Density Functional Theory Calculations of Core-Electron Binding Energies at the K-edge of Heavier Elements

Nicholas A. Besley*

School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.

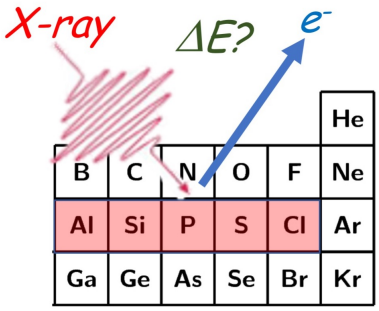
E-mail: nick.besley@nottingham.ac.uk

Abstract

The capability to determine core-electron binding energies (CEBEs) is vital in the analysis of X-ray photoelectron spectroscopy, and the continued development of light sources has made inner shell spectroscopy of heavier elements increasingly accessible. Density functional theory is widely used to determine CEBEs of lighter elements (boron–fluorine). It is shown that good performance of exchange-correlation functionals for these elements does not necessarily translate to the calculation of CEBEs for the heavier elements from the next row of the periodic table, and in general larger errors are observed. Two strategies are explored that improve the accuracy of the calculated CEBEs. The first is to apply element and functional dependent energy corrections and the second is a re-parameterisation of a short-range corrected functional. This functional is able to reproduce experimental phosphorus and sulphur K-edge CEBEs with an average error of 0.15 eV demonstrating the importance of reducing the self-interaction error associated with the core electrons, and represents progress towards

density functional theory calculations that performs equally well for ionisation at the K-edge of all elements.

Table of Contents Graphic



Introduction

The improved capabilities of modern synchrotron sources and free-electron lasers has led to an increasing interest in exploiting X-ray spectroscopy to probe problems in chemistry, physics, biology and materials science.[1–4] X-ray photoelectron spectroscopy (XPS) is one of the most widely used techniques, and is capable of probing the chemical environment of nuclei and informing on factors such as chemical bonding and electronic state. Quantum chemical calculations can play an important role in interpreting and analysing experimental data,[5–9] and these calculations need to be able to reliably distinguish between chemical environments that can be quite similar. Density functional theory (DFT) is the most commonly used method for these calculations since its relatively low computational cost allows it to be applied to study large systems.

The key quantities in simulations of X-ray photoelectron spectra are the core-electron binding energies (CEBEs), which correspond to the ionisation energies of electrons in core orbitals. There are several different approaches to computing CEBEs. The simplest approach is based upon Koopmans' theorem,[10] wherein the CEBE is approximated as the negative of the orbital energy. This approach is commonly used in the context of Kohn-Sham DFT, although it has been shown that Koopmans' theorem does not rigorously apply for Kohn-Sham eigenvalues.[11] The removal of a core-electron leads to significant changes in the valence electronic structure owing to an increase in the effective nuclear charge. Koopmans' theorem does not consider any final state effects, and as a consequence this relaxation of the electron density is not accounted for. A number of methods have been developed that incorporate final state effects with different levels of approximation,[12] such as the GW formalism.[13–16]

In Δself-consistent field (ΔSCF) approaches the CEBE is evaluated directly as the energy difference between the ground and core-ionised states.[17] In this approach, the final ionised state is considered explicitly and the relaxation of the electron density is described fully.

However, this comes at the cost of the need to perform a separate SCF calculation for each CEBE required. Calculation of the core-ionised state requires some method to prevent the variational collapse during the SCF process, but different schemes exist that work well for these states.[18–20] Upon ionisation of a core electron there is generally a contraction of the electron density, particularly around the site of the ionisation, owing to the increase in effective nuclear charge. It has been shown that small or moderately sized basis sets lack sufficient flexibility to describe this change, and can lead to an unbalanced description of the ground and core-ionised states resulting in significant errors in the calculated CEBEs.[21,22] These errors become increasingly significant as the nuclear charge of the ionised nuclei becomes larger, and to achieve good convergence of the calculated CEBEs with respect to basis set completeness requires large basis sets such as cc-pVQZ or cc-pCVTZ.[22] Different strategies have emerged to design moderately sized basis sets that can reproduce the basis set limit CEBEs, these include uncontracting the basis functions,[18,23] augmenting the basis set with basis functions from the element with one greater nuclear charge,[24] basis sets optimized for properties related to core-excitations[25] and multiwavelet formalisms.[8]

For ΔSCF calculations of CEBEs with DFT, large variations in CEBE can arise depending on the choice of exchange-correlation functional, and there is clearly an interest in determining which functionals provide the closest agreement with experiment. Towards this goal, several studies have compared CEBEs evaluated using different functionals with experimental data.[26–34] Chong and co-workers examined a range of functionals for a large set of experimental CEBEs at the C, N, O and F K-edges.[27] It was found that the most accurate values were obtained from a combination of Perdew-Wang (PW86 exchange and PW91 correlation) functionals in conjunction with empirical relativistic correction, yielding a mean absolute deviation (MAD) of 0.16 eV. The good performance of the Perdew-Wang functionals was also observed in another study.[28] More recent studies have revealed good performance of the PBE0 functional for amino acids,[31] the TPSS functional on a set of 68 molecules[30] and

the strongly constrained and appropriately normed (SCAN) non-empirical semi-local meta-GGA exchange-correlation functional in a study considering over 100 molecules.[32] These studies have focused on ionisation at the K-edge of lighter elements (boron to fluorine).

The development of light sources means that it is now possible to probe the deeper core-levels of heavier nuclei more readily. The more tightly bound core electrons of heavier nuclei leads to additional challenges for calculations of CEBEs, including the increased significance of relativistic effects and the breaking of a more strongly correlated electron pair. Another factor in DFT calculations of CEBEs is the self-interaction error associated with the approximate treatment of exchange. For electrons in more compact core-orbitals, the consequences of the self-interaction error are likely to become more significant. As a result, it should not be assumed that computational methods that are successful for CEBEs of lighter elements will also perform well as the nuclear charge increases. There has been relatively few studies that examine the calculation of CEBEs for heavier elements. A notable exception is the work of Segala and Chong who studied the K-edge of sulphur and phosphorus containing molecules.[29] The best performing functionals were found to be Becke00x[35] and BmTau1,[36] which alludes to the fact that more advanced functionals may be required to treat these systems.

In this contribution, the accuracy of different exchange-correlation functionals in DFT ΔSCF calculations of CEBEs are assessed with a focus on ionisation at the K-edge of the heavier elements aluminium to chlorine, and strategies for achieving a high level of accuracy are explored. A range of functionals representing different classes of functional, encompassing more recent developments such as meta-GGA, double hybrid, long-range corrected and short-range corrected functionals, are considered. It is common to assess the accuracy of the different functionals through a comparison with experimental data. This approach is limited by the availability of high-quality experimental gas-phase data which can preclude

a broad and balanced assessment that includes ionisation of the nuclei of elements across the periodic table. This problem becomes particularly acute for heavier elements where the amount of experimental data is more sparse. In this study, the DFT calculations are assessed relative to wavefunction based calculations. This has the additional advantage that it is possible to directly compare non-relativistic calculations and so removing ambiguities that are introduced through the requirement of an accurate treatment of relativistic effects.

Computational Methods

The DFT CEBEs were determined through the difference between the ground and core-ionised state energies ($E_{CEBE} = E_{CI} - E_{GS}$). For the core-ionised state, the ground state orbitals with a vacancy introduced into the relevant core-obital form the initial orbital guess, with the maximum overlap method[37] used to maintain the core-hole in the subsequent SCF calculation. A hybrid basis set comprising cc-pCVTZ for the element being ionised and cc-pVTZ for the remaining elements was used, denoted cc-pCVTZ:cc-pVTZ. Basis sets of this quality have been shown to provide CEBEs which are close to the complete basis set limit for DFT calculations.[22] The DFT calculations are assessed relative to those from ΔMP2 calculations with higher quality basis sets.

The functionals used have been chosen to represent different general classes of functional. B3LYP[38] and PBE0[39] are standard hybrid functionals, the PW86(exchange) + PW91(correlation) combination which has previously been shown to perform well for core-ionization energies,[27] CAM-B3LYP[40] is a long-range corrected functional, SCAN[41] is a meta-GGA functional and B2GP-PLYP[42] is a double hybrid functional. Short-range corrected functionals are also studied. These functionals were developed for time-dependent density functional theory (TDDFT) calculations of X-ray absorption spectra and are based upon a reversal of the principle of long-range corrected functionals and use a high fraction of

Hartree-Fock (HF) exchange in the short range (low $r_{12} = |\mathbf{r}_1\text{-}\mathbf{r}_2|$), and have been parameterised for C–F (SRC2r1) and Si–Cl (SRC2r2).[43,44] The structures of the molecules were optimised using B3LYP/6-311G** and all the calculations were performed with the Q-Chem software.[45]

Results and Discussion

Before examining the DFT calculations, the quality of the reference data from the wavefunction based calculations is explored. There has been considerable progress in the calculation of CEBEs within coupled cluster theory formalisms.[46–50] CEBEs can be determined using equation-of-motion coupled cluster based formalisms.[48] However, recent benchmark studies for CEBEs (C, O and N) have shown that it is necessary to go beyond single and double excitations (CVS-EOM-IP-CCSD) and include triple and quadrupole excitations.[50] The use of these methods for the molecules studied here for the heavier elements with the large basis sets that are necessary to ensure convergence with respect to the basis set are too computationally expensive. Furthermore, these calculations exploit a frozen core orbital which may become significant for heavier nuclei. In this study the DFT calculations are assessed relative to ΔMP2 calculations. These calculations use a hybrid cc-pCVQZ:cc-pVQZ basis set, with the exception of ionisation of fluorine and aluminium where a hybrid cc-pCV5Z:cc-pVQZ basis set is used to ensure covergence. Table 1 shows a comparison of CEBEs computed with ΔMP2 with experiment. In order to compare with experiment the computed values need to be corrected to account for relativistic effects. A range of values for these corrections have been suggested in the literature. In this work corrections of +0.05, +0.12, +0.24, +0.42, +0.71, +5.28 and +7.12 eV for boron, carbon, nitrogen, oxygen, fluorine, phosphorus and sulphur, respectively. The values for boron to fluorine are taken from the most recent work[51] and we note that these are larger than older values,[27] and the energy correction for phosphorus is an average of two values reported in other studies,[51,52] while the value for

sulphur is an average from three previous studies.[51–53]

Table 1: Comparison of CEBEs (in eV) computed with ΔMP2 with experiment. Experimental data taken from references[29,32] and references therein. A hybrid cc-pCVQZ:cc-pVQZ (cc-pCV5Z:cc-pVQZ for H**F**) basis set is used, and the site of the core-hole is indicated by bold font. MAD: Mean absolute deviation.

Molecule	Exp.	MP2
H_3**B**CO	195.15	194.91 (-0.24)
BF_3	202.83	202.89 (0.06)
CH_4	290.80	290.88 (0.08)
CH_3OH	292.51	292.66 (0.15)
NH_3	405.57	405.49 (-0.08)
CH_3C**N**	405.64	405.73 (0.09)
CH_3**O**H	539.11	539.57 (0.46)
H_2**O**	539.90	540.35 (0.45)
H**F**	694.23	694.41 (0.18)
P$(CH_3)_3$	2149.90	2149.87 (0.03)
PH_3	2150.69	2151.06 (0.37)
PF_3	2156.18	2156.63 (0.45)
CH_3**S**H	2477.97	2478.14 (0.17)
H_2**S**	2478.58	2478.85 (0.27)
SO_2	2483.90	2483.83 (-0.07)
MAD (B-F)	-	0.20
MAD (P,S)	-	0.23

For the ionisation of the lighter elements (B–F) the MAD is 0.20 eV. This value can be compared with the accuracy of ΔCCSD and ΔCCSD(T) calculations. The application of ΔCCSD techniques to core-ionisation is problematic owing to the presence of a virtual core orbital leading to difficulties in converging the calculations for the core-ionised state. This can be circumvented by keeping the unoccupied core orbital frozen in coupled cluster calculations.[49] For a set of small molecules considering ionisation of carbon–fluorine MADs of 0.17 eV and 0.18 eV where obtained for CVS-ΔCCSD/cc-pVTZ and CVS-ΔCCSD(T)/cc-pVTZ calculations, respectively.[49] However, the good performance of CVS-ΔCCSD(T)/cc-pVTZ benefits from a systematic cancelation of errors arising from the basis set and corrections to the CVS approximation. Including core-correlation functions in the basis set leads worsens the agreement with experiment necessitating a correction for the CVS approximation to be introduced. Currently, the accuracy of these approaches to the calculation of CEBEs

for heavier nuclei is unknown. The accuracy of ΔMP2 is only marginally worse than the CVS-ΔCCSD(T) methods and is straightforward to apply to study ionisation of the heavier nuclei, and the results in Table 1 show a small increase in the MAD for the ionisation of heavier elements (Al–Cl). Overall, ΔMP2 provides a good level of accuracy for both the lighter elements (B–F) and heavier elements (Al–Cl) and provides a reasonable reference for the assessment of the DFT calculations.

Table 2 shows the calculated ΔMP2 K-edge CEBEs for a range of small molecules that involve ionisation of boron–fluorine nuclei. These values have not been corrected for relativistic effects. Also shown are the errors (a negative error corresponds to the DFT CEBE being too low) with respect to these values for the functionals studied along with the associated MADs. Initially we consider the B3LYP, PBE0, PW86+PW91 and SCAN functionals which have MADs of 0.37 eV, 0.84 eV, 0.22 eV and 0.35 eV, respectively. These functionals have been studied previously in the context of computing CEBEs, and the good performance of the PW86+PW91 and SCAN functionals observed previously is also evident in these results. The overall MAD for these functionals is slightly greater than reported in previous work.[27,32] This may be associated with the comparison being made with calculations rather than experiment. B3LYP is also shown to perform better than PBE0. An important factor for interpreting XPS experiments is the capability to predict the energy shift between different chemical environments accurately. This is reflected in the MAD(ΔE) values, which provide a measure of the relative error in the CEBEs. For example, for boron K-edge the contribution to MAD(ΔE) will comprise of the average error in the energy difference between the CEBEs for BH_3 and BH_2F, BH_3 and BF_3, and BH_2F and BF_3. A similar process is applied for the five different nuclei making MAD(ΔE) an average over 15 values in total. This highlights the complex nature of assessing functionals for the calculation of CEBEs since the functionals have similar MAD(ΔE) values despite very different MADs.

Table 2: The error in CEBEs computed using different exchange-correlation functionals compared with reference MP2 data for ionisation of lighter nuclei. A hybrid cc-pCVTZ:cc-pVTZ basis set is used, and the site of the core-hole is indicated by bold font. MAD: Mean absolute deviation, all values in eV.

Molecule	MP2	B3LYP	PBE0	PW86+PW91	SCAN	SRC2r2	CAM-B3LYP	B2GP-PLYP
$\mathbf{B}H_3$	197.63	0.07	-0.62	-0.03	-0.20	-1.48	-0.05	-0.02
$\mathbf{B}H_2F$	199.29	-0.06	-0.80	-0.35	-0.41	-1.69	-0.14	0.12
$\mathbf{B}F_3$	202.84	-0.08	-0.90	-0.70	-0.60	-1.89	-0.08	0.53
$\mathbf{C}H_4$	290.76	0.14	-0.51	0.22	-0.03	-1.18	0.04	0.26
$\mathbf{C}H_3OH$	292.54	-0.10	-0.74	-0.09	-0.22	-1.48	-0.17	0.10
$\mathbf{C}O_2$	297.09	0.71	0.05	0.11	0.39	-0.81	0.87	1.79
$CH_3\mathbf{N}H_4$	404.72	0.20	-0.33	0.52	0.19	-0.71	0.16	0.29
$\mathbf{N}H_3$	405.25	0.12	-0.46	0.40	0.08	-0.77	0.08	0.25
$CH_3C\mathbf{N}$	405.49	-0.21	-0.77	-0.03	-0.32	-1.14	-0.18	-0.05
$CH_3\mathbf{O}H$	539.15	-0.63	-1.17	-0.14	-0.49	-0.98	-0.60	-0.62
$H_2C\mathbf{O}$	539.70	-0.82	-1.39	-0.27	-0.72	-1.20	-0.79	-0.99
$H_2\mathbf{O}$	539.93	-0.62	-1.22	-0.18	-0.53	-0.96	-0.57	-0.51
$CH_3\mathbf{F}$	692.64	-0.76	-1.38	-0.13	-0.51	-0.57	-0.63	-0.84
$BH_2\mathbf{F}$	693.65	-0.62	-1.28	0.00	-0.39	-0.44	-0.47	-0.71
H**F**	693.70	-0.35	-1.03	0.19	-0.20	-0.18	-0.21	-0.27
MAD	-	0.37	0.84	0.22	0.35	1.03	0.34	0.49
MAD(ΔE)	-	0.26	0.28	0.26	0.27	0.29	0.28	0.48

The results also provide an opportunity to assess the impact of more recent developments in exchange-correlation functionals on the calculation of CEBEs. The introduction of the long-range correction in CAM-B3LYP leads to an improvement in MAD compared with B3LYP, and overall CAM-B3LYP is one of the best performing functionals. The relatively high MADs for the B2GP-PLYP double hybrid functional is surprising, and reflects the relative importance of the DFT treatment of the exchange energy compared with the correlation energy. The performance of the short-range corrected functional SRC2r2 is poor compared with the other functionals. This functional has been parameterised to predict X-ray absorption energies for the K-edge of elements from the next row of the periodic table so accurate CEBEs would not necessarily follow. We note that the accuracy of this functional is better than the more appropriately parameterised SRC2r1 functional.[43] However, in the context of the study of heavier nuclei, an interesting observation for this functional is that the error in the CEBEs tends to reduce as the nuclear charge increases, while for the most of the other functionals the error increases. Overall, for the CEBEs of these elements, the results suggest

that PW86+PW91, SCAN and CAM-B3LYP are reliable and accurate functionals.

Table 3: The error in CEBEs computed using different exchange-correlation functionals compared with reference MP2 values for the ionisation of heavier nuclei. A hybrid cc-pCVTZ:cc-pVTZ basis set is used for the DFT calculations, and the site of the core-hole is indicated by bold font. MAD: Mean absolute deviation, all values in eV.

Molecule	MP2	B3LYP	PBE0	PW86+PW91	SCAN	SRC2r2	CAM-B3LYP	B2GP-PLYP
$\mathbf{Al}H_3$	1565.07	-1.52	-2.36	-1.18	-0.34	-0.42	-1.16	-0.24
$\mathbf{Al}H_2Cl$	1565.83	-1.60	-2.42	-1.36	-0.42	-0.52	-1.17	-0.16
$\mathbf{Al}H_2F$	1566.01	-1.51	-2.33	-1.25	-0.33	-0.48	-1.12	-0.11
$\mathbf{Si}H_4$	1843.19	-1.84	-2.72	-1.59	-0.46	-0.42	-1.44	-0.13
$H_3\mathbf{Si}OH$	1843.96	-1.93	-2.81	-1.73	-0.57	-0.55	-1.49	-0.15
$H_3\mathbf{Si}Cl$	1844.30	-1.88	-2.72	-1.70	-0.48	-0.49	-1.42	-0.06
$\mathbf{P}H_3$	2145.78	-2.50	-3.28	-2.23	-0.78	-0.68	-2.06	-0.48
$H_3\mathbf{P}O$	2148.33	-2.66	-3.50	-2.66	-0.99	-0.91	-2.13	-0.40
$H_2\mathbf{P}OOH$	2149.06	-2.72	-3.57	-2.75	-1.11	-1.01	-2.14	-0.34
$CH_3\mathbf{S}H$	2471.02	-3.15	-3.81	-2.82	-1.08	-0.87	-2.65	-0.94
$H_2C\mathbf{S}$	2471.15	-3.20	-3.86	-2.84	-1.36	-0.95	-2.72	-1.13
$H_2\mathbf{S}$	2471.73	-3.10	-3.81	-2.80	-1.03	-0.79	-2.61	-0.84
$CH_3\mathbf{Cl}$	2820.31	-3.71	-4.33	-3.37	-1.33	-0.87	-3.15	-1.30
$HCO\mathbf{Cl}$	2820.60	-3.53	-4.17	-3.29	-1.17	-0.73	-2.96	-0.96
$H\mathbf{Cl}$	2821.44	-3.61	-4.29	-3.33	-1.25	-0.76	-3.06	-1.11
MAD	-	2.57	3.33	2.33	0.85	0.70	2.09	0.56
MAD(ΔE)	-	0.09	0.09	0.13	0.13	0.12	0.07	0.13

The primary purpose of this study is to explore the performance of the functionals for the calculation of the CEBEs of heavier elements. Table 3 shows an extension of this comparison to consider ionisation of nuclei from the next row of the periodic table. From the results, a quite different picture of the relative accuracy of the functionals emerges. Initially, we consider the three most traditional functionals, B3LYP, PBE0 and PW86+PW91. For all of these three functionals there is a significant increase in the MAD, and this change is substantially greater than the error associated with the ΔMP2 calculations. This manifests itself as an underestimation of the CEBE, which increases with the nuclear charge of the ionised element. The most likely origin of this error is the self-interaction error associated with the core electrons which are in an increasingly compact orbital as the nuclear charge increases. The importance of the self-interaction error in the calculation of X-ray spectroscopy has been highlighted previously.[54–57] In the calculation of the CEBE this error will not can-

cel owing to their being a different number of core electrons in the ground and core-ionised states. However, even though there is an increase in MAD, there is a significant decrease in MAD(ΔE) compared with the lighter elements. This suggests that the error for these heavier elements is more systematic in nature.

The two functionals that performed best for the lighter elements, namely CAM-B3LYP and SCAN, also perform well for the heavier elements. CAM-B3LYP shows an improvement compared with B3LYP, but the MAD of 2.09 eV is higher than for the value of 0.85 eV for SCAN, while CAM-B3LYP has the lowest MAD(ΔE) value of 0.07 eV, compared with 0.13 eV for SCAN. The notable change in relation to the lighter elements is the two worst performing functionals SRC2r2 and B2GP-PLYP have the lowest MAD for the heavier elements. There are a lot of formal similarities between B2GP-PLYP, and MP2 so good performance of this functional may be expected. More surprising is the performance of SRC2r2. As observed earlier, this functional tended to perform better for heavier nuclei and this is reflected here. This results in the MAD decreasing for this functional while it increases for other functionals. The form of the SRC functionals will reduce the self-interaction error associated with the core-electrons, and this is the most likely reason for the observed behaviour of this functional. However, there remains a systematic underestimation of the CEBEs with this functional.

Even for the best performing functionals, the MAD remains relative high compared with the corresponding values for the lighter nuclei. Ideally it would be possible to achieve a comparable or better performance for the heavier nuclei. Two strategies are explored to improve the accuracy of the DFT calculations. The first builds upon the observation that the error appears to be systematic in nature and it is possible to apply an element and functional specific energy shift. This approach has been used in calculations of X-ray absorption and X-ray emission spectroscopies.[58,59] The second approach is to modify a functional so that

it accurately predicts the CEBEs. Both of these approaches have their drawbacks. The first requires the appropriate energy shifts to be determined for each functional for a given element, which can be problematic in the absence of experimental data or very reliable calculations. The second approach increases the empirical nature of the calculations and leads to property specific exchange-correlation functionals. Here energy shifts are derived for each element for a given functional from the average the error observed for the three molecules for each element studied in Table 3, for example the shift for chlorine is derived from the average error for $CH_3\mathbf{Cl}$, HCO**Cl** and H**Cl**. It would be possible to determine more representative shifts through a consideration of a much wider set of molcules. The parameterisation of the SRC2r2 functional is also revisited in order to optimise its accuracy. In this functional the exchange-correlation term is partitioned as[43]

$$\begin{aligned} E_{\rm xc}^{\rm SRC2} &= C_{\rm SHF} E_{\rm x}^{\rm SR-HF}(\mu_{\rm SR}) + (1 - C_{\rm SHF}) E_{\rm x}^{\rm SR-DFT}(\mu_{\rm SR}) \\ &+ C_{\rm LHF} E_{\rm x}^{\rm LR-HF}(\mu_{\rm LR}) + (1 - C_{\rm LHF}) E_{\rm x}^{\rm LR-DFT}(\mu_{\rm LR}) + E_{\rm c}^{\rm DFT} \end{aligned} \quad (1)$$

which has four parameters, $C_{\rm SHF}$, $C_{\rm LHF}$, $\mu_{\rm SR}$ and $\mu_{\rm LR}$, which determine the amount of HF exchange in the short and long range. In the parameterisation here, the parameters describing the long-range HF exchange are fixed at $C_{\rm LHF}$=0.20 and $\mu_{\rm LR}$=1.80 a_0^{-1} which means that in the long-range the functional is similar to B3LYP. The two remaining parameters $C_{\rm SHF}$ and $\mu_{\rm SR}$ were then varied to minimise the MAD with respect to the ΔMP2 data for the molecules in Table 3.

The results of the energy-shifted calculations are shown in Table 4. As expected the MAD is reduced significantly and all of the four functionals considered B3LYP, CAM-B3LYP, SCAN and SRC2r2 all have MADs less than 0.1 eV. B3LYP and CAM-B3LYP give the lowest MADs which reflects the consistently good predictions for the relative values of the CEBEs

Table 4: The error in CEBEs computed for energy-shifted and optimised exchange-correlation functionals compared with ΔMP2 for ionisation of heavier nuclei. A hybrid cc-pCVTZ:cc-pVTZ basis set is used for the DFT calculations, and the site of the core-hole is indicated by bold font. The energy shifts applied are (Al, Si, P, S, Cl) B3LYP: +1.54, +1.88, +2.62, +3.15, +3.62, CAM-B3LYP: +1.15, +1.45, +2.11, +2.66, +3.06, SCAN: +0.36, +0.50, +0.96, +1.16, +1.25, SRC2r2: +0.47, +0.49, +0.60, +0.87, +0.79. MAD: Mean absolute deviation, all values in eV.

Molecule	MP2	B3LYP	CAM-B3LYP	SCAN	SRC2r2	SRC2xps
AlH_3	1565.07	0.02	0.00	0.02	0.05	0.13
AlH_2Cl	1565.83	-0.06	0.01	-0.06	-0.05	0.04
AlH_2F	1566.01	0.03	0.00	0.03	-0.01	-0.01
SiH_4	1843.19	0.04	0.02	0.04	0.07	0.16
H_3**Si**OH	1843.96	-0.05	-0.04	-0.07	-0.06	0.01
H_3**Si**Cl	1844.30	0.00	0.01	0.02	0.00	0.06
PH_3	2145.78	0.12	0.08	0.18	-0.08	-0.05
H_3**P**O	2148.33	-0.04	-0.05	-0.03	-0.31	-0.34
H_2**P**OOH	2149.06	-0.10	-0.04	-0.15	-0.41	-0.47
CH_3**S**H	2471.02	0.00	0.01	0.08	0.00	-0.17
H_2C**S**	2471.15	-0.05	-0.09	-0.20	-0.08	-0.23
H_2**S**	2471.73	0.05	0.09	0.13	0.08	-0.09
CH_3**Cl**	2820.31	-0.09	-0.10	-0.08	-0.08	-0.08
HCO**Cl**	2820.60	0.09	0.00	0.08	0.06	0.02
H**Cl**	2821.44	0.01	0.11	0.00	0.03	0.03
MAD	-	0.05	0.04	0.08	0.09	0.12

from these functional, as shown by their MAD(ΔE) values (Table 3). Also shown are the results for the SRC2 functional with C_{SHF}=0.99 and μ_{SR}=2.21 a_0^{-1} (with C_{LHF}=0.20 and μ_{LR}=1.80 a_0^{-1}) which emerged as optimal in the fitting, this functional is denoted SRC2xps. During the fitting procedure there were many combinations of C_{SHF} and μ_{SR} that led to a similar MAD, however, the values chosen showed a good level of robustness with respect to small changes in the parameter values. The accuracy of this functional is competitive with the energy-shifted values (MAD=0.12 eV), and most of the observed error arises from H_3**P**O and H_2**P**OOH. The MAD(ΔE) value for this functional is 0.11 eV, which is also similar to the other functionals, as shown in Table 3. The effect of the increased fraction of HF exchange is shown by comparing the ground state orbital energies with those for B3LYP, which has the same fraction of HF exchange in the mid- and long-range. There is a large effect on the energy of the core-orbital, which is lowered by over 2 eV, while the energy change in the highest occupied orbital is much smaller (< 0.01 eV). The premise for the re-parameterised

short-range corrected functional is that the error in the calculated CEBEs arising from the self-interaction error associated with the core-electrons is reduced or removed, which means that the error in the calculated CEBEs should not be strongly dependent on nuclear charge. However, while there is an improvement in the accuracy of the CEBEs for the lighter elements, the overall performance is not competitive with the best performing functionals in Table 2.

Table 5: The error in calculated CEBEs compared with experiment for a set of phosphorus and sulphur containing molecules. Experimental data from reference.[29] The following energy shifts were applied for phosphorus and sulphur, B3LYP: +2.22 eV and +3.01 eV, CAM-B3LYP: +1.59 eV and +2.49 eV, SCAN: +0.62 eV and +1.00 eV, SRC2r2: +0.55 eV and +0.79 eV. A hybrid cc-pCVTZ:cc-pVTZ basis set is used, and the site of the core-hole is indicated by bold font. MAD: Mean absolute deviation, all values in eV.

Molecule	Exp.	B3LYP	CAM-B3LYP	SCAN	SRC2r2	SRC2xps
PMe_3	2149.9	-0.28	-0.41	-0.12	-0.17	0.10
PH_3	2150.69	0.09	-0.10	0.21	0.24	0.32
Me**P**Cl_2	2152.77	-0.16	-0.19	-0.11	-0.08	0.07
PCl_3	2153.98	0.02	0.05	0.09	0.10	0.07
PSCl_3	2154.94	-0.11	0.02	0.02	0.00	0.07
POCl_3	2155.39	0.02	0.11	0.19	0.06	0.02
PF_3	2156.18	0.11	0.10	0.02	0.00	0.08
PSF_3	2156.89	0.05	0.12	-0.15	0.02	0.09
POF_3	2157.67	0.14	0.17	-0.03	-0.02	0.14
PF_5	2159.26	0.09	0.12	-0.12	-0.13	0.27
MeP**S**Cl_2	2476.8	-0.01	-0.04	0.15	0.03	0.35
Me_2**S**	2477.42	0.02	0.01	0.10	0.08	0.01
P**S**Cl_3	2477.42	-0.09	-0.12	0.08	-0.08	0.25
(Me**S**)$_2$	2477.72	0.02	0.02	0.15	0.07	0.38
Me**S**H	2477.97	0.03	0.01	0.10	0.10	0.01
P**S**F_3	2478.26	-0.05	-0.06	0.03	-0.10	0.22
H_2**S**	2478.58	0.19	0.15	0.24	0.27	0.18
Me_2**S**O	2480.89	-0.04	-0.34	-0.34	-0.15	0.08
SOCl_2	2483.69	0.11	0.35	-0.19	0.05	0.21
SO_2	2483.9	-0.20	0.09	-0.35	-0.24	0.03
MAD	-	0.09	0.13	0.14	0.10	0.15

A more stringent test of the functionals is to compare the calculated CEBEs with values from experiment. Table 5 shows the accuracy of the energy-shifted CEBEs and the values from the newly parameterised SRC2 functional for a set of experimental values at the phosphorus and sulphur K-edges. In this comparison a relativistic correction has been applied

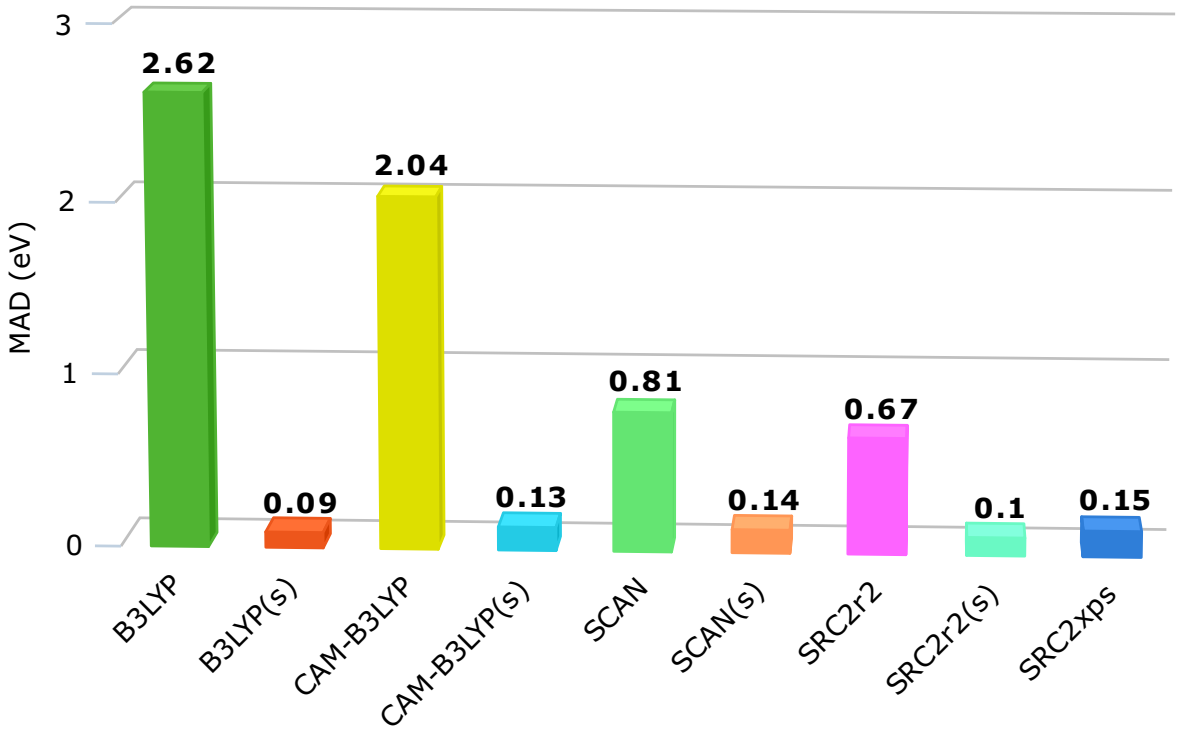


Figure 1: MADs for the uncorrected and energy-corrected CEBEs for the molecules in Table 5. The energy corrected values are denoted by (s).

as described above. The energy shifts applied are derived to minimise the error with experiment for the molecules considered in Table 5. There is some variation in the value for these shifts compared with those applied in Table 4, suggesting that a much larger data set may be required to determine reliable energy shifts that can be applied widely. All of the approaches give a similar level of accuracy, with the smallest error found for the energy-corrected B3LYP. The key result is the SRC2xps performs at a comparable level to the other functionals that have had a functional specific energy correction applied. This is illustrated in Figure 1 which gives the MADs for the uncorrected and energy-corrected values for the different functionals, along with those for SRC2xps. This shows that the magnitude of the energy corrections required decreases in the following order B3LYP>CAM-B3LYP>SCAN>SRC2r2, and no energy correction is required for SRC2xps.

Conclusions

Advances in light sources has made the X-ray spectroscopy of heavier elements more accessible. CEBEs are important in the analysis of X-ray photoelectron spectroscopy and this study has addressed the problem of computing accurate CEBEs for the elements Al–Cl using DFT. Comparison with experimental data for phosphorus and sulphur containing molecules shows that ΔMP2 calculations in conjunction with large basis sets can reproduce experimental CEBEs accurately, and this method is used as a basis to assess the DFT calculations. Through an assessment of a range of exchange-correlation functionals spanning different classes of functional it is shown that functionals that perform well for the calculation of CEBEs for lighter elements (B–F) do not necessarily perform well for heavier elements, and for commonly used functionals there is an increase in the size of the error. For a test set of molecules involving ionisation from Al–Cl, the best performing functionals with the lowest MAD with respect to ΔMP2 calculations are SCAN (meta-GGA), SRC2r2 (short-range corrected) and B2GP-PLYP (double hybrid). The functionals B3LYP and CAM-B3LYP have higher MAD but have low MAD(ΔE) indicating that they predict the relative values of the CEBEs correctly and that the error is systematic. It is demonstrated that the accuracy of the calculations can be improved through the application of element and functional dependent energy corrections, and the B3LYP and CAM-B3LYP functionals perform particularly well in this regard. A re-parameterised short-range corrected functional shows a comparable level of accuracy and is able to reproduce experimental phosphorus and sulphur K-edge CEBEs with an average error of 0.15 eV. This demonstrates the importance of reducing the self-interaction error associated with the core electrons in calculations of CEBE, and represents progress towards a DFT calculations that performs equally well for the ionisation at the K-edge of all elements.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

No acknowledgements/funding to declare.

References

(1) Lancaster, K. M.; Roemelt, M.; Ettenhuber, P.; Hu, Y.; Ribbe, M. W.; Neese, F.; Bergmann, U.; DeBeer, S. X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* **2011**, *334*, 974–977.

(2) Milne, C.; Penfold, T.; Chergui, M. Recent experimental and theoretical developments in time-resolved X-ray spectroscopies. *Coord. Chem. Rev.* **2014**, *277-278*, 44 – 68.

(3) Chergui, M.; Collet, E. Photoinduced structural dynamics of molecular systems mapped by time-resolved X-ray methods. *Chem. Rev.* **2017**, *117*, 11025–11065.

(4) Buzzi, M.; Först, M.; Mankowsky, R.; Cavalleri, A. Probing dynamics in quantum materials with femtosecond X-rays. *Nat.Rev. Mater.* **2018**, *3*, 299–311.

(5) Trinh, Q. T.; Tan, K. F.; Borgna, A.; Saeys, M. Evaluating the structure of catalysts using core-level binding energies calculated from first principles. *J. Phys. Chem. C* **2013**, *117*, 1684–1691.

(6) Lüder, J.; de Simone, M.; Totani, R.; Coreno, M.; Grazioli, C.; Sanyal, B.; Eriksson, O.; Brena, B.; Puglia, C. The electronic characterization of biphenylene, experimental and theoretical insights from core and valence level spectroscopy. *J. Chem. Phys.* **2015**, *142*, 074305.

(7) Santos, A. R.; Hanson-Heine, M. W. D.; Besley, N. A.; Licence, P. The impact of sulfur functionalisation on nitrogen-based ionic liquid cations. *Chem. Commun.* **2018**, *54*, 11403–11406.

(8) Pi, J. M.; Stella, M.; Fernando, N. K.; Lam, A. Y.; Regoutz, A.; Ratcliff, L. E. Predicting core level photoelectron spectra of amino acids using density functional theory. *J. Phys. Chem. Lett.* **2020**, *11*, 2256–2262.

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

(10) Koopmans, T. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Phys. Ther.* **1934**, *1*, 104–113.

(11) Pueyo Bellafont, N.; Bagus, P. S.; Illas, F. Prediction of core level binding energies in density functional theory: Rigorous definition of initial and final state contributions and implications on the physical meaning of Kohn-Sham energies. *J. Chem. Phys.* **2015**, *142*, 214102.

(12) Hanson-Heine, M. W. D.; George, M. W.; Besley, N. A. A scaled CIS(D) based method for the calculation of valence and core electron ionization energies. *J. Chem. Phys.* **2019**, *151*, 034104.

(13) Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* **1965**, *139*, A796–A823.

(14) Aryasetiawan, F.; Gunnarsson, O. The GW method. *Rep. Prog. Phys.* **1998**, *61*, 237–312.

(15) Golze, D.; Wilhelm, J.; van Setten, M. J.; Rinke, P. Core-level binding energies from GW: An efficient full-frequency approach within a localized basis. *J. Chem. Theory Comput.* **2018**, *14*, 4856–4869.

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

(17) Bagus, P. S. Self-consistent-field wave functions for hole states of some Ne-like and Ar-like ions. *Phys. Rev.* **1965**, *139*, A619–A634.

(18) Besley, N. A.; Gilbert, A. T. B.; Gill, P. M. W. Self-consistent-field calculations of core excited states. *J. Chem. Phys.* **2009**, *130*, 124308.

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

(20) Oosterbaan, K. J.; White, A. F.; Head-Gordon, M. Non-orthogonal configuration interaction with single substitutions for core-excited states: An extension to doublet radicals. *J. Chem. Theory Comput.* **2019**, *15*, 2966–2973.

(21) Tolbatov, I.; Chipman, D. M. Comparative study of Gaussian basis sets for calculation of core electron binding energies in first-row hydrides and glycine. *Theor. Chem. Acc.* **2014**, *133*, 1560.

(22) Fouda, A. E. A.; Besley, N. A. Assessment of basis sets for density functional theory-based calculations of core-electron spectroscopies. *Theor. Chem. Acc.* **2018**, *137*, 6.

(23) Sarangi, R.; Vidal, M. L.; Coriani, S.; Krylov, A. I. On the basis set selection for calculations of core-level states: different strategies to balance cost and accuracy. *Mol. Phys.* **2020**, *118*, e1769872.

(24) Hanson-Heine, M. W.; George, M. W.; Besley, N. A. Basis sets for the calculation of core-electron binding energies. *Chem. Phys. Lett.* **2018**, *699*, 279 – 285.

(25) Ambroise, M. A.; Jensen, F. Probing basis set requirements for calculating core ionization and core excitation spectroscopy by the Δself-consistent-field approach. *J. Chem. Theory Comput.* **2019**, *15*, 325–337.

(26) Cavigliasso, G.; Chong, D. P. Accurate density-functional calculation of core-electron binding energies by a total-energy difference approach. *J. Chem. Phys.* **1999**, *111*, 9485–9492.

(27) Takahata, Y.; Chong, D. P. DFT calculation of core-electron binding energies. *J. Electron Spectrosc. Relat. Phenom.* **2003**, *133*, 69 – 76.

(28) Takahashi, O.; Pettersson, L. G. M. Functional dependence of core-excitation energies. *J. Chem. Phys.* **2004**, *121*, 10339–10345.

(29) Segala, M.; Chong, D. P. K-shell core-electron binding energies for phosphorus- and sulfur-containing molecules calculated by density functional theory. *J. Electron Spectrosc. Relat. Phenom.* **2010**, *182*, 141 – 144.

(30) Pueyo Bellafont, N.; Viñes, F.; Illas, F. Performance of the TPSS functional on predicting core level binding energies of main group elements containing molecules: A good choice for molecules adsorbed on metal surfaces. *J. Chem. Theory Comput.* **2016**, *12*, 324–331.

(31) Tolbatov, I.; Chipman, D. M. Benchmarking density functionals and Gaussian basis sets for calculation of core-electron binding energies in amino acids. *Theor. Chem. Acc.* **2017**, *136*, 82.

(32) Kahk, J. M.; Lischner, J. Accurate absolute core-electron binding energies of molecules, solids, and surfaces from first-principles calculations. *Phys. Rev. Materials* **2019**, *3*, 100801.

(33) Besley, N. A. Modeling of the spectroscopy of core electrons with density functional theory. *WIREs Comput. Mol. Sci.* **2021**, *n/a*, e1527.

(34) Chong, D. P. Computational study of the electron spectra of vapor-phase indole and four azaindoles. *Molecules* **2021**, *26*.

(35) Becke, A. D. Simulation of delocalized exchange by local density functionals. *J. Chem. Phys.* **2000**, *112*, 4020–4026.

(36) Proynov, E.; Chermette, H.; Salahub, D. R. New τ-dependent correlation functional combined with a modified Becke exchange. *J. Chem. Phys.* **2000**, *113*, 10013–10027.

(37) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-consistent field calculations of excited states using the maximum overlap method (MOM). *J. Phys. Chem. A* **2008**, *112*, 13164–13171.

(38) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(39) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

(40) 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.

(41) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015**, *115*, 036402.

(42) Karton, A.; Tarnopolsky, A.; Lamére, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly accurate first-principles benchmark data sets for the parametrization and validation of density functional and other approximate methods. Derivation of a robust, generally

applicable, double-hybrid functional for thermochemistry and thermochemical kinetics. *J. Phys. Chem. A* **2008**, *112*, 12868–12886.

(43) Besley, N. A.; Peach, M. J. G.; 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.* **2009**, *11*, 10350–10358.

(44) Robinson, D.; Besley, N. A. Modelling the spectroscopy and dynamics of plastocyanin. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9667–9676.

(45) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.

(46) Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The Hitchhiker's Guide to Fock space. *Ann. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(47) Coriani, S.; Koch, H. Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework. *J. Chem. Phys.* **2015**, *143*, 181103.

(48) Vidal, M. L.; Feng, X.; Epifanovsky, E.; Krylov, A. I.; Coriani, S. New and efficient equation-of-motion coupled-cluster framework for core-excited and core-ionized states. *J. Chem. Theory Comput.* **2019**, *15*, 3117–3133.

(49) Zheng, X.; Cheng, L. Performance of delta-coupled-cluster methods for calculations of core-ionization energies of first-row elements. *J. Chem. Theory Comput.* **2019**, *15*, 4945–4955.

(50) 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.* **2019**, *15*, 1642–1651.

(51) Keller, L.; Blum, V.; Rinke, P.; Golze, D. Relativistic correction scheme for core-level binding energies from GW. *J. Chem. Phys.* **2020**, *153*, 114110.

(52) Ehara, M.; Kuramoto, K.; Nakatsuji, H. Relativistic effects in K-shell ionizations: SAC-CI general-R study based on the DK2 Hamiltonian. *Chem. Phys.* **2009**, *356*, 195 – 198.

(53) Mijovilovich, A.; Pettersson, L. G. M.; Mangold, S.; Janousch, M.; Susini, J.; Salome, M.; de Groot, F. M. F.; Weckhuysen, B. M. The interpretation of sulfur K-edge XANES spectra: A case study on thiophenic and aliphatic sulfur compounds. *J. Phys. Chem. A* **2009**, *113*, 2750–2756.

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

(55) Tu, G.; Rinkevicius, Z.; Vahtras, O.; Ågren, H.; Ekström, U.; Norman, P.; Carravetta, V. Self-interaction-corrected time-dependent density-functional-theory calculations of x-ray-absorption spectra. *Phys. Rev. A* **2007**, *76*, 022506.

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

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

(58) Lestrange, P. J.; Nguyen, P. D.; Li, X. Calibration of energy-specific TDDFT for modeling K-edge XAS spectra of light elements. *J. Chem. Theory Comput.* **2015**, *11*, 2994–2999.

(59) Fouda, A. A. E.; Besley, N. A. Improving the predictive quality of time-dependent

density functional theory calculations of the X-ray emission spectroscopy of organic molecules. *J. Comp. Chem.* **2020**, *41*, 1081–1090.