

# Basis Sets for the Calculation of Core-Electron Binding Energies

Magnus W. D. Hanson-Heine,<sup>a</sup> Michael W. George<sup>a,b</sup> and Nicholas A. Besley<sup>a\*</sup>

<sup>a</sup>*School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK,* <sup>b</sup>*Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, 199 Taikang East Rd., Ningbo 315100, Zhejiang, China.*

---

## Abstract

Core-electron binding energies (CEBEs) computed within a  $\Delta$ self-consistent field approach require large basis sets to achieve convergence with respect to the basis set limit. It is shown that supplementing a basis set with basis functions from the corresponding basis set for the element with the next highest nuclear charge ( $Z+1$ ) provides basis sets that give CEBEs close to the basis set limit. This simple procedure provides relatively small basis sets that are well suited for calculations where the description of a core-ionised state is important, such as time-dependent density functional theory calculations of X-ray emission spectroscopy.

*Keywords:* Core electron binding energies, basis set, x-ray photoelectron spectroscopy, density functional theory

---

## 1. Introduction

Calculations of core-electron spectroscopy have become increasingly important because of advances in X-ray light sources, such as free electron lasers. These sources can deliver short femtosecond pulses of X-rays which can probe ultrafast chemical processes [1–3]. One widely used technique is X-ray photoelectron spectroscopy (XPS) which involves the ionisation of core electrons. XPS is a powerful analytical tool that can provide information on elemental

---

<sup>1</sup>nick.besley@nottingham.ac.uk (Nicholas Besley)

composition, as well as the chemical environment and electronic state of the constituent atoms of a material. Computational simulations can often assist  
10 in deciphering the information contained within an XPS spectrum. A critical component in calculations of XPS are the core-electron binding energies (CEBEs). CEBEs correspond to the energy required to ionise a core electron, and are usually calculated as the difference between the energy of the ground state of a molecule or material and the energy with a core electron removed.  
15 The most common approach to calculating CEBEs is using a  $\Delta$ self-consistent field ( $\Delta$ SCF) approach [4–8].

For molecular systems, calculations of CEBEs most commonly use gaussian basis functions. For these calculations, the calculated CEBEs are highly dependent on the basis set used, and several groups have assessed the accuracy of  
20 different basis sets for the calculation of CEBEs [7, 9–15]. One conclusion from this work is that the inclusion of core-valence correlation functions is important for the correlation consistent basis sets of Dunning and co-workers and that the cc-pCVTZ basis set performs well. A recent study considered a wide range of  
25 basis sets and CEBEs of molecules containing first and second row elements [15]. It was found that for first row elements a number of relatively small basis sets gave CEBEs that were in close agreement with CEBEs calculated with much larger basis sets. However, for the core ionisation of second row elements modest sized basis sets performed poorly. The best performing basis sets in terms of  
30 accuracy and size were the individual gauge for localised orbitals (IGLO-II and IGLO-III) basis sets. These basis sets were developed for NMR spectroscopy and their good performance in calculations of core-electron spectroscopies has been identified previously [16, 17]. Other basis sets designed for NMR spectroscopy calculations, including the pcSseg-n basis sets of Jensen [18], have also  
35 been shown to perform well in calculations of CEBEs [15]. A limitation of these basis sets is that they are only available for the elements hydrogen, boron to fluorine and aluminium to chlorine. For the study of elements outside of these, for example s-block elements or transition metals, alternative basis sets need to

be used. With standard, widely used families of basis sets it is necessary to use  
40 very large basis sets to achieve convergence with respect to the basis set limit  
in calculations of CEBEs. It is therefore desirable to have small or moderately  
sized basis sets that are available for all (or most) elements that give CEBEs  
that are near the basis set limit. It has been shown that it is only necessary  
to use large basis sets for the core-ionised atom and that accurate CEBEs can  
45 be computed using basis sets of cc-pCVTZ quality for the core-ionised atom  
and much smaller basis sets for the remaining atoms, which represents a com-  
putationally efficient scheme [19–21]. Another factor that is relevant for the  
calculation of core-electron excitation energies, where the core electron is ex-  
cited to a virtual orbital (rather than ionised), is that it is advantageous to have  
50 additional features, such as diffuse basis functions, readily available.

In this work we propose a simple strategy that can be applied to all com-  
monly used gaussian basis sets, and provides relatively small basis sets, that are  
comparable in size to 6-31G\* or cc-pVDZ, which reproduce the CEBEs com-  
55 puted with much larger and extensive basis sets with a high degree of accuracy.  
The modification proposed is predicated on the observation that to calculate  
an accurate CEBE it is necessary to be able to describe both the neutral and  
core-ionised species accurately. Large basis sets have the inherent flexibility to  
achieve this, whereas smaller basis sets lead to an unbalanced treatment where  
60 the core-ionised state is described relatively poorly. This is illustrated in Table  
1, which shows the change in the total energy between a range of moderately  
sized basis sets and the cc-pCVQZ basis set for several molecules calculated  
using a  $\Delta$ SCF approach [8] and density functional theory (DFT) with the PBE  
exchange-correlation functional [22]. The data shows that for the split-valence  
65 and correlation-consistent basis sets the error is significantly greater for the  
core-ionised state in comparison to the ground state. For the ground state, the  
change in energy is not highly dependent on the molecule. In contrast for the  
core-ionised state, the energy change is much larger for ionisation from the core  
orbitals of the second row elements. This is exemplified for HCl with the cc-

Molecule	Basis set	$\Delta E_{\text{GS}} / \text{eV}$	$\Delta E_{\text{CI}} / \text{eV}$
H <sub>2</sub> O	6-31G*	1.69	4.57
	6-311G*	1.04	1.39
	cc-pVDZ	1.38	4.34
	cc-pCVDZ	1.32	2.64
	cc-pVTZ	0.31	0.74
	cc-pCVTZ	0.26	0.27
	IGLO-II	0.53	0.50
HF	6-31G*	2.12	5.06
	6-311G*	1.08	1.47
	cc-pVDZ	1.70	4.99
	cc-pCVDZ	1.63	3.18
	cc-pVTZ	0.38	0.93
	cc-pCVTZ	0.34	0.36
	IGLO-II	0.62	0.58
H <sub>2</sub> S	6-31G*	1.64	19.81
	6-311G*	0.80	1.80
	cc-pVDZ	0.88	18.14
	cc-pCVDZ	0.78	2.63
	cc-pVTZ	0.31	11.10
	cc-pCVDZ	0.17	0.32
	IGLO-II	1.45	1.73
HCl	6-31G*	1.67	20.00
	6-311G*	0.75	9.58
	cc-pVDZ	0.85	37.10
	cc-pCVDZ	0.75	2.58
	cc-pVTZ	0.28	30.07
	cc-pCVTZ	0.16	0.31
	IGLO-II	1.68	1.93

Table 1: Computed difference in total DFT energy from the cc-pCVQZ basis set for the ground ( $\Delta E_{\text{GS}}$ ) and core-ionised ( $\Delta E_{\text{CI}}$ ) states.

70 pVTZ basis set where the energy change for the ground state is 0.28 eV, while  
the corresponding energy change for the core-ionised state is 30.07 eV. This  
will result in an error of about 30 eV in the computed CEBE compared with  
the cc-pCVQZ value, and demonstrates why large basis sets are particularly  
important for CEBEs of second row elements. The data also shows the reason  
75 for the accuracy of the IGLO basis sets for the calculation of CEBEs. For the  
IGLO-II basis set the energy variation from cc-pCVQZ is highly consistent for  
the ground and core-ionised states. The inclusion of core-valence correlation  
basis functions in the cc-pCVDZ and cc-pCVTZ basis sets [23–25] also leads to  
a significant improvement in the description of the core-ionised state. However,  
80 it is necessary to use the cc-pCVTZ basis set to achieve an accuracy comparable  
with IGLO-II. This is consistent with earlier observations that the cc-pCVTZ  
basis set is accurate for calculations of CEBEs. In this work we show that  
accurate CEBEs can be computed with small basis sets, such as 6-31G\* and  
cc-pVDZ, by supplementing the basis set with basis functions for the element  
85 with the next highest nuclear charge ( $Z+1$ ) than the element being ionised. The  
rationale for this is the core-ionised state has an effective nuclear charge of  $Z+1$   
and these basis functions are better able to describe this state. In particular,  
the exponents of the basis functions for the  $Z+1$  element will be larger and more  
suitable for the tighter orbitals of the core-ionised state.

90

## 2. Computational Details

The accuracy of the basis sets is assessed by considering the CEBEs for the  
ionisation of the 1s electron of the following molecules: BF<sub>3</sub>, CH<sub>4</sub>, H<sub>2CO, NH<sub>3</sub>,  
HCN, HCN, H<sub>2CO, CO, SiH<sub>4</sub>, PH<sub>3</sub>, PF<sub>3</sub>, H<sub>2S, SO<sub>2</sub> and HCl, where the un-  
95 derlined element is core-ionised. This set of molecules includes excitation from  
first and second row elements. The molecular structures were optimised with  
second-order Møller-Plesset perturbation theory (MP2) with the cc-pVTZ basis  
set. CEBEs were computed with DFT using the PBE exchange-correlation func-</sub></sub></sub>

tional. It has been shown elsewhere that the basis set dependence of computed  
100 CEBEs is not sensitive to the functional used [15]. In addition to DFT, CEBEs  
have also been computed with MP2. The core-ionised state is computed using  
the maximum overlap method (MOM) [26] to prevent variational collapse and  
maintain the core-ionised state during the SCF procedure. The CEBEs are as-  
sessed by computing the mean absolute deviation (MAD) from values evaluated  
105 with the cc-pCVQZ basis set, with the cc-pVQZ basis set used for hydrogen. A  
number of studies have performed detailed comparisons between DFT and MP2  
CEBEs and experiment [13, 14, 19, 27, 28]. This work shows that it is possible  
to achieve an agreement to within about 0.2 eV if large basis sets are used, and  
relativistic effects are treated for heavier nuclei. Consequently, here we focus on  
110 a comparison with values computed with a large basis set.

The standard basis sets considered include the split-valence basis sets 6-31G\*  
and 6-311G\*, the cc-pVDZ correlation-consistent basis set, and the IGLO-II and  
IGLO-III basis sets [29–37]. Several non-standard basis sets are also considered.  
115 Previous studies have shown that uncontracting the core basis function can im-  
prove calculated CEBEs [8], and we consider two such basis sets u6-31G\* and  
u6-311G\* where the 'u' denotes the core basis function has been uncontracted.  
The objective of this study is to produce basis sets with relatively few basis  
functions that reproduce CEBEs close to the basis set limit by supplementing  
120 standard basis sets with basis functions from the Z+1 element. The first vari-  
ant of these basis sets corresponds to adding just the core basis function from  
the Z+1 element to give the (core Z+1)6-31G\*, (core Z+1)6-311G\* and (core  
Z+1)cc-pVDZ basis sets. In the second variation of these basis sets, the core  
and valence (but not polarisation) basis functions from the Z+1 element are in-  
125 cluded to give (Z+1)6-31G\*, (Z+1)6-311G\* and (Z+1)cc-pVDZ basis sets. The  
reason for not including Z+1 polarisation functions is to keep the size of the ba-  
sis sets to a minimum. Examples of these basis sets, highlighting the additional  
basis functions included, are given for carbon and silicon in the Supporting In-  
formation. In the calculations shown here, the modified basis set has been used

130 for all atoms. However, as discussed later, an alternative strategy is to use the  
modified basis set only for the element that has a core electron ionised, while  
retaining the standard basis set for the remaining elements. This approach is  
more efficient since it reproduces the values of the calculations where the mod-  
ified basis set is used throughout but with considerably fewer additional basis  
135 functions. For transition metal complexes the cc-pwCVQZ basis set is used for  
the transition metal atoms with the cc-pCVQZ for the remaining atoms, this  
basis set is referred to as simply cc-pCVQZ for brevity. The basis sets were  
obtained from the Basis Set Exchange [38, 39]. For some of the metal com-  
plexes it was necessary to use orbitals generated for the core-ionised state using  
140 the PBE0 functional as a starting guess in order to converge the calculation  
for the PBE functional. The structures of the transition metal complexes were  
optimised using the PBE0 functional with the Stuttgart Relativistic Small Core  
basis set for the transition metal atoms[40] and the 6-311G\* basis set for the  
non-metal atoms. All calculations were performed with the Q-CHEM software  
145 package [41].

### 3. Results and Discussion

Table 2 shows the MADs for the CEBEs calculated with PBE and MP2 for  
the first row nuclei, the maximum errors and standard deviations are included  
150 in the Supporting Information. For the PBE calculations, the cc-pVDZ and  
6-31G\* basis sets perform poorly with a MAD of about 3 eV, while the 6-311G\*  
and IGLO basis sets are considerably more accurate. In particular, the IGLO  
basis sets have an error of less than 0.1 eV compared with the cc-pCVQZ basis  
set. These findings are consistent with previous studies of CEBEs of first row  
155 nuclei [15]. For the non-standard basis sets, uncontraction of the core orbital  
basis function for the 6-31G\* only results in a modest reduction in the error with  
a MAD of 2.28 eV, with a particularly large error for H<sub>2</sub>CO. Addition of the core  
orbital for the Z+1 element in the (core Z+1)6-31G\* and (core Z+1)cc-pVDZ

Basis set	MAD <sub>PBE</sub> / eV	MAD <sub>MP2</sub> / eV
6-31G*	2.88	2.08
(core Z+1)6-31G*	2.44	1.16
(Z+1)6-31G*	0.13	0.44
u6-31G*	2.28	1.29
6-311G*	0.40	0.25
(core Z+1)6-311G*	0.33	0.27
(Z+1)6-311G*	0.03	0.54
u6-311G*	0.12	0.46
cc-pVDZ	2.93	2.16
(core Z+1)cc-pVDZ	1.44	0.96
(Z+1)cc-pVDZ	0.13	0.33
IGLO-II	0.07	3.36
IGLO-III	0.05	0.64

Table 2: Mean absolute deviations relative to the cc-pCVQZ basis set for CEBEs of first row elements for the molecules  $\underline{\text{B}}\text{F}_3$ ,  $\underline{\text{C}}\text{H}_4$ ,  $\text{H}_2\underline{\text{C}}\text{O}$ ,  $\underline{\text{N}}\text{H}_3$ ,  $\text{HCN}$ ,  $\text{HCN}$ ,  $\text{H}_2\underline{\text{C}}\text{O}$  and  $\underline{\text{C}}\text{O}$  computed with DFT with the PBE exchange-correlation functional and MP2.

basis sets results in MADs that remain too high. We note that uncontraction  
160 of the core orbital and addition of the Z+1 core basis function does improve  
on the already low MAD for 6-31G\*. Addition of the Z+1 core and valence  
basis functions in the (Z+1)6-31G\* and (Z+1)cc-pVDZ results in a low MAD  
of about 0.1 eV which is comparable in accuracy to the IGLO basis sets. The  
need to include both core and valence orbitals will be discussed in more detail  
165 later. These trends are also observed for the MP2 calculations. In general the  
MP2 basis set error is larger than for DFT. For MP2 the IGLO basis sets have a  
significant error of 0.64 eV for IGLO-III and a particularly large error of 3.36 eV  
for IGLO-II. However, the (Z+1)6-31G\* and (Z+1)cc-pVDZ maintain a good  
level of performance with MADs of 0.44 eV and 0.33 eV, respectively.

170

For core ionisation of the second row nuclei, the good performance of the



(Z+1) basis sets becomes more evident. It is for calculations of CEBEs of heavier nuclei that the deficiencies of standard basis sets becomes increasingly apparent. Table 3 shows the MADs for the CEBEs calculated with PBE and MP2 for the second row nuclei, the maximum errors and standard deviations are included in the Supporting Information. For the DFT calculations, standard small basis sets have MADs of about 15 eV, and even the 6-311G\* basis set that performed well for the first row nuclei has a MAD of over 2 eV. The (Z+1)6-31G\* and (Z+1)cc-pVDZ basis sets show a remarkable improvement in accuracy compared with their standard versions with MADs of 0.41 eV and 0.23 eV. These MADs are comparable with those for the IGLO basis sets. The inclusion of just the Z+1 core orbital basis function or uncontraction of the core orbital basis functions do not lead to a significant improvement for the ionisation of the second row core electrons. For the MP2 calculations excellent performance of the Z+1 basis sets is also observed, and for these calculations the MADs for the IGLO basis sets are significantly greater than for the (Z+1)6-31G\* and (Z+1)cc-pVDZ basis sets.

The preceding analysis has shown that for DFT calculations of CEBEs for the first and second row p-block elements the (Z+1)6-31G\* basis set are considerably more accurate with respect to cc-pCVQZ than the standard 6-31G\* basis set, and comparable in accuracy to the IGLO basis sets. The MADs for the MP2 calculations tend to be greater than the corresponding DFT values, however, for these calculations the MADs for the (Z+1)6-31G\* basis set are lower than for the IGLO basis sets. We now consider CEBEs for first and second row s-block elements. The IGLO basis sets have not been defined for these elements and it is currently necessary to use large basis sets to calculate CEBEs for these systems reliably [15]. Table 4 shows CEBEs for the 1s electrons of molecules containing Li, Be, Na and Mg computed with PBE and MP2 with the cc-pCVQZ, 6-31G\* and (Z+1)6-31G\* basis sets. For both PBE and MP2 calculations with the 6-31G\* basis set, there is a large increase in the error with respect to cc-pCVQZ for the Na and Mg K-edges compared with Li and Be, consistent with the observations for the p-block elements. The CEBEs com-

Basis set	MAD <sub>PBE</sub> / eV	MAD <sub>MP2</sub> / eV
6-31G*	15.43	17.01
(core Z+1)6-31G*	16.67	15.58
(Z+1)6-31G*	0.23	0.18
u6-31G*	15.67	14.58
6-311G*	2.42	2.13
(core Z+1)6-311G*	2.24	1.74
(Z+1)6-311G*	0.32	0.28
u6-311G*	2.28	1.83
cc-pVDZ	14.79	16.62
(core Z+1)cc-pVDZ	13.16	12.17
(Z+1)cc-pVDZ	0.41	0.45
IGLO-II	0.41	0.84
IGLO-III	0.23	0.90

Table 3: Mean absolute deviations relative to the cc-pCVQZ basis set for CEBEs of second row elements for the molecules  $\underline{\text{Si}}\text{H}_4$ ,  $\underline{\text{P}}\text{H}_3$ ,  $\underline{\text{P}}\text{F}_3$ ,  $\text{H}_2\underline{\text{S}}$ ,  $\underline{\text{S}}\text{O}_2$  and  $\underline{\text{H}}\underline{\text{C}}\underline{\text{I}}$  computed with DFT with the PBE exchange-correlation functional and MP2.

puted with PBE/(Z+1)6-31G\* are in excellent agreement with the larger basis set, and have a MAD of just 0.3 eV compared with 10.1 eV for 6-31G\*. For the corresponding MP2 calculations a similar trend is observed, although the MAD for the (Z+1)6-31G\* basis set has a larger value of 1.5 eV, primarily due to the CEBEs for two Mg K-edges.

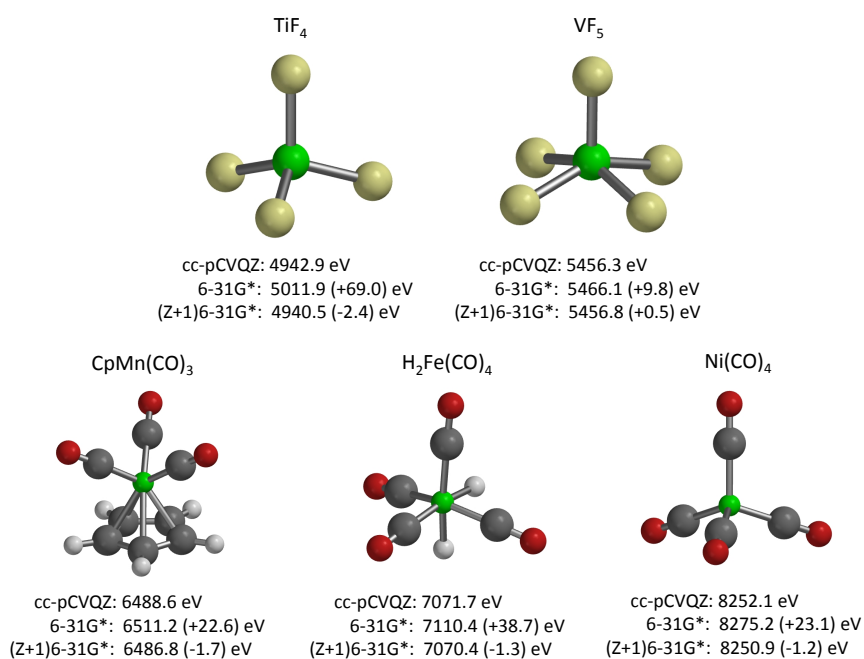


Figure 1: Calculated CEBEs for transition metal complexes.

We now consider CEBEs for the K-edge of transition metals from the first row of the d-block. For these systems only DFT calculations have been performed owing to its lower computational cost compared with MP2, and most calculations of CEBEs in the literature use DFT. The energies presented have not been corrected for relativistic effects which are large for the K-edge of transition metals [42], however, these effects will not alter the comparison between

Molecule	PBE/cc-pCVQZ	PBE/6-31G*	PBE/(Z+1)6-31G*
LiCH <sub>3</sub>	63.3	65.3 (+2.0)	63.3 (0.0)
LiCl	64.6	67.1 (+2.5)	64.6 (0.0)
BeF <sub>2</sub>	122.6	123.9 (+1.3)	122.5 (-0.1)
Be(OH) <sub>2</sub>	120.2	121.3 (+1.1)	120.1 (-0.1)
NaCH <sub>3</sub>	1083.8	1096.6 (+12.8)	1084.3 (+0.5)
NaF	1075.8	1097.5 (+21.7)	1076.1 (+0.3)
MgH <sub>2</sub>	1306.9	1324.8 (+17.9)	1306.7 (-0.2)
Mg(OH) <sub>2</sub>	1309.2	1330.5 (+21.3)	1308.0 (-1.2)
MAD	-	10.1	0.3
Molecule	MP2/cc-pCVQZ	MP2/6-31G*	MP2/(Z+1)6-31G*
LiCH <sub>3</sub>	63.8	64.7 (+0.9)	63.0 (-0.8)
LiCl	65.0	66.4 (+1.4)	64.6 (-0.4)
BeF <sub>2</sub>	123.6	123.9 (+0.3)	122.7 (-0.9)
Be(OH) <sub>2</sub>	121.4	121.1 (-0.3)	120.5 (-0.9)
NaCH <sub>3</sub>	1078.1	1097.5 (+19.4)	1077.1 (-1.0)
NaF	1079.3	1098.4 (+19.1)	1078.2 (-1.1)
MgH <sub>2</sub>	1305.8	1326.4 (+20.6)	1309.4 (+3.6)
Mg(OH) <sub>2</sub>	1307.9	1327.6 (+19.7)	1311.1 (+3.2)
MAD	-	10.2	1.5

Table 4: Computed CEBEs with the deviation from cc-pCVQZ in parenthesis for the K-edge of s-block elements. Energies are in eV.

215 different basis sets. The calculations with the 6-31G\* basis show a wide range of  
errors compared with the cc-pCVQZ calculation, varying from +9.8 eV to +69.0  
eV. Clearly calculations with this basis set would not be reliable. In contrast,  
the corresponding errors for the (Z+1)6-31G\* basis are reduced greatly and vary  
220 between -2.4 eV and +0.5 eV. Taking Ni(CO)<sub>4</sub> as an example, the calculation  
with the cc-pCVQZ basis set has a total of 841 basis functions compared with  
156 and 245 basis functions for 6-31G\* and (Z+1)6-31G\*, respectively. Con-  
sequently, the (Z+1)6-31G\* basis set reproduces the values of the cc-pCVQZ  
basis set despite being less than one third of the size. In practice the (Z+1)6-  
31G\* basis set can be used more efficiently. In the calculations presented the  
225 (Z+1)6-31G\* basis set is used for all elements in the molecule or complex. An  
alternative strategy is to use the (Z+1)6-31G\* basis set just for the element that  
has a core electron ionised while retaining the 6-31G\* basis set for the remain-  
ing elements. For Ni(CO)<sub>4</sub>, this calculation has a total of 173 basis functions  
and the computed CEBE is essentially unchanged from the calculation with the  
230 (Z+1)6-31G\* basis set for all elements. Consequently, it is possible to greatly  
improve the accuracy of the computed CEBE with the addition of relatively few  
basis functions.

Molecules that have energetically equivalent core orbitals present an interest-  
235 ing case for the calculation of CEBEs, in particular with regard to augmenting  
the basis set for the core-ionised atom only. Table 5 shows CEBEs computed  
using DFT for a range of molecules that have symmetrically equivalent atoms.  
For these systems, it can be necessary to break the symmetry of the wavefunc-  
tion (or molecule) to allow the core hole to localise on one atom and achieve  
240 convergence for the core-ionised state. The (Z+1)6-31G\* basis sets works well  
for these systems, with the calculated CEBEs within 0.2 eV of the values for  
the cc-pCVQZ basis set, improving significantly the values for the 6-31G\* basis  
set. However, if the (Z+1)6-31G\* basis set is used only for the atom being  
ionised (i.e. just a single, carbon, nitrogen or fluorine atom) a significant error  
245 is observed and the values are not an improvement on 6-31G\*.

Molecule	cc-pCVQZ	6-31G*	(Z+1)6-31G*	<sup>†</sup> (Z+1)6-31G*
C <sub>2</sub> H <sub>2</sub>	289.6	292.1	289.7	292.2
C <sub>2</sub> H <sub>4</sub>	289.1	291.5	289.2	291.6
C <sub>2</sub> H <sub>6</sub>	289.0	291.4	289.1	291.4
N <sub>2</sub>	408.8	411.6	409.0	412.0
F <sub>2</sub>	694.7	697.9	694.7	698.0

Table 5: Computed PBE CEBEs for the K-edge molecules with energetically equivalent core orbitals. Energies are in eV.<sup>†</sup>The (Z+1)6-31G\* is used for one heavy atom only, 6-31G\* used for the remaining atoms.

Now we examine the molecular orbitals associated with the different basis sets. SiH<sub>4</sub> is used as an example, but the results are representative of other systems. For the ground state there is only a small difference (less than 0.007 a.u.) in the orbital energies of the occupied orbitals between the DFT calculations with the cc-pCVQZ, 6-31G\* and (Z+1)6-31G\* basis sets. However, for the core ionised state some significant differences in the orbital energies emerge for the 6-31G\* basis set for the silicon 1s, 2s and 2p orbitals. For example, the energies of the occupied 1s orbitals for the cc-pCVQZ, (Z+1)6-31G\* and 6-31G\* basis set are -71.716, -71.719 and -73.071 a.u., respectively. Similarly, for the 2s orbitals the energies are -6.384, -6.394 and -6.784 a.u. Figure 2 shows the radial behaviour of the core orbitals centred on the silicon atom. For the cc-pCVQZ basis set the 1s and 2s orbitals have a greater amplitude at the nuclei in the core-ionised state, while the 6-31G\* basis set shows no significant difference between the two states. The orbitals for the (Z+1)6-31G\* basis set are qualitatively correct with a distinctly larger amplitude for the core-ionised state, although the orbitals are less sharply peaked. The poor description of the 2s orbital by the 6-31G\* basis set provides an explanation for why adding just the core orbital basis function for the Z+1 element was not sufficient, since the additional basis functions will not improve the description of the 2s orbital. For the first row elements, analysis of the molecular orbital coefficients shows that

Molecule	cc-pCVQZ	(Z+1)6-31G*	IGLO-II	6-31G*
HCN	401.6 (0.0356)	402.4 (0.0353)	401.8 (0.0355)	405.8 (0.0327)
	401.2 (0.0341)	401.9 (0.0349)	401.4 (0.0339)	405.3 (0.0315)
	394.2 (0.0008)	394.9 (0.0009)	394.5 (0.0008)	398.5 (0.0009)
H <sub>2</sub> S	2501.1 (0.0065)	2502.8 (0.0065)	2502.8 (0.0064)	2537.9 (0.0044)
	2498.3 (0.0045)	2500.0 (0.0046)	2500.0 (0.0046)	2535.1 (0.0032)
	2496.4 (0.0041)	2498.0 (0.0042)	2498.2 (0.0042)	2533.1 (0.0030)
TiF <sub>4</sub>	5005.4 (0.0002)	5002.4 (0.0002)	-	5076.0 (0.0000)
	5003.9 (0.0001)	5000.8 (0.0001)	-	5073.9 (0.0001)
	4986.8 (0.0003)	4983.7 (0.0003)	-	5056.7 (0.0001)

Table 6: Computed X-ray emission energies for the PBE exchange-correlation functional with oscillator strength in parenthesis. Energies are in eV.

the core orbital in the core-ionised state in the (Z+1)6-31G\* calculation has a large coefficient for the Z+1 2s basis function (for example, 0.2 in CH<sub>4</sub>). This underlies the poor performance of the (core Z+1)6-31G\* basis set for these first row elements, since this basis function will not be included. For the highest occupied molecular orbital (HOMO) the 6-31G\* basis set does show the same trend in the variation between the two states as the large basis set, but is not as accurate as the (Z+1)6-31G\* basis set.

Another area where an accurate description of the core-ionised state is important is the simulation of X-ray emission spectroscopy with time-dependent density functional theory (TDDFT). X-ray emission spectra can be simulated by applying TDDFT to a Kohn-Sham determinant with a core-hole [42–45]. This approach has been applied to study organic molecules and inorganic complexes, and it has been shown that these calculations are highly dependent on the basis set used [15, 42]. TDDFT is also commonly used to study X-ray absorption spectroscopy [46, 47]. In contrast to X-ray emission spectroscopy, standard basis sets perform well for these calculations since they are based upon

the linear response of the ground state density. Table 6 shows X-ray emission  
285 energies and the associated intensities computed with TDDFT with the PBE  
functional for the cc-pCVQZ, 6-31G\* and (Z+1)6-31G\* basis sets. TDDFT cal-  
culations with the PBE functional will not accurately reproduce experimental  
values, but here only the variation between the basis sets is of interest. The tran-  
sition energies predicted by the 6-31G\* basis set show a large variation from  
290 the cc-pCVQZ basis set, in particular for TiF<sub>4</sub> where a qualitatively incorrect  
spectrum is predicted. The (Z+1)6-31G\* basis set predicts values much closer  
to cc-pCVQZ, and in close agreement with the IGLO-II basis set for HCN and  
H<sub>2</sub>S. In calculations of X-ray emission spectroscopy it is common to shift the  
computed spectrum to align with experiment. For the HCN and H<sub>2</sub>S molecules,  
295 the 6-31G\* basis set calculations do predict the energy differences between the  
transitions, but the relative intensity of the bands is predicted less accurately.  
For the transition metal complex, TiF<sub>4</sub>, the spectrum is predicted poorly with  
the smaller basis set. This illustrates that the choice of basis set is important  
for these calculations.

#### 300 4. Conclusions

The calculation of CEBEs is important in the simulation and interpretation  
of XPS. Computed CEBEs converge slowly with respect to the basis set and  
large basis sets are necessary for accurate CEBEs. This high dependence on the  
basis set is associated with small basis sets not providing a balanced treatment of  
305 ground and core-ionised states. Including basis functions from the basis sets for  
the Z+1 element provides a simple solution to this problem that can be applied  
generally. Furthermore, it is sufficient to include the additional basis functions  
for the element where the core-ionisation occurs which means that relatively few  
additional basis functions are required. It is shown that these modified basis  
310 sets result in a large improvement in CEBEs computed with small basis sets  
compared to those of much larger basis sets. CEBEs computed with DFT for  
the K-edge of first and second row elements with the (Z+1)6-31G\* have a MAD



of 0.3 eV relative to cc-pCVQZ, while (Z+1)cc-pVDZ has a MAD of less than 0.5 eV. This represents an accuracy similar to the IGLO basis sets that are known to be accurate for calculations of CEBEs. For MP2 calculations the error for the (Z+1)6-31G\* basis set was less than that for the IGLO basis sets. A limitation of the IGLO basis sets is that they are only available for a subset of elements. It is shown that the (Z+1)6-31G\* basis set also performs well for s-block elements and first row transition metals, for which IGLO basis sets are not defined. It is also shown that these basis sets lead to a large improvement in performance for calculations of XES with TDDFT. Overall, the inclusion of basis functions from the Z+1 element, as exemplified by (Z+1)6-31G\*, provides relatively small basis sets that are well suited for calculations where the description of a core-ionised state is important.

### Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council [Grant No. EP/N002148/1].

### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at

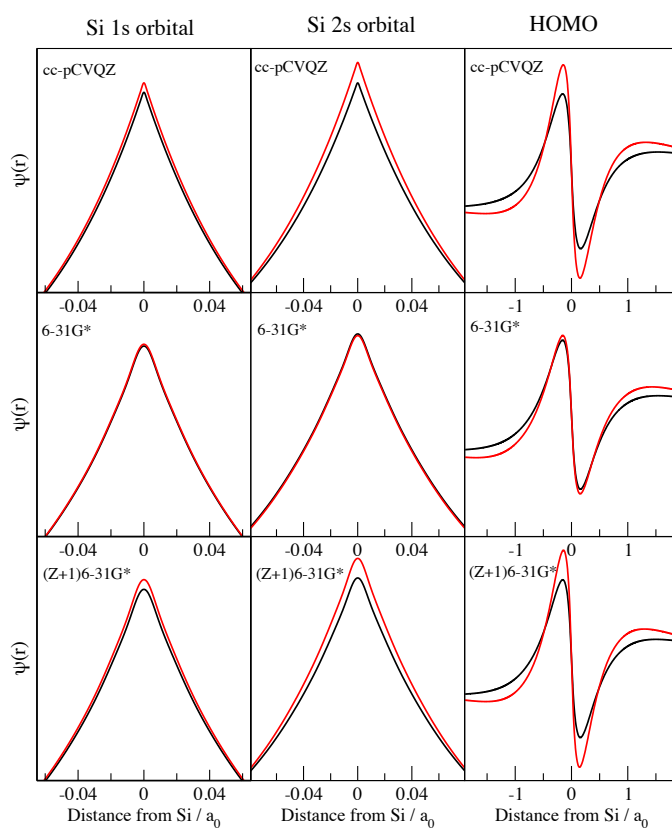


Figure 2: Plot of the radial behaviour of the silicon 1s orbital in  $\text{SiH}_4$ , (a) cc-pCVQZ, (b) 6-31G\* and (c) (Z+1)6-31G\*. The ground state is shown in black and the core-ionised state in red.

## References

- [1] C. Bressler, M. Chergui, Ultrafast x-ray absorption spectroscopy, *Chem. Rev.* 104 (2004) 1781–1812.
- [2] L. X. Chen, X. Zhang, M. L. Shelby, Recent advances on ultrafast x-ray spectroscopy in the chemical sciences, *Chem. Sci.* 5 (2014) 4136–4152.
- [3] M. Chergui, E. Collet, Photoinduced structural dynamics of molecular systems mapped by time-resolved x-ray methods, *Chem. Rev.* 117 (2017) 11025–11065.
- [4] P. S. Bagus, Self-consistent-field wave functions for hole states of some Ne-like and Ar-like ions, *Phys. Rev.* 139 (1965) A619–A634.
- [5] P. Deutsch, L. Curtiss, Ab initio calculation of the K-shell excitation and ionization energies of CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>O, and HF, *Chem. Phys. Lett.* 39 (1976) 588 – 592.
- [6] L. Triguero, O. Plashkevych, L. G. M. Pettersson, H. Ågren, Separate state vs. transition state Kohn-Sham calculations of X-ray photoelectron binding energies and chemical shifts, *J. Electron Spectrosc. Relat. Phenom.* 104 (1999) 195 – 207.
- [7] Y. Takahata, D. P. Chong, DFT calculation of core-electron binding energies, *J. Electron Spectrosc. Relat. Phenom.* 133 (2003) 69 – 76.
- [8] N. A. Besley, A. T. B. Gilbert, P. M. W. Gill, Self-consistent-field calculations of core excited states, *J. Chem. Phys.* 130 (2009) 124308.
- [9] G. Cavgliasso, D. P. Chong, Accurate density-functional calculation of core-electron binding energies by a total-energy difference approach, *J. Chem. Phys.* 111 (1999) 9485–9492.
- [10] D. P. Chong, C. Bureau, Accurate density-functional calculation of core-electron binding energies with a scaled polarized triple-zeta basis set: VII.

Effects of poorer geometry and poorer basis sets, *J. Electron Spectrosc. Relat. Phenom.* 106 (2000) 1 – 6.

- [11] S. Carniato, P. Millié, Accurate core electron binding energy calculations using small 6-31G and TZV core hole optimized basis sets, *J. Chem. Phys.* 116 (2002) 3521–3532.
- [12] D. P. Chong, Y. Takahata, Density functional theory calculation of electron spectra of formaldehyde, *Chem. Phys. Lett.* 418 (2006) 286 – 291.
- [13] I. Tolbatov, D. M. Chipman, Comparative study of gaussian basis sets for calculation of core electron binding energies in first-row hydrides and glycine, *Theor. Chem. Acc.* 133 (2014) 1560.
- [14] I. Tolbatov, D. M. Chipman, Benchmarking density functionals and Gaussian basis sets for calculation of core-electron binding energies in amino acids, *Theor. Chem. Acc.* 136 (2017) 82.
- [15] A. E. A. Fouda, N. A. Besley, Assessment of basis sets for density functional theory-based calculations of core-electron spectroscopies, *Theor. Chem. Acc.* 137 (2018) 6.
- [16] A. Mijovilovich, L. G. M. Pettersson, S. Mangold, M. Janousch, J. Susini, M. Salome, F. M. F. de Groot, B. M. Weckhuysen, The interpretation of sulfur k-edge xanes spectra: A case study on thiophenic and aliphatic sulfur compounds, *J. Phys. Chem. A* 113 (2009) 2750–2756.
- [17] T. Fransson, I. Zhovtobriukh, S. Coriani, K. T. Wikfeldt, P. Norman, L. G. M. Pettersson, Requirements of first-principles calculations of x-ray absorption spectra of liquid water, *Phys. Chem. Chem. Phys.* 18 (2016) 566–583.
- [18] F. Jensen, Segmented contracted basis sets optimized for nuclear magnetic shielding, *J. Chem. Theory Comput.* 11 (2015) 132–138.

- [19] J. Shim, M. Klobukowski, M. Barysz, J. Leszczynski, Calibration and applications of the  $\Delta$ MP2 method for calculating core electron binding energies, *Phys. Chem. Chem. Phys.* **13** (2011) 5703–5711.
- [20] B. Kovac, I. Ljubic, A. Kivimaki, M. Coreno, I. Novak, Characterisation of the electronic structure of some stable nitroxyl radicals using variable energy photoelectron spectroscopy, *Phys. Chem. Chem. Phys.* **16** (2014) 10734–10742.
- [21] B. Kovac, I. Ljubic, A. Kivimaki, M. Coreno, I. Novak, The study of the electronic structure of some N-heterocyclic carbenes (NHCs) by variable energy photoelectron spectroscopy, *Phys. Chem. Chem. Phys.* **17** (2015) 10656–10667.
- [22] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77** (1996) 3865–3868.
- [23] D. E. Woon, T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon, *J. Chem. Phys.* **103** (1995) 4572–4585.
- [24] K. A. Peterson, T. H. Dunning Jr., Accurate correlation consistent basis sets for molecular core-valence correlation effects: The second row atoms Al–Ar, and the first row atoms B–Ne revisited, *J. Chem. Phys.* **117** (2002) 10548–10560.
- [25] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning Jr., A. K. Wilson, Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg, *Theor. Chem. Acc.* **128** (2011) 69–82.
- [26] A. T. B. Gilbert, N. A. Besley, P. M. W. Gill, Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM)., *J. Phys. Chem. A* **112** (2008) 13164–13171.

- 410 [27] Y. Takahata, D. P. Chong, DFT calculation of core-electron binding energies, *J. Electron Spectrosc. Relat. Phenom.* 133 (2003) 69 – 76.
- [28] M. Segala, D. P. Chong, K-shell core-electron binding energies for phosphorus- and sulfur-containing molecules calculated by density functional theory, *J. Electron Spectrosc. Relat. Phenom.* 182 (2010) 141 – 144.
- 415 [29] W. J. Hehre, R. Ditchfield, J. A. Pople, Self-consistent molecular orbital methods. XII. further extensions of gaussian type basis sets for use in molecular orbital studies of organic molecules, *J. Chem. Phys.* 56 (1972) 2257–2261.
- [30] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions, *J. Chem. Phys.* 72 (1980) 650–654.
- 420 [31] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, Self-consistent molecular orbital methods. XXIII. A polarization type basis set for second row elements, *J. Chem. Phys.* 77 (1982) 3654–3665.
- 425 [32] V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, 6-31G\* basis set for atoms K through Zn, *J. Chem. Phys.* 109 (1998) 1223–1229.
- [33] T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen, *J. Chem. Phys.* 90 (1989) 1007–1023.
- 430 [34] D. E. Woon, T. H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. III. the atoms aluminum through argon, *J. Chem. Phys.* 98 (1993) 1358–1371.
- [35] N. B. Balabanov, K. A. Peterson, Systematically convergent basis sets for transition metals. I. all-electron correlation consistent basis sets for the 3d elements Sc-Zn, *J. Chem. Phys.* 123 (2005) 064107.
- 435

- [36] S. Huzinaga, Gaussian type functions for polyatomic systems. I, *J. Chem. Phys.* 42 (1965) 1293–1302.
- [37] M. Schindler, W. Kutzelnigg, Theory of magnetic susceptibilities and NMR  
440 chemical shifts in terms of localized quantities. II. Application to some  
simple molecules, *J. Chem. Phys.* 76 (1982) 1919–1933.
- [38] D. Feller, The role of databases in support of computational chemistry  
calculations, *J. Comp. Chem.* 17 (1996) 1571–1586.
- [39] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi,  
445 J. Chase, J. Li, T. L. Windus, Basis set exchange: A community database  
for computational sciences, *J. Chem. Inf. Model.* 47 (2007) 1045–1052.
- [40] A. Bergner, M. Dolg, W. Küchle, H. Stoll, H. Preuss, Ab initio energy-  
adjusted pseudopotentials for elements of groups 13-17, *Mol. Phys.* 80  
(1993) 1431–1441.
- 450 [41] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kuss-  
mann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey,  
P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Ku, A. Lan-  
dau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz,  
R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman,  
455 D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard,  
E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova,  
C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden,  
M. Diedenhofen, R. A. DiStasio Jr., H. Do, A. D. Dutoi, R. G.  
Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya,  
460 J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G.  
Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev,  
J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk,  
C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko,  
C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F.  
465 Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J.

- Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stck, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt,  
470 O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-  
475 Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, *Mol. Phys.* 113 (2015) 184–215.  
480
- [42] I. P. E. Roper, N. A. Besley, The effect of basis set and exchange-correlation functional on time-dependent density functional theory calculations within the Tamm-Dancoff approximation of the x-ray emission spectroscopy of transition metal complexes, *J. Chem. Phys.* 144 (2016) 114104.
- 485 [43] N. A. Besley, Equation of motion coupled cluster theory calculations of the x-ray emission spectroscopy of water, *Chem. Phys. Lett.* 542 (2012) 42 – 46.
- [44] J. D. Wadey, N. A. Besley, Quantum chemical calculations of x-ray emission spectroscopy, *J. Chem. Theory Comput.* 10 (2014) 4557–4564.
- 490 [45] Y. Zhang, S. Mukamel, M. Khalil, N. Govind, Simulating valence-to-core x-ray emission spectroscopy of transition metal complexes with time-dependent density functional theory, *J. Chem. Theory Comput.* 11 (2015) 5804–5809.
- [46] N. A. Besley, F. A. Asmuruf, Time-dependent density functional theory



495 calculations of the spectroscopy of core electrons, *Phys. Chem. Chem. Phys.*  
12 (2010) 12024–12039.

[47] N. A. Besley, Fast time-dependent density functional theory calculations of the x-ray absorption spectroscopy of large systems, *J. Chem. Theory Comput.* 12 (2016) 5018–5025.