) of 23 kJ/mol for the Thiel benchmark set, while 10 kJ/mol accuracy is only reached when coupled cluster geometries are used in TD-B2GP-PLYP//SCS-CC2 calculations.<sup>48</sup> Though spin-component scaling of the MP2 contribution in TD-DHDFT has been shown to lead to further increases in accuracy,<sup>47–49</sup> these single-point excitation calculations are still out-performed by EOM-CC energies.

Despite these successes of TD-DHDFT, analytical excited state gradients are not available for TD-DHDFT methods, making them impractical for  $S_1$  state geometry optimisation and frequency calculations.

### Equation of Motion Coupled Cluster (EOM-CC)

Equation of motion of coupled cluster (EOM-CC) methods provide a systematic route to accurate excited energies, though at high computational cost scaling based upon the order of contributions included in the cluster operator. Inclusion of quadruples leads to almost perfect agreement ( $\leq 1$  kJ/mol) to extrapolated full configuration interaction (exFCI)  $E^{\text{vert}}$  energies,<sup>13</sup> but the  $\mathcal{O}(N^{10})$  scaling of CCSDTQ makes the method infeasible for all but the smallest molecules. While CCSDT formally scales as  $\mathcal{O}(N^8)$  this can be reduced to  $\mathcal{O}(N^7)$  if the CC3 method is used.<sup>50</sup> CC3 reproduces EOM-CCSDT singlet valence excitation energies well, particularly when a triple- $\zeta$  basis set is used.<sup>8,51</sup> If triples are included perturbatively using CCSDR(3) a more manageable cost scaling can be achieved.<sup>52,53</sup> The use of diffuse functions or basis sets larger than triple- $\zeta$  is found to be important for Rydberg excited states, and relatively unimportant for valence states like the  $n \rightarrow \pi^*$  excitations studied here.<sup>21,54</sup>

Geometry optimisation and frequency calculations with EOM-CC including triples contributions is impractical, even with analytical gradients,<sup>55</sup> for all but the smallest, most symmetric, molecules. However, analytic gradients make (EOM-)CCSD geometry optimisations<sup>56</sup> viable for smaller carbonyl species. The deviation of excited state CC2 and EOM-CCSD geometries

from CC3 geometries has been characterised by Budzak *et al.* for several small carbonyls, and ranges from 0.01–0.02 Å.<sup>57</sup>

Loos *et al.* have shown that theoretical  $E^{0-0}$  energy predictions can achieve chemical accuracy, though their original protocol required the use of CCSDR(3)/def2-TZVPP geometries and CC3/aug-cc-pVTZ energies.<sup>21</sup> The expense of CCSDR(3) optimisations restricted these calculations to the smallest carbonyls: formaldehyde, acetaldehyde, acetone, acrolein and propynal. In a follow-up study, Loos and Jacquemin were able to reduce the computational expense by using (EOM-)CCSD geometries, and were able to retain chemical accuracy.<sup>58</sup> The authors caution that this protocol relies on a near-systematic cancellation of errors: underestimated reorganization energies are almost exactly compensated by overestimation of vertical excitation energies.

**Composite methods for  $E^{0-0}$**  A recent review of literature calculations of  $E^{0-0}$  excitation energies demonstrates that errors with hybrid TD-DFT functionals fall in the range of 20–40 kJ/mol,<sup>20</sup> outside the 10 kJ/mol accuracy desired here. Even CC2  $E^{0-0}$  energies are outside this accuracy, with deviations with respect to experiment in the 10–20 kJ/mol range.<sup>20</sup> Some success using affordable methods has been found in combining TD-DFT optimised geometries and frequencies with low cost EOM-CC vertical excitation energies. This is demonstrated in the work of Oruganti *et al.*<sup>59</sup> where  $\sim 9$  kJ/mol deviations are found with a combined CC2/DFT'' calculation of  $E^{0-0}$  energies. The aforementioned composite approach of Loos and Jacquemin using coupled cluster methods for both geometries and energies achieved ‘chemically accurate’  $S_0 \rightarrow S_1$   $E^{0-0}$  predictions, however with the caveat of cancellation of errors mentioned above.<sup>58</sup>

However, no broadly applicable computational protocol has been demonstrated to achieve the desired sub-10 kJ/mol error in predicted  $E^{0-0}$  energies.<sup>58,60</sup> This current work is only concerned only with  $n \rightarrow \pi^*$  excitations of carbonyls, and ultimately only as means to the end of predicting photolysis thresholds. The methods used in this work aim to exploit the

similarities within carbonyl classes and resulting trends in EOM-CC  $E^{\text{vert}}$  energies to determine an affordable protocols for predicting  $n \rightarrow \pi^*$   $E^{0-0}$  energies to within 10 kJ/mol of experimental values. The results here are therefore only applicable to carbonyl valence excitations, and are not proposed to be generalisable to other excitation types.

## Computational Methods

### Software and common settings

Geometry optimisations and frequency calculations were performed with ORCA<sup>61</sup> version 4.0.1.2 or 4.1.0, except for when the (EOM-)CCSD or M06-2X methods were used, in which case calculations were performed with Gaussian 16.<sup>62</sup>  $S_0 \rightarrow S_1$  vertical excitations were computed using ORCA for *bt*-PNO-STEOM-CCSD calculations, and Dalton 2018.1<sup>63</sup> for CC2, CCSD, CCSDR(3), and CC3 calculations. All  $S_0$  quantities were calculated with ORCA using the B2GP-PLYP functional.

The Weigend and Aldrichs triple- $\zeta$  split valence basis set, def2-TZVP,<sup>64</sup> was used for all (TD-)DFT geometry optimisations and frequency calculations. CCSD geometry optimisations used the aug-cc-pVTZ basis set.<sup>65,66</sup> Vertical excitation transition energies were calculated with the def2-TZVPP basis set; except for *bt*-PNO-STEOM-CCSD calculations for which the larger canonical aug-cc-pVQZ<sup>65</sup> basis set and associated auxiliary basis sets<sup>46,67</sup> were used. All DFT calculations were dispersion-corrected using the D3(BJ)<sup>68,69</sup> scheme. All energies are ZPVE-corrected, using an appropriate scaling factor.<sup>70</sup> The resolution of the identity (RI) approach was used to speed up electronic structure calculations in ORCA. DFT calculations in ORCA employed the RI-JK scheme, which is efficient for “smaller” molecules.<sup>71</sup> Auxiliary basis sets derived from the def2-TZVP canonical basis set were used when required: def2/JK<sup>72</sup> for the exchange and Coulomb integrals and def2-TZVP/C<sup>73</sup> for the correlation energy. The RI-JK scheme has not been implemented in ORCA for TD-DFT methods. In this case, the RIJ-

COSX scheme<sup>71</sup> was used for the  $S_1$  TD-DFT calculations reported below.

The DFT calculations employed tight SCF convergence of  $10^{-8}$  a.u. A finer numerical grid (the ‘Grid5’ keyword in ORCA, ‘Integral=UltraFine’ in the Gaussian 16 default) was used to ensure the reliability of the geometry optimisations and numerical frequencies. The DFT exchange grid was kept at the ORCA default coarse grid to maintain computational efficiency.

In the methods described below, basis set, D3(BJ) dispersion correction, resolution of the identity, time-dependent, and the equation-of-motion, formalisms are taken as implicit for relevant electronic states.

### Geometries — DFT and CCSD

#### $S_0$ minima for vertical excitation calculations

A reliable  $S_0$  minimum structure is required for computing vertical excitation energies since EOM-CC methods are highly sensitive to the quality of the geometry.<sup>2</sup> CCSD optimisation and frequency calculations were performed, however for a few carbonyl species with 6 carbons, this was not feasible. Following benchmarking of B2GP-PLYP optimised geometries against CCSD geometries for “small” carbonyls, the B2GP-PLYP functional was used to optimise the  $S_0$  geometries of the “large” carbonyls. Analytic gradients were used for all B2GP-PLYP optimisations. Frequencies were obtained by two-sided numerical differentiation with a step length of 0.005 bohr.

#### $S_1$ Excited state minima

TD-DFT calculations, employing the Tamm-Dancoff approximation,<sup>38</sup> were used to describe  $S_1$  excited states. The lowest 5 roots were determined to ensure the reliability of the  $S_1$  energies. Because it lacks analytic gradients within the TD-DFT formalism, the B2GP-PLYP functional was not used for  $S_1$  optimisations. Instead,  $S_1$  geometry optimisations were performed with both the GH B3LYP functional,<sup>30,31</sup> and the RSH CAM-B3LYP functional.<sup>29</sup> EOM-CCSD

geometries were also obtained for “small” carbonyl structures.

## Zero-point corrections

All reported energies  $E^{0-0}$  energies include ZPVE correction. These ZPVEs are scaled by an appropriate scaling factor: 0.9752 for B3LYP/def2-TZVP calculations; 0.976 for CAM-B3LYP/def2-TZVP calculations; and 0.9754 for M06-2X/def2-TZVP calculations.<sup>70</sup> The same scaling factor is used for all electronic surfaces, implicitly assuming the neglect of electronic-state-specific ZPVE scaling factors to be a minor error compared to other errors in the energetic components of  $E^{0-0}$  values.

## Vertical excitation energies

Vertical excitation energies were calculated at  $S_0$  optimised minimum energy geometries, using either the B2GP-PLYP or CCSD methods, as outlined above. These are respectively denoted  $[EOM-CC]/B2GP-PLYP$  or  $[EOM-CC]/CCSD$  in all plots and discussions herein — where  $[EOM-CC]$  is replaced by the level of theory used for the vertical excitation energy calculations.  $S_0$  minimum structures that have symmetry elements (*e.g.*  $C_s$  point group structures with a mirror plane) were fully symmetrised before being used in vertical excitation energy calculations. The point groups of the  $S_0$  minima and state symmetries of the  $S_1$  excited states are listed for each molecule in Table S3. In three test molecules, formaldehyde, acetone, and glyoxal, the inclusion of 1–5 roots in the EOM-CC calculations returned virtually identical results (Table S4) and so, unlike the TD-DFT calculations, only a single root was determined in EOM-CC calculations.

Several EOM-CC methods were used for vertical excitation energies in order to assess the impact of including higher order terms in the cluster operator, and the use of a perturbative triples correction. In order of cost (and presumed methodological rigour and accuracy) the different EOM-CC methods considered are:

- *bt*-PNO-STEOM-CCSD, an approximation to CCSD that exploits locality us-

ing pair natural orbitals to achieve non-iterative  $\mathcal{O}(N^5)$  scaling. Abbreviated as STEOM below;

- CC2, an iterative approximation to the full CCSD solution;
- CCSD, the full iterative solution to the EOM-CC formalism solving for both single and double excitations;
- CCSDR(3), full iterative singles and doubles solution with a non-iterative perturbative triples estimate;
- CC3, an iterative approximation to the full CCSDT solution using a more efficient calculation of triples

The lower scaling of the STEOM method in comparison to other EOM-CC methods allowed use of the cc-pVQZ basis set. Since a large basis set was required, the integration grids were increased to their finest pre-defined settings (`Grid7`, `GridX7` in ORCA) and radial integration was manually set to an even finer grid (`IntAcc 6.0` in ORCA). Within the STEOM approximation, the cut-off between PNOs was set to the most accurate threshold (`TightPNO` in ORCA). Automated active space selection was used,<sup>74</sup> and all calculations reported have an orbital percentage active character value of  $> 98.0$ , indicating the validity of the active space.

## Results and Discussion

### Vertical excitation energies — $E^{\text{vert}}$

The  $E^{\text{vert}}$  energies for each “small” carbonyl molecule ( $\leq 4$  carbons) are shown in Figure 2, where the results of various EOM-CC methods are correlated against the highest level of theory consistently achievable across the dataset: CCSDR(3)//CCSD shown on the y-axis.

The effect of geometry on  $E^{\text{vert}}$  is highlighted by the use of open circles ( $\circ$ ) for CCSD geometries, and open squares ( $\square$ ) for B2GP-PLYP geometries. As can be observed, the datapoints



for both geometries nearly overlap, demonstrating that B2GP-PLYP  $S_0$  minimum energy geometries are sufficient for performing the  $E^{\text{vert}}$  calculations here. This is supported by an average root-mean-square deviation (RMSD) of  $<0.01$  Å between geometries using these two different methods, and a maximum RMSD of  $<0.03$  Å.

The various EOM-CC methods are shown in Figure 2 in different colours: CCSDR(3) in gray (●), CCSD in blue (●), CC2 in green (●), and STEOM in red (●). It is clear that  $E^{\text{vert}}$  energies deviate from the CCSDR(3) results based upon the EOM-CC method used, and in systematically similar ways across classes of carbonyl species. EOM-CCSD overestimates  $E^{\text{vert}}$ , though the overestimation is larger for  $\alpha, \beta$ -unsaturated carbonyls than for saturated species. CC2 displays an overestimation for saturated species, but an underestimation for  $\alpha, \beta$ -unsaturated species. Both these results indicate the saturated species and  $\alpha, \beta$ -unsaturated carbonyl species must be treated differently in correlating against higher-level EOM-CC results, due to the different delocalisation environment at the chromophore. Finally, the STEOM energies, despite being used by the authors in for earlier work in which the  $S_1$  energies weren't as critical,<sup>1</sup> are seen in Figure 2 to not systematically follow the CCSDR(3)//CCSD energies and so are unlikely to be sufficient for applications where accurate  $S_1$  energies are important — such as the accessible  $S_1$  Norrish Type II reaction in carbonyls.<sup>75–78</sup>

CC3//CCSD calculations were only achievable for six molecules: formaldehyde, acetaldehyde, acrolein, acetone, glycolaldehyde, and glyoxal, and are shown as crosses (×). Additionally, near-exact exFCI  $E^{\text{vert}}$  values are available for formaldehyde and acetaldehyde,<sup>13</sup> and are shown as asterisks (\*). The CCSDR(3)//CCSD diagonal in Figure 2 is an overestimation of these higher-level  $E^{\text{vert}}$  results, however the methods are so computationally demanding, and data on carbonyl species so few, that for the practical photochemical protocol desired here these highest-level results cannot be used to correlate. Instead, it is noted that the EOM-CC  $E^{\text{vert}}$  energies without triples contributions show the most

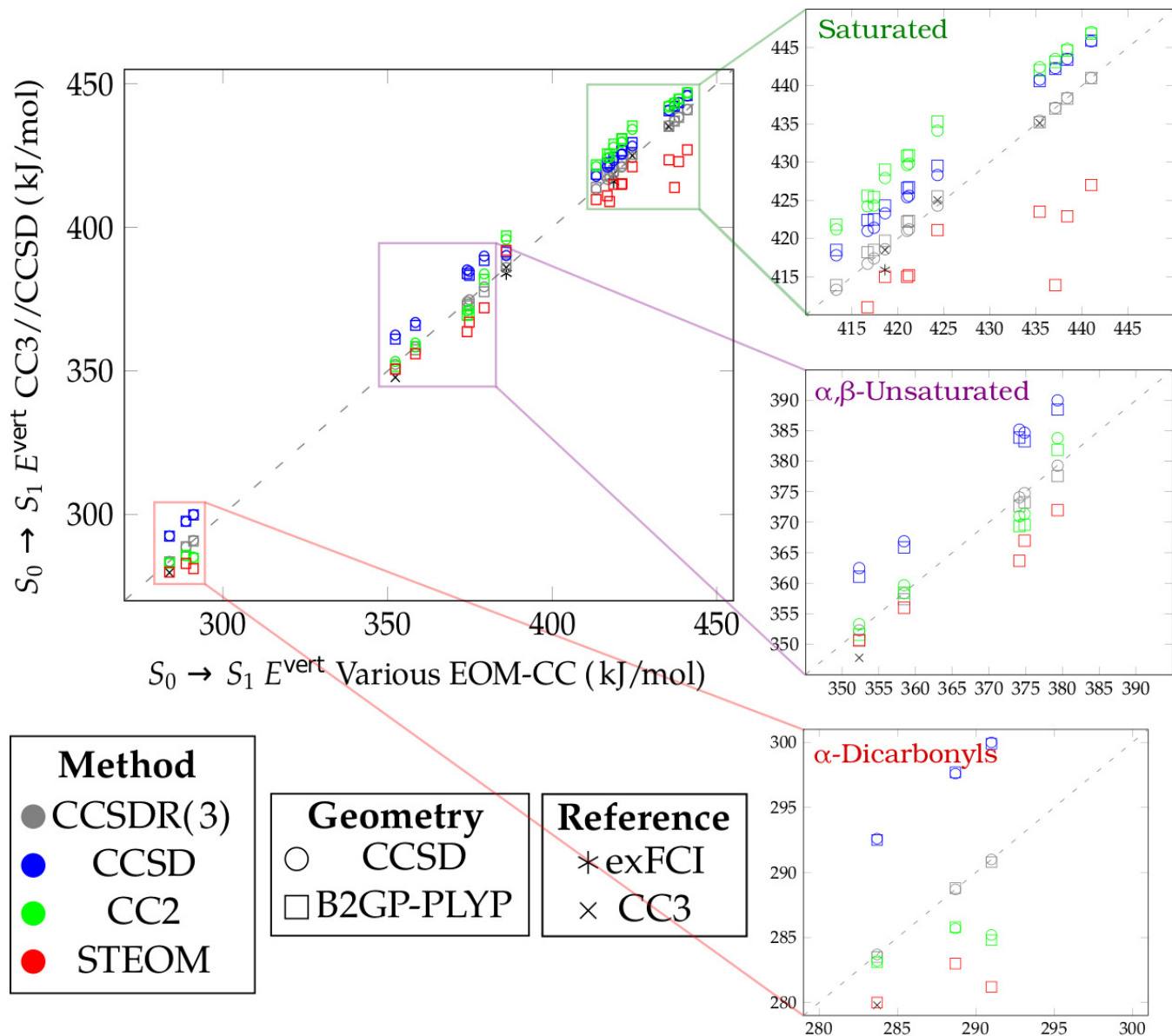
deviation, and differ from CCSDR(3) more than CCSDR(3) differs from available highest-level results.

For this current work, CCSDR(3)//CCSD results are considered here to form the ‘benchmark’  $E^{\text{vert}}$  energies achievable across all “small” carbonyls, to which lower level of theories can be compared and practical methods for calculating  $E^{\text{vert}}$  for “large” carbonyls can be determined. The inclusion of the triples contribution is thus important for accurate carbonyl  $E^{\text{vert}}$  predictions, to correct for overestimates when only doubles contributions are included. Indeed, in a recent cross-comparison of TD-DFT and CC excitation energies<sup>79</sup> Suellen *et al.* emphasize that “*only CC methods including contributions for the triples are accurate enough to offer a safe reference to benchmark TD-DFT*”.

The perturbative triples correction in CCSDR(3) has previously been demonstrated to work well for excited states dominated by single excitations,<sup>53,80</sup> and the Table S3 shows that the % singles for all species is  $>90\%$  and the  $||T_1/T_2||$  norm of the CCSD wavefunction is always above 2.0. Table S3 also contains the energy change upon inclusion of triples ( $\Delta E$  triples), which tends to be  $\sim 4\text{--}5$  kJ/mol for saturated species, and  $\sim 8\text{--}13$  kJ/mol for  $\alpha, \beta$ -unsaturated species. When CC3 energies are achievable, they vary from CCSDR(3) by  $<1$  kJ/mol for saturated species, but for  $\alpha, \beta$ -unsaturated species the differences between CCSDR(3) and CC3 is close to  $3\text{--}4$  kJ/mol. This provides cautionary data about using perturbative triples when studying  $\alpha, \beta$ -unsaturated chromophores, and accords with the higher % doubles observed for  $\alpha, \beta$ -unsaturated species ( $\sim 9\text{--}10\%$ ) compared to saturated species ( $\sim 7\text{--}8\%$ ).

Most noticeable in Figure 2 is the near systematic deviations of most EOM-CC methods from CCSDR(3)//CCSD diagonal. This indicates that when an approximation of the triples contribution to  $E^{\text{vert}}$  is needed for a molecule where CCSDR(3) calculations may be too expensive (*e.g.* hexanal), a lower-level EOM-CC method can be used and then scaled to account for the deviation to CCSDR(3) according to this data on 20 smaller carbonyl species.





**Figure 2:** Correlation between calculated  $E^{\text{vert}}$  values calculating using various levels of EOM-CC theory. The CCSDR(3)//CCSD results are the diagonal “benchmark” values here, as the highest consistently achievable level of theory. Saturated and unsaturated species behave differently: the  $\alpha,\beta$ -unsaturated  $E^{\text{vert}}$  energies are most sensitive to the level of theory.

### Unscaled vertical excitation energies

For brevity, only the  $E^{\text{vert}}$  data for small carbonyl are shown here in the correlation plot of Figure 2, but bar plots of each  $E^{\text{vert}}$  energy is shown in the supporting information, organised by carbonyl class and displayed for each individual molecule. Experimental  $E^{\text{FC}}$  energies (from  $\lambda_{\text{max}}$ ), where available, are also displayed alongside the calculated  $E^{\text{vert}}$  data, lending assurance that the correct excitation has been modelled, and demonstrating that  $E^{\text{FC}}$  differs as an estimate of  $E^{\text{vert}}$  by about 10–15 kJ/mol, despite

not being directly comparable quantities.

**Consistency of  $E^{\text{vert}}$  within carbonyl classes** Excitation energies are consistent within carbonyl structural classes. Saturated carbonyl  $E^{\text{vert}}$  energies fall into a 415–445 kJ/mol range, with the exception in the formaldehyde in the 385–400 kJ/mol range. Ketones have consistently  $\sim 20$  kJ/mol higher  $E^{\text{vert}}$  energies, compared to the equivalent aldehyde, in keeping with the effects of their  $\alpha$ -alkyl auxochromes. Alkyl chain extension is predicted to

have only a marginal effect on  $E^{\text{vert}}$ . This can be seen by comparing the  $E^{\text{vert}}$  values of acetaldehyde to pentanal in Figure S2, or acetone to pentan-2-one in Figure S4. Chain branching is also predicted to lead to  $E^{\text{vert}}$  differences of only a few kJ/mol, as can be seen by comparing the  $E^{\text{vert}}$  of 2-methylpropanal and propanal, as well as pivaldehyde and 2-methylpropanal (Figures S2 and S3). Different functional groups in the carbonyl molecule do not dramatically alter  $E^{\text{vert}}$ , provided they are sufficiently separated from the C=O moiety. This is shown in the similar  $E^{\text{vert}}$  energies for glycolaldehyde (Figure S3) and the corresponding unfunctionalised aldehyde, acetaldehyde (Figure S2). These results reinforce that  $E^{\text{vert}}$  values are consistent within carbonyl classes at a given level of EOM-CC theory.

### Scaling factors for EOM-CC calculations

CCSDR(3)//CCSD  $E^{\text{vert}}$  energies could be practically calculated for all only the 20 “small” carbonyls, whereas excitation energies for the additional 18 “large” carbonyls could only be computed using lower levels of theory. Again, for brevity bar plots of the individual energies for “large” carbonyls are relegated to the supporting information.

Ideally, scaling factors would be determined such that these lower-level EOM-CC results can be scaled to an energy which approximates the effects of triples contributions. The different trends in EOM-CC  $E^{\text{vert}}$  predictions between saturated and unsaturated species necessitates different scaling factors for those two classes.

These desired scaling factors are determined from a linear regression between the lower-levels of theory and the CCSDR(3)//CCSD  $E^{\text{vert}}$  values, and are reported in Table S6 of the supporting information. A single “Total” scaling factor is reported for regression across all 20 species in the “small” carbonyl dataset. Scaling factors for saturated carbonyls were determined by linear regression across the 12 saturated carbonyl species, and for unsaturated carbonyls across the 8  $\alpha$ ,  $\beta$ -unsaturated species (including  $\alpha$ -dicarbonyls) in the “small” carbonyl dataset. The standard error of the slope of each regression line is also reported in Table S6.

In Table S6 the only scaling factors significantly above 1.0 are the STEOM scaling factors, indicating correction for the underestimated STEOM  $E^{\text{vert}}$  energies observed in Figures 2 for all carbonyls. These STEOM scaling factors also have among the largest standard errors in linear regression, indicating a significant amount of intrinsic error in the method, which was also observed in error distributions in  $E^{\text{vert}}$  benchmarks by Loos *et al.*<sup>13</sup> The CCSDR(3)//B2GP-PLYP scaling factors are all close to 1.0, with small standard errors, demonstrating that the effect of CCSD *vs.* B2GP-PLYP geometries is minor. This is also reflected in the fact that scaling factors for a given level of EOM-CC theory are effectively identical for CCSD and B2GP-PLYP geometries, most to within the standard errors of the slopes.

The CC2 and CCSD scaling factors differ significantly for saturated and unsaturated carbonyls, in keeping with the increased importance of triples contributions in unsaturated carbonyls. For CC2, the unsaturated scaling factors are slightly above 1.0 (1.01), whereas for saturated carbonyls they are less than 1.0 (0.98). In contrast, CCSD scaling factors are smaller for unsaturated carbonyls than for saturated carbonyls. This indicates that the use of approximate *vs.* “full” iterative doubles contributions results in different errors for these two carbonyl classes that have qualitatively different frontier orbitals. In particular, the scaling corrects the  $E^{\text{vert}}$  overestimate by CCSD for unsaturated carbonyl energies with a scaling factor less than 0.98. The use of a “Total” scaling factor over all species appears to be a poor compromise between values appropriate for saturated and  $\alpha$ ,  $\beta$ -unsaturated species.

**Errors of scaled EOM-CC energies** The  $E^{\text{vert}}$  energies obtained after applying the scaling factors in Table S6 are compared to the unscaled ‘benchmark’ CCSDR(3)//CCSD energies across the “small” carbonyl dataset. Table 1 reports the mean unsigned error (MUE), mean signed error (MSE), and maximum absolute error (MaxAE), of scaled and unscaled energies with reference to unscaled CCSDR(3)//CCSD energies. All scaled  $E^{\text{vert}}$  energies are reported

in the supporting information in Tables S7 and S8, and included in the bar plots for individual molecules.

**Table 1:** Deviation of scaled EOM-CC vertical excitation energies relative to CCSDR(3)//CCSD values. The mean unsigned error (MUE), mean signed error (MSE), and maximum absolute error (MaxAE) obtained using each scaling factor is reported. (kJ/mol)

Geometry Energy	B2GP-PLYP				CCSD	
	<i>bt</i> -STEOM	CC2	CCSD	CCSDR(3)	CC2	CCSD
Unscaled						
MUE	7.7	6.2	6.6	0.9	5.8	6.5
MSE	-7.1	4.7	6.6	0.1	4.8	6.5
MaxAE	23.2	11.0	9.8	1.7	9.8	11.1
Saturated & unsaturated scaling factors (Table S6)						
MUE	4.3	2.1	0.6	0.6	1.9	0.4
MSE	0.0	0.0	0.1	0.0	-0.1	0.0
MaxAE	15.2	5.0	1.7	1.0	5.6	1.7

Using a single “Total” scaling factor is an improvement to unscaled energies, and using separate scaling factors for saturated and unsaturated species reduces the error further still. From Table 1, scaling  $E^{\text{vert}}$  with the appropriate factor in Table S6 improves the agreement to CCSDR(3)//CCSD energies. The MUE of CCSDR(3)//B2GP-PLYP energies to CCSDR(3)//CCSD energies is less than 1 kJ/mol with or without the use of scaling factors, though using both saturated and unsaturated scaling factors also decreases the MaxAE to 1 kJ/mol. This indicates that if CCSDR(3)//B2GP-PLYP  $E^{\text{vert}}$  calculations can be achieved for a molecule, scaling is unimportant.

The general underestimation delivered by unscaled STEOM  $E^{\text{vert}}$  calculations (MSE -7.1 kJ/mol) is corrected through scaling (MSE 0.0 kJ/mol), however the MUE remains  $\sim 4$  kJ/mol, and the MaxAE is still unacceptably large after scaling, at  $\sim 15$  kJ/mol. From this, it is demonstrated that while STEOM is a computationally efficient method that delivers improved  $E^{\text{vert}}$  energies compared to many TD-DFT methods, the intrinsic error in the method means that STEOM energies cannot be effectively scaled to estimate the inclusion of triples contributions.

For CC2 and CCSD  $E^{\text{vert}}$  energies, the unscaled MUEs of 6–7 kJ/mol are greatly decreased by scaling. Scaling with a single “Total”

scaling factor reduces the MUE of CC2 energies to  $\sim 4$  kJ/mol, though MaxAEs of  $\sim 10$  kJ/mol are still observed. For CCSD, using the “Total” scaling factor decreases MUEs to an acceptable  $\sim 2$  kJ/mol, though MaxAEs are still  $>4$  kJ/mol. The CCSD  $E^{\text{vert}}$  energies in Figures S2–S6 showed the greatest difference in energetic trends between saturated and unsaturated carbonyls, and so these CCSD energies benefit most from using two separate scaling factors. The CCSD  $E^{\text{vert}}$  energies scaled by the saturated and unsaturated scaling factors show MUEs of  $\sim 0.5$  kJ/mol with both CCSD and B2GP-PLYP  $S_0$  geometries, and the resulting negligible MSE errors indicate no systematic overestimation or underestimation across the “small” carbonyl dataset. The 1.7 kJ/mol MaxAEs of scaled CCSD  $E^{\text{vert}}$  energies, found for both B2GP-PLYP and CCSD geometries, means CCSD energies can be reliably scaled to approximate CCSDR(3)//CCSD ‘benchmark’  $E^{\text{vert}}$  energies.

Since  $E^{\text{vert}}$  energies scaled using both a saturated and an unsaturated scaling factor deliver much lower deviations, these energies are the only scaled  $E^{\text{vert}}$  energies discussed below. We report in Table 2 below our best estimates for the  $E^{\text{vert}}$  energies of all 38 carbonyl species studied here. The difference between the scaled CCSD//B2GP-PLYP energies and the CCSDR(3)//CCSD values they were scaled against are notably highest for the complex  $\alpha$ ,  $\beta$ -unsaturated carbonyls. Since higher-level calculations on larger carbonyls were not practical, the scaled CCSD//B2GP-PLYP energies are reported as the best estimates for  $E^{\text{vert}}$  in these 18 species.

Having achieved estimates of  $E^{\text{vert}}$  for larger carbonyl species, with some measure of triples correction through scaling CCSD energies, we turn our attention to the other quantities in Figure 1 necessary for calculating  $E^{0-0}$ .

## $S_1$ Relaxation energies — $E^{\text{relax}}$

As shown in Figure 1, the adiabatic energy,  $E^{\text{adiabatic}}$ , is the difference of the single point energies at the  $S_0$  and  $S_1$  minima, without ZPVE correction:  $E^{\text{adiabatic}} = E^{\text{vert}} - E^{\text{relax}}$ . Ideally,

**Table 2:** Best estimates of  $E^{\text{vert}}$  for each carbonyl.

<i>EOM-CC singlepoint Geometry</i>	<i>Method</i>		<i>Diff.</i>
	CCSDR(3) CCSD	Scaled CCSD B2GP-PLYP	
<i>Small carbonyls</i>	$E^{\text{vert}}$ (kJ/mol)		
Formaldehyde	386.0	386.5	0.5
Acetaldehyde	418.6	419.1	0.5
Propanal	413.3	413.3	0.0
Butanal	421.0	421.3	0.3
Pentanal	421.2	421.4	0.2
2-Methylpropanal	416.7	417.2	0.5
Pivaldehyde	417.4	417.3	-0.1
Acetone	435.4	435.1	-0.3
Butanone	438.4	437.9	-0.5
Pentan-2-one	437.1	436.7	-0.4
Pentan-3-one	441.0	440.3	-0.7
Glycolaldehyde	424.3	424.2	-0.1
Acrolein	352.3	352.0	-0.3
Methacrolein	358.4	356.8	-1.6
Crotonaldehyde	379.3	378.8	-0.5
MVK <sup>a</sup>	374.1	374.2	0.1
MIPK <sup>b</sup>	374.8	373.6	-1.2
Glyoxal	283.7	285.1	1.4
Methylglyoxal	288.7	290.3	1.7
Diacetyl	291.0	292.4	1.4
<b>MUE</b>			<b>0.6</b>
<b>MSE</b>			<b>0.1</b>
<i>Large carbonyls</i>			
2-Methylpentanal	—	418.0	—
3-Methylpentanal	—	412.2	—
4-Methylpentanal	—	421.5	—
2-Ethylbutanal	—	420.6	—
3-3-Dimethylbutanal	—	403.7	—
Hexanal	—	419.8	—
2-Methylpent-2-enal	—	384.7	—
3-Methylpent-2-enal	—	367.2	—
Hex-2-enal	—	377.4	—
Hex-3-enal	—	410.7	—
Hexan-2-one	—	437.6	—
Hexan-3-one	—	439.1	—
3-Methylpentan-2-one	—	428.4	—
4-Methylpentan-2-one	—	435.0	—
Hex-3-en-2-one	—	378.4	—
Hex-1-en-3-one	—	375.9	—
Hex-4-en-3-one	—	384.1	—
2-Oxobutanal	—	283.8	—

<sup>a</sup> Methyl vinyl ketone.<sup>b</sup> Methyl isopropenyl ketone.

both calculations would be done with an EOM-CC level of theory that includes triples, whose importance was demonstrated above. However, the inclusion of triples for  $E^{\text{vert}}$  calculations was only feasible for  $S_0$  minima of species in the “small” carbonyl dataset. The pyramidalisation of the carbonyl moiety in the  $S_1$  excited state reduces symmetry and hence makes CC-triples calculations on  $S_1$  minima infeasible. Since de-

termination of  $E^{\text{relax}}$  involves subtraction of two  $S_1$  energies, it may be hoped that the systematic biases intrinsic to any given excited state method will cancel, whereas  $E^{\text{vert}}$  needed a high level of theory and could not rely on error cancellation. If the method used for  $E^{\text{relax}}$  is affordable, the same level of theory can be used to optimise  $S_1$  transition states, and determine  $S_1$  ZPVEs, allowing straightforward calculation of an intrinsic reaction barrier relative to the  $S_1$  minimum.

$E^{\text{relax}}$  energies are reported in Table 3.  $E^{\text{relax}}$  energies are calculated both using same TD-DFT method as the optimisation (*e.g.* B3LYP//B3LYP), and those where TD-DFT geometries were used for EOM-CCSD single point energies (*e.g.* CCSD//B3LYP). The highest level of theory practical for the “small” carbonyl dataset was CCSD//CCSD, although EOM-CCSD  $S_1$  optimisation was infeasible for species with five carbons: pentanal, pivaldehyde, pentan-2-one, and pentan-3-one. The MUE of the  $E^{\text{relax}}$  energies calculated by each method is reported at the bottom of Table 3, relative to CCSD//CCSD values.

**Table 3:**  $S_1$  relaxation energies ( $E^{\text{relax}}$ ) for species in the “small” carbonyl dataset, calculated at various levels of theory, as indicated. (kJ/mol)

<i>Singlepoint</i>	B3LYP	CAM-B3LYP	CCSD	CCSD	CCSD
<i>//</i>	<i>//</i>	<i>//</i>	<i>//</i>	<i>//</i>	<i>//</i>
<i>Geometry</i>	B3LYP	CAM-B3LYP	B3LYP	CAM-B3LYP	CCSD
Formaldehyde	29	25	35	36	31
Acetaldehyde	43	39	49	51	48
Propanal	39	36	46	48	45
Butanal	45	41	51	53	50
Pentanal	45	43	52	56	—
2-Methylpropanal	41	35	48	51	38
Pivaldehyde	41	39	48	51	—
Acrolein	33	26	26	33	31
Methacrolein	34	26	25	32	31
Crotonaldehyde	34	26	25	33	42
Acetone	50	47	57	60	57
Butanone	51	48	57	61	58
Pentan-2-one	46	41	49	52	—
Pentan-3-one	52	49	58	61	—
MVK	48	40	40	45	46
MIPK	45	39	40	46	42
Glycolaldehyde	76	69	73	76	72
Glyoxal	8	7	7	8	7
Methylglyoxal	16	13	12	14	11
Diacetyl	18	15	15	16	13
<b>MUE<sup>a</sup></b>	<b>4.2</b>	<b>6.1</b>	<b>3.6</b>	<b>3.7</b>	<b>—</b>

<sup>a</sup> Mean unsigned error compared to CCSD//CCSD

The MUE with respect to CCSD//CCSD of the purely DFT  $E^{\text{relax}}$  values is  $\sim 4$  kJ/mol for B3LYP and  $\sim 6$  kJ/mol for CAM-B3LYP. The MUE with respect to CCSD//CCSD  $E^{\text{relax}}$  de-

creases to 3.6 kJ/mol and 3.7 kJ/mol when CCSD single point energy calculations are performed on B3LYP and CAM-B3LYP geometries, respectively. Aside from crotonaldehyde and 2-methylpropanal, the CCSD//DFT  $E^{\text{relax}}$  energies differ from the CCSD//CCSD values by  $\leq 5$  kJ/mol. Crotonaldehyde and 2-methylpropanal display  $E^{\text{relax}}$  differences between methods that are  $>10$  kJ/mol. Using ‘doubles-quality’ EOM-CC calculations was also seen to be problematic for  $E^{\text{vert}}$  values of crotonaldehyde because of its larger delocalised  $\pi$ -system.

In comparison to the CC3//CCSD  $E^{\text{relax}}$  energies calculated by Loos and Jacquemin to obtain ‘chemically accurate’  $E^{0-0}$  energies through error cancellation, the CCSD//CCSD  $E^{\text{relax}}$  energies lack triples contributions and are all smaller than CC3//CCSD energies by: 10 kJ/mol for formaldehyde, 9 kJ/mol for acetaldehyde, 11 kJ/mol for acetone, 14 kJ/mol for acrolein, and 4 kJ/mol for glyoxal.<sup>58</sup> Consequently, the combination of the best  $E^{\text{vert}}$  and  $E^{\text{relax}}$  energies here will lead to an overestimate of the  $E^{0-0}$  values of Loos and Jacquemin — while the scaled  $E^{\text{vert}}$  energies in Table 2 are able to match CC3//CCSD  $E^{\text{vert}}$  energies, the CCSD//CCSD  $E^{\text{relax}}$  energies in Table 3 underestimate the CC3//CCSD values.

The MUEs of  $\sim 4\text{--}6$  kJ/mol in Table 3 indicate that  $E^{\text{relax}}$  will be a major source of error in the  $E^{0-0}$  energy prediction. It is hoped the underestimation of  $E^{\text{relax}}$  due to the lack of triples contributions is systematic and can, potentially, be accounted for. The final component in an  $E^{0-0}$  prediction is the  $\Delta E^{\text{ZPVE}}$  energy, described below.

## $S_1 - S_0$ Zero-point vibrational energy differences — $\Delta E^{\text{ZPVE}}$

As shown in Figure 1,  $E^{0-0}$  can be obtained by applying ZPVE correction to  $E^{\text{adiabatic}}$ :  $E^{0-0} = E^{\text{adiabatic}} + \Delta E^{\text{ZPVE}}$ . While frequency calculations for ZPVEs are the most computationally demanding step, several publications show that  $\Delta E^{\text{ZPVE}}$  is particularly insensitive to the DFT functional used.<sup>58,81,82</sup> The MUE between different DFT functionals has been found to be less than 1 kJ/mol, when a triple- $\zeta$  or larger basis

set is used.<sup>10</sup> The same DFT methods used for  $S_1$  geometry optimisations should therefore be suitable for evaluating  $\Delta E^{\text{ZPVE}}$ . Table 4 below reports  $\Delta E^{\text{ZPVE}}$  energies of ‘small’ carbonyls, calculated from geometry optimisations and frequency calculations using the B3LYP, CAM-B3LYP, and M06-2X functionals. All ZPVEs have been scaled with the appropriate (ground state) scaling factor.

**Table 4:** Zero-point vibrational energy differences,  $\Delta E^{\text{ZPVE}}$ , between the  $S_0$  and  $S_1$  minimum energy geometries, calculated using different density functionals. (kJ/mol)

Method	B3LYP	CAM-B3LYP	M06-2X
Formaldehyde	-8.3	-8.9	-9.0
Acetaldehyde	-7.0	-7.0	-7.2
Propanal	-8.0	-7.5	-7.4
Butanal	-7.6	-7.3	-7.5
Pentanal	-6.8	-5.8	-7.9
2-Methylpropanal	-8.4	-8.1	-7.9
Pivaldehyde	-8.5	-8.0	-8.3
Acrolein	-7.2	-7.5	-7.9
Methacrolein	-7.5	-7.8	-8.9
Crotonaldehyde	-9.1	-8.6	-8.4
Acetone	-6.2	-5.7	-5.2
Butanone	-6.4	-6.0	-5.5
Pentan-2-one	-7.1	-5.7	-5.4
Pentan-3-one	-6.1	-4.8	-6.1
MVK	-5.9	-5.4	-6.3
MIPK	-7.2	-6.5	-6.5
Glycolaldehyde	-6.5	-6.3	-6.1
Glyoxal	-6.0	-6.2	-7.5
Methylglyoxal	-4.9	-5.1	-6.3
Diacetyl	-4.8	-5.2	-5.2
Comparison	B3LYP CAM-B3LYP	B3LYP M06-2X	CAM-B3LYP M06-2X
MSE	<b>0.3</b>	<b>0.0</b>	<b>-0.4</b>
MUE	<b>0.5</b>	<b>0.7</b>	<b>0.6</b>
MaxAE	<b>1.4</b>	<b>1.7</b>	<b>2.1</b>

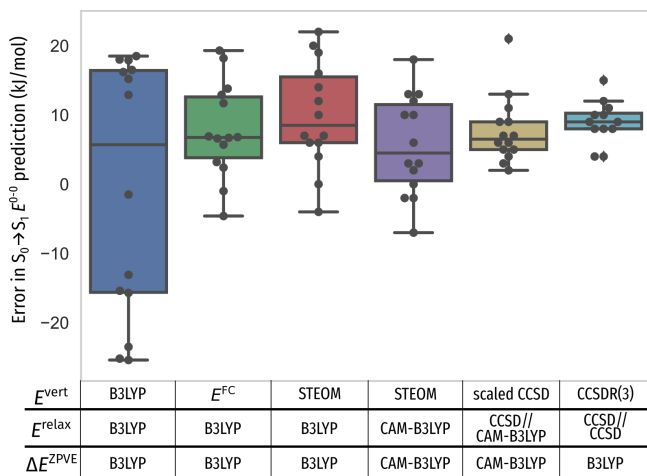
The MUEs between  $\Delta E^{\text{ZPVE}}$  energies evaluated using different functionals are all  $\sim 0.5$  kJ/mol, and the MaxAEs range from 1.4–2.1 kJ/mol. Pentan-2-one has the largest MaxAE for two comparisons: between B3LYP and CAM-B3LYP (1.4 kJ/mol), and B3LYP and M06-2X (1.7 kJ/mol). Pentanal has the largest MaxAE for one comparison: between CAM-B3LYP and M06-2X (2.1 kJ/mol). These molecules are among the largest of the ‘small’ carbonyl dataset, and so it can be expected that the magnitude of deviations in  $\Delta E^{\text{ZPVE}}$  would be greater as they contain more normal modes. Since the MaxAEs using different functionals are all  $\sim 2$  kJ/mol or less, it can be concluded that  $\Delta E^{\text{ZPVE}}$  is not a major source of error in  $E^{0-0}$ .

# $S_0 \rightarrow S_1$ 0-0 Excitation energies — $E^{0-0}$

It now remains to assess how well these component energies combine to provide an  $E^{0-0}$  energy prediction. The  $E^{0-0}$  energy is the first quantity directly comparable between theory and experiment, whereas other components had to be assessed relative to higher-level theory. The only published procedure able to achieve chemically accurate  $E^{0-0}$  predictions without relying on compensation of errors requires CC3 calculations for both energies and geometries.<sup>21</sup> In the current paper, the components of  $E^{0-0}$  are calculated with lower levels of theory than CC3 in order to access larger molecules, so it can be expected these  $E^{0-0}$  predictions will have larger error distributions and be reliant on error cancellation. The scaled  $E^{\text{vert}}$  energies here provide good agreement with “CC-triples” quality energies to within 1 kJ/mol, and  $\Delta E^{\text{ZPVE}}$  is relatively insensitive to level of theory, so the level of theory used for  $E^{\text{relax}}$  is expected to be the remaining bottleneck in terms of computational cost and accuracy.

The error distributions of  $S_0 \rightarrow S_1$   $E^{0-0}$  predictions for the “small” carbonyls, using different protocols, are shown with box and whisker plots in Figure 3. The 14 experimental  $S_0 \rightarrow S_1$   $E^{0-0}$  energies of Godunov and Yakovlev from Table S1 are used for the comparison to the theoretical results, and the raw error data is overlaid as black dots. The error data are horizontally offset to avoid overlapping data points. The components used in each  $E^{0-0}$  protocol are shown in the column underneath. The first column shows the error distribution when all components are calculated with (TD-)DFT (B3LYP), and the second when the  $E^{\text{vert}}$  is replaced with the  $E^{\text{FC}}$  energy from an available  $\lambda_{\text{max}}$  measurement (Table S2). When CCSD//CCSD is used for  $E^{\text{relax}}$  only 12 data points are available, since  $S_1$  minima were unable to be optimised with EOM-CCSD two “small” carbonyls containing five carbon atoms.

From Figure 3 it can be seen that purely B3LYP calculations of the components in  $E^{0-0}$  are insufficient for accurate predictions. The error spread of this protocol is large, leading to absolute errors almost double the 10 kJ/mol



**Figure 3:** Box and whisker plots of errors in  $S_0 \rightarrow S_1$   $E^{0-0}$  predictions using different protocols. The method used for each component in  $E^{0-0}$  is shown below each column.

accuracy desired. The error spread can be significantly reduced by using more accurate methods for  $E^{\text{vert}}$ . A simple composite scheme was used in previous work by the authors,<sup>1</sup> where  $E^{\text{vert}}$  is replaced with an inexpensive EOM-CC method (STEOM) and all other quantities were calculated with (TD-)DFT. This procedure (columns 3 and 4 in Figure 3) significantly decreases the error distribution, regardless of whether CAM-B3LYP or B3LYP is used. Use of  $E^{\text{FC}}$  for  $E^{\text{vert}}$  (column 2 in Figure 3) is also an improvement over pure TD-DFT calculations, providing reasonable and computationally inexpensive estimates of  $E^{0-0}$  for carbonyls when a  $\lambda_{\text{max}}$  measurement is available.

In terms of *de novo*  $E^{0-0}$  prediction, use of the highest level of theory components (CCSDR(3)  $E^{\text{vert}}$  + CCSD//CCSD  $E^{\text{relax}}$  + B3LYP  $\Delta E^{\text{ZPVE}}$ , column 6) results in a narrow error distribution, however the errors are clustered around an  $\sim 9$  kJ/mol overestimate. This highest-level protocol is impractical for “large” species. When the computational cost is lowered through use of scaled  $E^{\text{vert}}$  energies and CCSD//DFT  $E^{\text{relax}}$  values, the error spread is still significantly lower than the other protocols. This can be seen by comparing column 5 in Figure 3 with the four leftmost columns.

The scaled CCSD protocol in column 5 of Figure 3 delivers an outlier for crotonaldehyde, with an error above 20 kJ/mol. As mentioned

above, extended  $\pi$ -bonded chromophores such as in crotonaldehyde are not well captured with perturbative triples estimates, and had the least accurate scaled  $E^{\text{vert}}$  energies. Crotonaldehyde also displayed one of the largest  $E^{\text{relax}}$  discrepancies between CCSD and DFT energies in Table 3.

Though use of higher-level methods decreases the error distributions in  $E^{0-0}$  predictions, examination of Figure 3 shows that all protocols, on average, overestimate the experimental  $E^{0-0}$  energies. The use of scaled  $E^{\text{vert}}$  energies has not lowered the median error significantly compared to a purely B3LYP protocol, as seen upon examination of the median lines in the box plots of Figure 3. However, without triples contributions (or scaling to triples) in  $E^{\text{vert}}$ , both under- and over-estimation occur as seen with TD-DFT and STEOM calculations. When  $E^{\text{vert}}$  includes the effects of triples then all errors are systematic overestimations, which was expected in the discussion of  $E^{\text{relax}}$  since compensation of errors could not be achieved because the  $E^{\text{relax}}$  values here lack triples contributions.

Since this overestimation of  $E^{0-0}$  appears systematic, an empirical energy shift correction can be used. Since both scaling and empirical correction are used, this  $E^{0-0}$  protocol relies on similarities between carbonyl excitation energies to access larger carbonyl species, and is not presented as generally applicable to other organic molecules. Here, the empirical correction adopted is a single energy shift value for each given protocol, equal to the negative of the MSE for that protocol, applied to the full set of molecules.

While inelegant, empirical correction factors are used to correct systematic deficiencies in composite methods, including the ‘Higher Level Correction’ ( $\Delta\text{HLC}$ ) in the popular  $Gn$  set of composite methods.<sup>83,84</sup> The original  $\Delta\text{HLC}$  correction in  $G1$  was a physically justified correction relating to isogyric reactions.<sup>85</sup> However, in later  $Gn$  protocols the  $\Delta\text{HLC}$  correction became a semi-empirical parameter, providing a molecule-independent correction to improve agreement with experiment.<sup>86</sup>

Theoretical  $E^{0-0}$  energies are reported in Table S11 of the supporting information for all

“small” carbonyls. Error statistics are reported at the bottom of Table S11, values in parentheses are after an empirical shift equal to the negative of the MSE. Without empirical correction, all protocols in Table S11, apart from purely B3LYP calculations, deliver MUEs in the range of 7–11 kJ/mol. The purely B3LYP protocol has the largest MUE (16.8 kJ/mol) and largest MaxAE (26.2 kJ/mol). The MaxAEs of the other methods are comparable, 16–21 kJ/mol. The MSE alone is a poor indicator of accuracy of a  $E^{0-0}$  protocol. For example, the purely B3LYP protocol both under- and over-estimates experimental  $E^{0-0}$  energies, resulting in a small (-0.5 kJ/mol) MSE but a large spread of errors.

It is preferable that the  $E^{0-0}$  contains a systematic error to experiment (that can be taken into account), rather than a large error spread such as was observed for (TD)-B3LYP calculations. Only the two rightmost protocols shown in Figure 3 and Table S11 deliver systematic overestimates, and correspondingly these protocols benefit the most from an empirical energy shift. The purely B3LYP calculations are unable to be improved by a molecule-independent energy shift, modest benefits are seen for the other methods using DFT  $E^{\text{relax}}$  energies, but the two highest-level protocols that have “CC-triple” quality  $E^{\text{vert}}$  achieve MUEs below 4 kJ/mol after empirical correction.

After empirical correction, the highest-level protocol in the rightmost column of Figure 3 achieves a  $\sim 2$  kJ/mol MUE, and a MaxAE below 10 kJ/mol. Since this protocol is not feasible for large carbonyls, the next most accurate protocol must be used to predict  $E^{0-0}$  energies for the “large” carbonyl dataset. This scaled CCSD  $E^{\text{vert}}$  + CCSD//CAM-B3LYP  $E^{\text{relax}}$  + CAM-B3LYP  $\Delta E^{\text{ZPVE}}$  protocol, once empirically corrected, also achieves a MUE under 4 kJ/mol, but the MaxAE is slightly above 10 kJ/mol. The MaxAE of this protocol is for the case of crotonaldehyde, and is again cautionary that the accuracy of predicting  $E^{0-0}$  for large  $\alpha,\beta$ -unsaturated species may be degraded in comparison to saturated species.

Here, the correction to the  $S_0 \rightarrow S_1$   $E^{0-0}$  energy of a composite protocol is molecule-



independent, but has not been benchmarked beyond carbonyl compounds. It is used to correct for coupled cluster triples contributions to the  $S_1$  relaxation energy, which was required in the systematic error correction of the composite protocol of Loos and Jacquemin.<sup>58</sup> Experimental  $E^{0-0}$  data is not available for the species in the “large” carbonyl dataset and so the best estimates of  $E^{0-0}$  must come from computational protocols. The theoretical best estimates for  $E^{0-0}$  of all 38 carbonyl species here can be found in Table S12 of the supporting information.

## Conclusions

Protocols for theoretical  $E^{0-0}$  energy prediction can be made affordable by avoiding the iterative calculation of CC-triples contributions, while still including the effect of triples by scaling EOM-CCSD energies to CCSDR(3) values. Saturated and  $\alpha, \beta$ -unsaturated species have qualitatively different chromophores, but once calculated  $E^{\text{vert}}$  energies scaled separately the MSE to CCSDR(3) is  $\sim 0.5$  kJ/mol with a MaxAE of  $\sim 2$  kJ/mol. Double-hybrid geometries can replace CCSD geometries when calculating  $E^{\text{vert}}$ , and once the resolution of the identity is used double-hybrid DFT calculations come at similar cost to standard DFT methods. The combination of EOM-CCSD//TD-DFT  $S_1$   $E^{\text{relax}}$  relax energies with CC-triples quality  $E^{\text{vert}}$  does lead to overestimation of experimental data, though in an apparently consistent and systematic fashion. The  $\Delta E^{\text{ZPVE}}$  component can be reliably calculated with appropriately scaled (TD-)DFT methods, as long as the excited state is a ‘well-behaved’ valence excitation.

By benchmarking to the available experimental data on carbonyl 0–0 excitation energies,<sup>12</sup> we have determined a protocol to accurately predict theoretical carbonyl  $E^{0-0}$  energies at affordable computational cost. This protocol allows the excited state energies of larger carbonyl species to be predicted accurately in cases where experimental data may be missing. We hope to apply the  $E^{0-0}$  energies in this work to further theoretical studies on larger carbonyl species, including calculation of the photolysis

thresholds and rates needed to model the central impact carbonyls have on earth’s atmospheric chemistry.<sup>87–91</sup>

**Acknowledgement** This work was supported by the Australian Research Council (grant DP160101792). It was also supported by grants of computer time under the Merit Allocation Scheme on the NCI National Facility at the Australian National University, as well as computer time on the computational cluster Katana supported by the Faculty of Science, UNSW Australia, and the computational cluster Artemis supported by the Sydney Informatics Hub at the University of Sydney. KNR recognises an Australian Government Research Training Program (RTP) scholarship.

## Supporting Information Available

The Supporting Information contains: experimentally measured 0–0 excitation energies; measured  $\lambda_{\text{max}}$  values of carbonyls; details of EOM-CC calculations, including point groups and % singles, doubles, and triples; unscaled and scaled  $E^{\text{vert}}$  energies for all carbonyls and EOM-CC methods; plots of  $E^{\text{vert}}$  energies for each carbonyl studied;  $S_1$   $E^{\text{relax}}$  energies;  $S_1/S_0$   $\Delta E^{\text{ZPVE}}$  energies;  $E^{0-0}$  energies calculated by different composite protocols; best theoretical estimates of  $E^{0-0}$  energies for each carbonyl; indicative timings for the  $E^{\text{vert}}$  and  $E^{0-0}$  calculations. (PDF)

Cartesian co-ordinates for all optimised structures (ZIP)

## References

- (1) Rowell, K. N.; Kable, S. H.; Jordan, M. J. T. Structural Effects on the Norrish Type I  $\alpha$ -Bond Cleavage of Tropospherically Important Carbonyls. *The Journal of Physical Chemistry A* **2019**, *123*, 10381–10396, PMID: 31675237.
- (2) Jacquemin, D. What is the Key for Accurate Absorption and Emission Calculations, Energy or Geometry? *J. Chem. Theory Comp.* **2018**, *14*, 1534–1543.

- (3) Laurent, A. D.; Jacquemin, D. TD-DFT benchmarks: A review. *Int. J. Quantum Chem.* **2013**, *113*, 2019–2039.
- (4) Jacquemin, D.; Perpe, E. A. Extensive TD-DFT Benchmark : Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420–2435.
- (5) Send, R.; Kühn, M.; Furche, F. Assessing excited state methods by adiabatic excitation energies. *J. Chem. Theory Comp.* **2011**, *7*, 2376–2386.
- (6) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, 134110.
- (7) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P.; Thiel, W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *J. Chem. Phys.* **2008**, *129*.
- (8) Silva-Junior, M. R.; Sauer, S. P. A.; Schreiber, M.; Thiel, W. Basis set effects on coupled cluster benchmarks of electronically excited states: CC3, CCSDR(3) and CC2. *Mol. Phys.* **2010**, *108*, 453–465.
- (9) Koch, W.; Bonfanti, M.; Eisenbrandt, P.; Nandi, A.; Fu, B.; Bowman, J.; Tanner, D.; Burghardt, I. Two-layer Gaussian-based MCTDH study of the S1  $\leftarrow$  S0 vibronic absorption spectrum of formaldehyde using multiplicative neural network potentials. *J. Chem. Phys.* **2019**, *151*, 064121.
- (10) Fang, C.; Oruganti, B.; Durbeej, B. How method-dependent are calculated differences between vertical, adiabatic, and 0-0 excitation energies? *J. Phys. Chem. A* **2014**, *118*, 4157–4171.
- (11) Goerigk, L.; Moellmann, J.; Grimme, S. Computation of accurate excitation energies for large organic molecules with double-hybrid density functionals. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4436.
- (12) Godunov, I. A.; Yakovlev, N. N.; Khimii, Z. S. Experimental structural and conformational studies of carbonyl molecules in the ground and lower excited states. *J. Struct. Chem.* **1995**, *36*, 238–253.
- (13) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comp.* **2018**, *14*, 4360–4379.
- (14) Krylov, A. I. Equation-of-Motion Coupled-Cluster Methods for Open-Shell and Electronically Excited Species: The Hitchhiker’s Guide to Fock Space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.
- (15) Siegbahn, P. E.; Almlöf, J.; Heiberg, A.; Roos, B. O. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *J. Chem. Phys.* **1981**, *74*, 2384–2396.
- (16) Veryazov, V.; Malmqvist, P. k.; Roos, B. O. How to Select Active Space for Multiconfigurational Quantum. *Int. J. Quantum Chem.* **2011**, *111*, 3329–3338.
- (17) Andersson, K.; Malmqvist, P.; Roos, B. O. Second order perturbation theory with a complete active space self consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.
- (18) Helmich-Paris, B. Benchmarks for Electronically Excited States with CASSCF Methods. *J. Chem. Theory Comp.* **2019**, *15*, 4170–4179.
- (19) Blase, X.; Duchemin, I. Calculations of  $n \rightarrow \pi^*$  Transition Energies: Comparisons Between TD- DFT, ADC, CC, CASPT2, and BSE/ GW Descriptions. **2017**, *121*, 6122–6134.
- (20) Loos, P.-F.; Jacquemin, D. Evaluating 0-0 Energies with Theoretical Tools: a Short Review. *ChemPhotoChem* **2019**, *3*, 684–696.
- (21) Loos, P.-F.; Galland, N.; Jacquemin, D. Theoretical 0–0 Energies with Chemical Accuracy. *J. Phys. Chem. Lett.* **2018**, *9*, 4646–4651.
- (22) Rowell, K. Structure–Activity Relationships for Carbonyl Photolysis. Ph.D. thesis, The University of New South Wales, 2020.
- (23) Shaw, M. F. Photochemical Formation of Enols from Carbonyls. Ph.D. thesis, The University of Sydney, 2017.
- (24) Chan, B.; Radom, L. W1X-1 and W1X-2: W1-quality accuracy with an order of magnitude reduction in computational cost. *J. Chem. Theory Comp.* **2012**, *8*, 4259–4269.
- (25) Adamo, C.; Jacquemin, D. The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chem. Soc. Rev.* **2013**, *42*, 845–856.
- (26) Bloino, J.; Baiardi, A.; Biczysko, M. Aiming at an accurate prediction of vibrational and electronic spectra for medium-to-large molecules: An overview. *Int. J. Quantum Chem.* **2016**, *116*, 1543–1574.

- (27) Goerigk, L.; Grimme, S. Assessment of TD-DFT methods and of various spin scaled CIS(D) and CC2 versions for the treatment of low-lying valence excitations of large organic dyes. *J. Chem. Phys.* **2010**, *132*, 0–9.
- (28) Parac, M.; Grimme, S. Comparison of multireference Møller-Plesset theory and time-dependent methods for the calculation of vertical excitation energies of molecules. *J. Phys. Chem. A* **2002**, *106*, 6844–6850.
- (29) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (30) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (31) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (32) Peach, M. J.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **2008**, *128*, 044118.
- (33) Mardirossian, N.; Parkhill, J. A.; Head-Gordon, M. Benchmark results for empirical post-GGA functionals: Difficult exchange problems and independent tests. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19325–19337.
- (34) Wiggins, P.; Williams, J. A.; Tozer, D. J. Excited state surfaces in density functional theory: A new twist on an old problem. *J. Chem. Phys.* **2009**, *131*.
- (35) Bremond, E.; Savarese, M.; Adamo, C.; Jacquemin, D. Accuracy of TD-DFT Geometries: A Fresh Look. *J. Chem. Theory Comp.* **2018**, *14*, 3715–3727.
- (36) Guido, C. A.; Jacquemin, D.; Adamo, C.; Menucci, B. On the TD-DFT Accuracy in Determining Single and Double Bonds in Excited-State Structures of Organic Molecules. *J. Phys. Chem. A* **2010**, *114*, 13402–13410.
- (37) Guido, C. A.; Knecht, S.; Kongsted, J.; Menucci, B. Benchmarking Time-Dependent Density Functional Theory for Excited State Geometries of Organic Molecules in Gas-Phase and in Solution. *J. Chem. Theory Comp.* **2013**, *9*, 2209–2220.
- (38) Hirata, S.; Head-gordon, M. Time-dependent density functional theory within the Tamm–Dancoff approximation. *Chem. Phys. Lett.* **1999**, *314*, 291–299.
- (39) Petrenko, T.; Kossmann, S.; Neese, F. Efficient time-dependent density functional theory approximations for hybrid density functionals: Analytical gradients and parallelization. *J. Chem. Phys.* **2011**, *134*.
- (40) Peach, M. J.; Williamson, M. J.; Tozer, D. J. Influence of triplet instabilities in TDDFT. *J. Chem. Theory Comp.* **2011**, *7*, 3578–3585.
- (41) Chantzis, A.; Laurent, A. D.; Adamo, C.; Jacquemin, D. Is the Tamm-Dancoff Approximation Reliable for the Calculation of Absorption and Fluorescence Band Shapes? *J. Chem. Theory Comp.* **2013**, *9*, 4517–4525.
- (42) Jacquemin, D.; Perpète, E. A.; Scuseria, G. E.; Ciofini, I.; Adamo, C. TD-DFT performance for the visible absorption spectra of organic dyes: Conventional versus long-range hybrids. *J. Chem. Theory Comp.* **2008**, *4*, 123–135.
- (43) Goerigk, L.; Grimme, S. Double-hybrid density functionals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 576–600.
- (44) Grimme, S.; Neese, F. Double-hybrid density functional theory for excited electronic states of molecules. *J. Chem. Phys.* **2007**, *127*, 154116.
- (45) Kendall, R. A.; Früchtl, H. A. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theo. Chem. Acc.* **1997**, *97*, 158–163.
- (46) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (47) Sancho-García, J. C.; Adamo, C. Double-hybrid density functionals: Merging wavefunction and density approaches to get the best of both worlds. *Phys. Chem. Chem. Phys.* **2013**, *15*, 14581–14594.
- (48) Schwabe, T.; Goerigk, L. Time-dependent double-hybrid density functionals with spin-component and spin-opposite scaling. *J. Chem. Theory Comp.* **2017**, 4307–4323.
- (49) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
- (50) Koch, H.; Christiansen, O.; Jørgensen, P.; Sanchez De Merás, A. M.; Helgaker, T. The CC3 model: An iterative coupled cluster ap-

- proach including connected triples. *J. Chem. Phys.* **1997**, *106*, 1808–1818.
- (51) Kánnár, D.; Szalay, P. G. Benchmarking coupled cluster methods on valence singlet excited states. *J. Chem. Theory Comp.* **2014**, *10*, 3757–3765.
  - (52) Christiansen, O.; Koch, H.; Jørgensen, P. Perturbative triple excitation corrections to coupled cluster singles and doubles excitation energies. *J. Chem. Phys.* **1996**, *105*, 1451–1459.
  - (53) Sauer, S. P.; Schreiber, M.; Silva-Junior, M. R.; Thiel, W. Benchmarks for electronically excited states: a comparison of noniterative and iterative triples corrections in linear response coupled cluster methods: CCSDR(3) versus CC3. *J. Chem. Theory Comp.* **2009**, *5*, 555–564.
  - (54) Kánnár, D.; Tajti, A.; Szalay, P. G. Accuracy of coupled cluster excitation energies in diffuse basis sets. *J. Chem. Theory Comp.* **2017**, *13*, 202–209.
  - (55) Gauss, J.; Stanton, J. F. Analytic first and second derivatives for the CCSDT- $n$  ( $n = 1-3$ ) models: a first step towards the efficient calculation of CCSDT properties. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2047–2060.
  - (56) Stanton, J. F.; Gauss, J. Analytic energy derivatives for the equation-of-motion coupled-cluster method: Algebraic expressions, implementation and application to the S1 state of HFCO. *Theo. Chem. Acc.* **1995**, *91*, 267–289.
  - (57) Budzak, S.; Scalmani, G.; Jacquemin, D. Accurate Excited-State Geometries: A CASPT2 and Coupled-Cluster Reference Database for Small Molecules. *J. Chem. Theory Comp.* **2017**, *13*, 6237–6252.
  - (58) Loos, P.-F.; Jacquemin, D. Chemically Accurate 0-0 Energies with not-so-Accurate Excited State Geometries. *J. Chem. Theory Comput.* **2019**, *15*, 2481–2491.
  - (59) Oruganti, B.; Fang, C.; Durbeej, B. Assessment of a composite CC2/DFT procedure for calculating 0–0 excitation energies of organic molecules. *Mol. Phys.* **2016**, *114*, 3448–3463.
  - (60) Loos, P.-F.; Scemama, A.; Jacquemin, D. The Quest for Highly Accurate Excitation Energies: A Computational Perspective. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2374–2383.
  - (61) Neese, F. Software update: The ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2017**, *8*, 4–9.
  - (62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.
  - (63) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P. et al. The Dalton quantum chemistry program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2014**, *4*, 269–284.
  - (64) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–305.
  - (65) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
  - (66) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
  - (67) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
  - (68) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comp. Chem.* **2011**, *32*, 1456–1465.
  - (69) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H.; Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
  - (70) Kesharwani, M. K.; Brauer, B.; Martin, J. M. Frequency and zero-point vibrational energy scale factors for double-hybrid density functionals (and other selected methods): Can anharmonic force fields be avoided? *J. Phys. Chem. A* **2015**, *119*, 1701–1714.
  - (71) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
  - (72) Weigend, F. Hartree–Fock Exchange Fitting

- Basis Sets for H to Rn. *J. Comp. Chem.* **2008**, *29*, 167–175.
- (73) Hellweg, A.; Hättig, C.; Höfener, S.; Kloppe, W. Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn. *Theo. Chem. Acc.* **2007**, *587*–597.
- (74) Dutta, A. K.; Nooijen, M.; Neese, F.; Izsák, R. Automatic active space selection for the similarity transformed equations of motion coupled cluster method. *J. Chem. Phys.* **2017**, *146*, 074103.
- (75) Wagner, P. J.; Klán, P. *CRC Handbook of Organic Photochemistry and Photobiology*, 2nd Edition; 2004; Chapter 52, pp 1–22.
- (76) Tölgyesi, M.; Nacs, A.; Bérces, T. Quenching of the photochemical decomposition of n-butyraldehyde in the vapour phase and in isooctane. *J. Photochem.* **1979**, *11*, 333–345.
- (77) Nau, W. M.; Greiner, G.; Wall, J.; Rau, H.; Olivucci, M.; Robb, M. A. The Mechanism for Hydrogen Abstraction by  $n, \pi^*$  Excited Singlet States: Evidence for Thermal Activation and Deactivation through a Conical Intersection. *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 98–101.
- (78) Feyter, S. D.; Diau, E. W.; Zewail, A. H. Femtosecond Dynamics of Norrish Type-II Reactions: Nonconcerted Hydrogen-Transfer and Diradical Intermediacy. *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 260–263.
- (79) Suellen, C.; Garcia Freitas, R.; Loos, P.-F.; Jacquemin, D. Cross Comparisons Between Experiment, TD-DFT, CC, and ADC for Transition Energies. *J. Chem. Theory Comp.* **2019**, *4581*–4590.
- (80) Watson, T. J.; Lotrich, V. F.; Szalay, P. G.; Perera, A.; Bartlett, R. J. Benchmarking for perturbative triple-excitations in EE-EOM-CC methods. *J. Phys. Chem. A* **2013**, *117*, 2569–2579.
- (81) Uppsten, M.; Durbeek, B. Quantum chemical comparison of vertical, adiabatic, and 0-0 excitation energies: The PYP and GFP chromophores. *J. Comp. Chem.* **2012**, *33*, 1892–1901.
- (82) Jacquemin, D.; Planchat, A.; Adamo, C.; Menucci, B. TD-DFT assessment of functionals for optical 0-0 transitions in solvated dyes. *J. Chem. Theory Comp.* **2012**, *8*, 2359–2372.
- (83) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.
- (84) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* **1998**, *109*, 7764–7776.
- (85) Wheeler, S. E.; Houk, K. N.; Schleyer, P. V.; Allen, W. D. A Hierarchy of homodesmotic reactions for thermochemistry. *J. Am. Chem. Soc.* **2009**, *131*, 2547–2560.
- (86) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
- (87) Lary, D. J.; Shallcross, D. E. Central role of carbonyl compounds in atmospheric chemistry. *J. Geophys. Res. Atmos.* **2000**, *105*, 19771–19778.
- (88) Vereecken, L.; Glowacki, D. R.; Pilling, M. J. Theoretical Chemical Kinetics in Tropospheric Chemistry: Methodologies and Applications. *Chem. Rev.* **2015**, *115*, 4063–4114.
- (89) Jenkin, M. E.; Saunders, S. M.; Pilling, M. J. The tropospheric degradation of volatile organic compounds: A protocol for mechanism development. *Atmos. Environ.* **1997**, *31*, 81–104.
- (90) Rickard, A.; Young, J. The Master Chemical Mechanism (MCM) v3.2. <http://mcm.leeds.ac.uk/MCM>.
- (91) Goldstein, A. H.; Galbally, I. E. Known and unexplored organic constituents in the earth’s atmosphere. *Environ. Sci. Technol.* **2007**, *41*, 1514–1521.