Supplementary Information

Efficient Basis Sets for Core Excited States Motivated by Slater's Rules

Jin Qian[1,2,3], Ethan J. Crumlin[1,2]*, David Prendergast[3]*

[1] Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

[2] Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

[3] Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States.

This file includes:

1. Method 2
1.1. Computational details 2
1.2. Core-Electron Binding Energy Calculations 2
1.3. Gaussian Basis-Set Orbital Representations 2
1.4. Construction of the core-excited basis sets, motivated by Slater's Rule 3
2. Results 6
2.1. Ground and core-excited state orbitals for carbon 6
2.2. Detailed results for carbon 7
2.3. The special case of CO 8
2.4. Comparison of results obtained through CCSD vs. B3LYP 9
2.5. Ground and core-excited state orbitals for B, N, and O 9
2.6. Energies and charges for B, N, and O 12
2.7. Core-excited Basis Sets 13
2.8. Example Q-chem input files for deltaSCF 19
3. Addition Information 23
4. References 23

1. Method

1.1. Computational details

All calculations were carried out using the Q-Chem code.[1] The exchange and correlation energies were described using the B3LYP hybrid functional.[2] Total energy differences are computed via the ΔSCF methodology, using the Maximum Overlap Method (MOM)[3] to stably describe core-excited state self-consistent fields.

1.2. Core-Electron Binding Energy Calculations

XPS spectra measure the binding energy of core-electrons via photoemission. From a computational perspective, we aim to reproduce this measurement through calculation of the core-electron binding energy (CEBE) via total energy difference between the core-excited final state (N-1 electrons) and the initial ground state (N electrons):

$$\Delta E = E_{final}(N-1) - E_{initial}(N) \tag{1}$$

where ΔE is a many-body estimate of the binding energy of the excited electron. We focus mostly on DFT or hybrid DFT/HF calculations based on a single reference electron configuration. In order to access particular orbital excitations, we express the total energies as functions of their orbital occupations, n_i corresponding to the resultant single-particle orbitals ϕ_i:

$$E = E(n_1, n_2, n_3, \ldots, n_N) \tag{2}$$

In this way, we can define the core-excited state energy as the self-consistent energy assuming that a particular core orbital remains unoccupied ($n_i = 0$). From Eqs. (1) and (2), the ΔSCF CEBE could be given by

$$\Delta E = E(0,1,1,\ldots,1) - E(1,1,1,\ldots,1)\ , \tag{3}$$

if, for example, the first orbital corresponds to the core orbital we wish to excite. This is effectively accomplished using the MOM[3], by occupying those orbitals that overlap most with a reference set. The emptied core orbital becomes the first orbital in the virtual subspace. Particular care must be taken when there are symmetry equivalent atoms to be core excited, since the reference (ground) state may include a linear combination of core orbitals from multiple atoms, which would not represent a core excitation of a single atom.

1.3. Gaussian Basis-Set Orbital Representations

Each single-electron orbital is expressed as a linear combination of atomic (i.e., localized) orbitals ϕ_i,

$$\psi = \sum_{i=1}^{N} C_i \phi_i \quad . \tag{4}$$

Each ϕ_i is a Gaussian Type Orbital (GTO), which may be only one or a linear combination of Gaussian Type Basis Functions (GTFs) chosen to efficiently describe the atomic orbitals and their interactions with other atoms. In Cartesian coordinates, centered on the nucleus of a particular atom A, the GTFs can be written as:

$$\varphi_{ijk}(x_A,y_A,z_A;a) = N x_A^i y_A^j z_A^k \exp\left(-a r_A^2\right)$$
(6)

where the normalization factor is

$$N = \left(\frac{2a}{\pi}\right)^{3/4}\left[\frac{(8a)^{i+j+k} i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2}$$
(7)

and $r_A^2 = x_A^2 + y_A^2 + z_A^2$. The sum of the indices, $i+j+k = 0,1,2,3$ corresponds to the orbital angular momentum, s,p,d,f, respectively. The Gaussian exponent, a, defines the spatial extent of the GTF, as it is the inverse of twice the square of the half width at half maximum of the Gaussian.

A given basis set includes a fixed number of Gaussian exponentes, $\{a_n\}$, which span various spatial ranges. In order to reduce the computational complexity, GTFs with different exponents can be combined in fixed linear combinations called contractions for a given angular momentum:

$$\phi_i = \sum_{n=1}^{M} K_n \varphi(a_n) \qquad (5)$$

The coefficients, K_n, are called contraction terms. The contractions reduce the actual number of GTOs employed in the calculation and, in certain cases (e.g., Dunning's basis sets), effectively describe the orbitals of the isolated atom, with the variational freedom to describe orbital variations due to intra- and inter-molecular binding provided by additional (typically uncontracted, single exponent) GTOs.

1.4. Construction of the core-excited basis sets, motivated by Slater's Rule

In this work, we made use of direct comparisons between two numerical descriptions of the atomic orbitals of the elements: (1) numerical orbitals from grid-based finite difference calculations and (2) orbitals constructed using a given Gaussian basis set. For the numerical orbitals, we chose as benchmarks Perdew-Burke-Ernzerhof (PBE)[4] solutions for the atomic orbitals as a representative example, generated with the atomic code of Vanderbilt[5], provided as part of the ultrasoft pseudopotential generation method. These orbitals are generated on a log-radial grid with sufficient resolution to provide a numerically converged estimate. In principle, one could also use a fully uncontracted basis to within a calculation for the isolated atom to generate similar benchmarks. Our intent here was to avoid any inherent basis set limitation and was driven by our experience in generating pseudopotentials.

These benchmark orbitals were compared with various contracted GTOs from split valence basis sets, such as those of Pople[6] and Dunning[7,8]. Using the Cartesian GTF definition from Eqs. (6) and (7), we plot the radial components of the 1s, 2s, and 2p orbitals of the isolated carbon atom in its ground state electron configuration in **Figure S1 a,b,c**, respectively. We found excellent agreement between the grid-based orbitals and specific contractions in Dunning's cc-pVTZ basis (which were, of course, designed to reproduce the atomic natural orbitals (ANOs)[9]). If we do the same for the atomic orbitals of the core-excited state, we see clear differences between our benchmark core-excited orbitals and both the contractions from ground state C and N (Z+1), as

might be expected (**Figure S1 d,e,f**). Therefore, when applying such contracted basis sets, we might easily expect systematic errors in calculated CEBEs due to deficiencies in these numerical representations.

To construct an improved basis set, we take inspiration from Slater's Rules[10], which appeared in the 1930s. Slater's Rules provide a physically motivated origin for the screening constant σ which relates the effective nuclear charge Z_{eff} and actual charge *Z*, where

$$Z_{eff} = Z - \sigma \tag{8}$$

The numerical values for σ , following Slater's numerical formulae, are provided in **Table S1**, where we also include calculated values assuming $1s^1$ excitations. Notably, in Slater's original manuscript[10], the K-edge excitation of Fe was discussed as an illustrative example, indicating that the Slater's Rules are at least qualitatively applicable to excited states. A later update by Clementi [11] in the 1960s, consisting of finely tuned, numerically optimized Z_{eff} parameters based on ground-state Hartree-Fock calculations, did not explicitly cover core-excited states.

Slater's original intent was to employ Z_{eff} to rescale the exponent in Slater-type orbitals, which are exponential, not Gaussian, and nodeless (i.e., describing the outer, bonding region of a given orbital). We can use dimensional analysis to match the modification $Z_{eff} \rightarrow Z_{eff}^{'}$ induced by core excitation within the exponents of a Gaussian basis set, as follows:

$$\alpha_n \rightarrow \alpha_n^{'} = \alpha_n \left(\frac{Z_{eff}^{'}}{Z_{eff}}\right)^2 \tag{9}$$

where $\left(\frac{Z_{eff}^{'}}{Z_{eff}}\right)^2$ is the scaling factor that we report in **Table S1** as a good initial guess for constructing the specialized core-excited state basis set.

Chong et. al[12] experimented with scaling parameters from Clementi (see **Table S2**), employing generalized transition state (GTS) calculations of CEBEs and concluded that scaled basis sets did not perform satisfactorily. Another similar attempt from Millie *et al.*[13] used the simulated annealing procedure to optimize the contracted gaussian basis sets for ground state and core hole state (by finding those that generate the lowest electronic total energy), but concluded that the resulting CEBEs were not satisfactory either (at least within a desirable ~0.1 eV error).

Rather than focusing on optimizing the total energy, here we focus on gaussian basis sets for core-excited states that best reproduce the atomic orbitals, drawing inspiration from Slater's Rules. Direct comparison of the orbitals can reveal where Slater's Rules succeed and where they fail. For example, given that these rules were initially derived for Slater-type orbitals, without nodes, we might expect that a single rescaled exponent might have difficulty describing the effective screening of nuclear charge when applied to the full orbital with nodes close to the nucleus and a broad tail far away. We can see that Dunning had evolved a similar intuition by choosing the same exponents for both 1s and 2s orbitals within the cc-pVTZ series, while Slater might indicate that the same exponents should be used for 2s and 2p, with 1s belonging to a

different shell. Our calculations confirmed that Dunning's convention provides better reproduction of the core-excited orbitals when rescaling coefficients. In detail, we would advise that all s orbitals (1s, 2s, …) adopt the same scaled exponents, defined initially by Slater rescaling appropriate to the n=1 shell; all p orbitals (2p, …) adopt Slater rescaling for n=2; 3d uses n=3; etc.

Finally, within a given ground state contraction, merely rescaling the exponents is not sufficient to reproduce the core-excited orbitals that would match this GTO (although in some cases, such as 1s, it is very close). In addition, we also reoptimize the contraction terms, K_n, via linear regression to minimize the least squares error with the benchmark orbitals. This defines a stable procedure to move from a given ground-state basis set to its core-excited version by reference to some benchmark core-excited orbitals:

1) identify which contracted GTOs correspond to the benchmark orbitals.
2) rescale the exponents according to our modified Slater's rules based on angular momentum rather than principal quantum number.
3) reoptimize the contraction terms starting from the originals, employing linear regression.

In principle, one can also avoid the necessity of using benchmarks by following a similar procedure (not reported in this work, but seen as essentially equivalent):

1) rescale the exponents according to our modified Slater's rules based on angular momentum rather than principal quantum number.
2) fully uncontract the basis by separating all unique exponents into their own GTOs.
3) compute the optimal total energy (with appropriate occupancy constraint, e.g., via MOM) for the core-excited state (using whichever method/Hamiltonian you prefer).
4) use the computed optimal orbital coefficients to define new contractions analogous to the original ground state basis set.

Element	Atomic Number	State	Reference	1s	2s	2p	Scaling (1s)	Scaling (2s)	Scaling (2p)
H	1	gs	Slater	1.00					
He	2	gs	Slater	1.70	0.30				
Li	3	gs	Slater	2.70	1.30				
Be	4	gs	Slater	3.70	1.95	1.95			
B	5	gs	Slater	4.70	2.60	2.60			
C	6	gs	Slater	5.70	3.25	3.25			
N	7	gs	Slater	6.70	3.90	3.90			
O	8	gs	Slater	7.70	4.55	4.55			
F	9	gs	Slater	8.70	5.20	5.20			
Ne	10	gs	Slater	9.70	5.85	5.85			
H	1	1s1	Slater	N/A					
He	2	1s1	Slater	2.00	1.15				
Li	3	1s1	Slater	3.00	2.15				

Be	4	1s1	Slater	4.00	2.80	2.80	1.17	2.06	2.06
B	5	1s1	Slater	5.00	3.45	3.45	1.13	1.76	1.76
C	6	1s1	Slater	6.00	4.10	4.10	1.11	1.59	1.59
N	7	1s1	Slater	7.00	4.75	4.75	1.09	1.48	1.48
O	8	1s1	Slater	8.00	5.40	5.40	1.08	1.41	1.41
F	9	1s1	Slater	9.00	6.05	6.05	1.07	1.35	1.35
Ne	10	1s1	Slater	10.0	6.70	6.70	1.06	1.31	1.31

Table S1. Effective charge Z_{eff} derived from Slater's Rule for elements H to Ne, a similar table using Clementi's parameters can be found in **Table S2**: 'gs' refers to the ground state, '1s1' to the core-excited state, scaling refers to Eqs. (9) as an initial guess for constructing a new core-excited state basis.

Element	Atomic Number	State	Reference	1s	2s	2p	Scaling (1s)	Scaling (2s)	Scaling (2p)
H	1	gs	Clementi	1					
He	2	gs	Clementi	1.69					
Li	3	gs	Clementi	2.69	1.28				
Be	4	gs	Clementi	3.68	1.91				
B	5	gs	Clementi	4.68	2.58	2.42			
C	6	gs	Clementi	5.67	3.22	3.14			
N	7	gs	Clementi	6.67	3.85	3.83			
O	8	gs	Clementi	7.66	4.49	4.45			
F	9	gs	Clementi	8.65	5.13	5.1			
Ne	10	gs	Clementi	9.64	5.76	5.76			
H	1	1s1	Clementi	N/A					
He	2	1s1	Clementi	2					
Li	3	1s1	Clementi	2.99	2.14	N/A			
Be	4	1s1	Clementi	3.99	2.78	N/A	1.18	2.12	N/A
B	5	1s1	Clementi	4.98	3.43	3.28	1.13	1.77	1.84
C	6	1s1	Clementi	5.97	4.07	3.95	1.11	1.60	1.58
N	7	1s1	Clementi	6.96	4.71	4.62	1.09	1.50	1.46
O	8	1s1	Clementi	7.96	5.35	5.28	1.08	1.42	1.41
F	9	1s1	Clementi	8.95	6	5.95	1.07	1.37	1.36
Ne	10	1s1	Clementi	9.94	6.64	6.62	1.06	1.33	1.32

Table S2 Effective charge and scaling factor derived from Clementi's Rules [11].

2. Results

2.1. Ground and core-excited state orbitals for carbon

The changes in the atomic orbitals when a core electron is excited are provided in detail in **Figure S1** for the carbon atom. We note the following key points: (i) Consistency is clearly observed between the 1s, 2s, 2p ground state orbitals constructed from cc-pVTZ (based on ANO) and the numerically converged, grid-based all-electron (AE) solution for the orbitals (the defined

benchmark), as shown in **Figure S1 a, b, c**. This is unsurprising, as these efficiently contracted correlation-consistent basis sets describe ground state properties accurately. (ii) Neither carbon nor nitrogen ground-state orbitals resemble the benchmark core-excited orbitals of carbon, with the discrepancies visualized in **Figure S1 d, e, f**. We note that the largest differences are not observed in the core (1s) orbital, but rather, in the more diffuse (2p) orbitals of higher principal quantum number and angular momentum. We can rationalize this based on the rescaling proposed in **Tables S1** and **S2,** which is much more pronounced in the outer orbitals, such as 2p. Therefore, despite the appeal of the Z+1 approximation, substituting the nitrogen ground state basis contractions for carbon does not reproduce the core-excited orbitals. Similarly, we would also not expect any linear combination of the GTOs of nitrogen and carbon ground-state basis sets to reproduce the core-excited orbitals, given that different orbitals have different rescaling. (iii) Utilizing Slater's Rules, rescaling of the exponents and optimizing the contraction terms, we constructed a new core-excited basis for carbon, with the same level of computational complexity as the original basis that compares excellently with the benchmark orbitals. This basis reduced the CEBE error to 0.1 eV in comparison with experiments for a wide range of carbon containing molecules, outlined below.

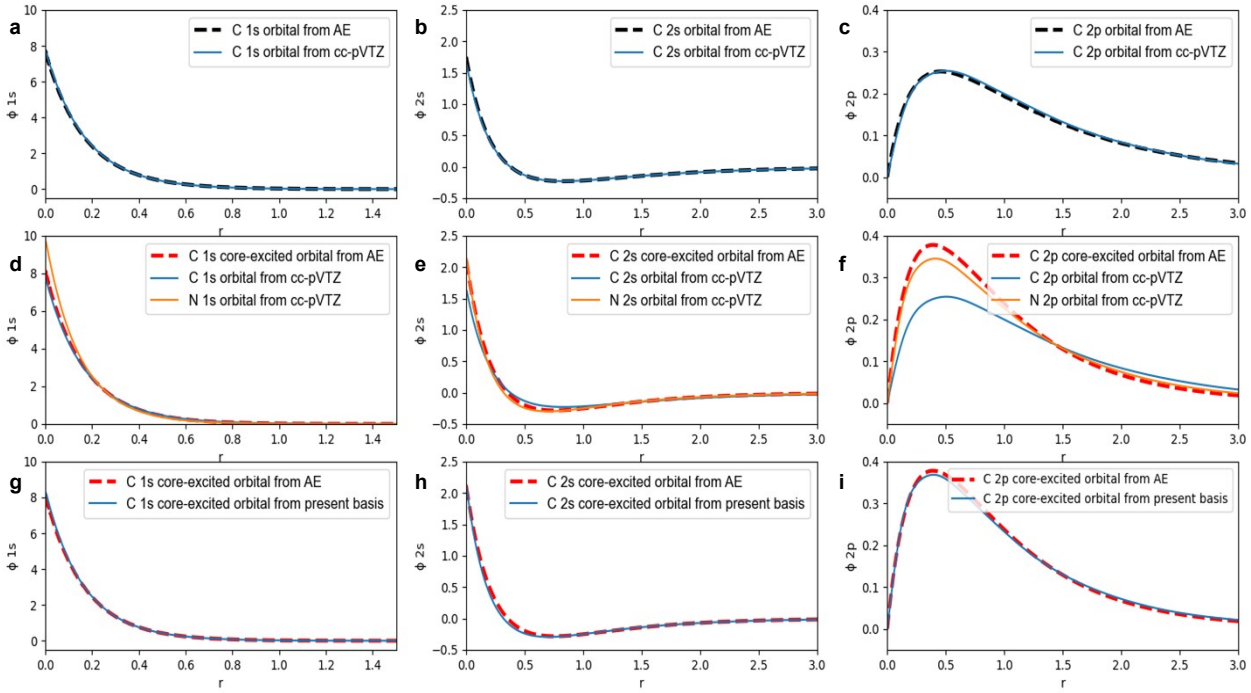

Figure S1 Plotting carbon orbitals generated from various basis sets benchmarking with grid-based all-electron (AE) solution, generated with the atomic code of Vanderbilt[5]. The first column (**a, d, g**) are zoom-in views of the 1s orbitals. The second column (**b, e, h**) are zoom-in views of the 1s orbitals. The third column (**c, f, i**) are zoom-in views of the 1s orbitals. The first row (**a, b, c**) are comparison of the ground state grid-based AE solution with those constructed from Dunning's cc-pVTZ basis. The second row (**d, e, f**) are comparison of the core-excited grid-based AE solution with those constructed from carbon and nitrogen (N+1) from Dunning's cc-pVTZ basis. The third row (**g, h, i**) are comparison of the core-excited grid-based AE solution with our newly constructed, Slater's Rule motivated basis sets.

2.2. Detailed results for carbon

We used the results in this table to plot **Figure 2** in the manuscript. We observe that the variational principle is consistently followed using our new Slater's Rule inspired basis sets.

Molecule	E(Initial State)	E(Final State) (C ground state)	E(Final State) (N ground state)	E(Final State) (Slater's Rule)	Charge on Carbon
CH_4	-40.5382	-29.8407	-29.8108	-29.8473	-0.4069
CH_3SiH_3	-331.2620	-320.5821	-320.5470	-320.5901	-0.3854
CH_3SH	-438.7558	-428.0340	-427.9984	-428.0416	-0.2820
HCN	-93.4614	-82.6630	-82.6273	-82.6724	-0.0875
CH_3OH	-115.7722	-105.0175	-104.9806	-105.0238	-0.0603
H_2NCN	-148.8455	-138.0502	-138.0098	-138.0606	-0.0494
CO	-113.3573	-102.4479	-102.3973	-102.4588	-0.0271
CCl_4	-1879.0244	-1868.1190	-1868.0693	-1868.1305	0.0638
H_2CO	-114.5494	-103.7146	-103.6774	-103.7220	0.1636
NH_2CHO	-169.9655	-159.1373	-159.0974	-159.1444	0.1743
HNCO	-168.7514	-157.8630	-157.8232	-157.8726	0.2587
CH_2F_2	-239.0818	-228.1868	-228.1436	-228.1919	0.3443
CO_2	-188.6606	-177.7054	-177.6624	-177.7150	0.3683
COF_2	-313.1411	-302.1225	-302.0759	-302.1293	0.4560
CHF_3	-338.3739	-327.3778	-327.3294	-327.3822	0.4829
CF_4	-437.6615	-426.5666	-426.5137	-426.5710	0.5446

Table S3 B3LYP total energies in Hartree for C-containing molecules in the initial (ground) state, employing the cc-pVTZ basis, and C $1s^1$ core-excited state MOM total energies computed using various cc-pVTZ (or similar) basis sets for the excited C atom: ground state C, ground state N, and our new basis inspired by Slater's Rules. The new basis set produces variationally improved total energies for the excited state. The Mulliken local charge on the core-excited C atom is provided and also used in **Figure 3** in the main manuscript.

2.3. The special case of CO

We revisit the peculiar case of CO in more details, as discussed in the manuscript. It has been reported by Thomas *et al.*[14] that the equilibrium bond length for core-ionized CO is 4.85 pm shorter than that of neutral CO; whereas for CF_4, the excitation of the C-F stretching mode is weak and the analysis shows the change in equilibrium CF bond length upon ionization is no more than 0.54pm. This anharmonicity analysis[14,15] in combination with the previous computational findings of the sensitive CO polarizability in the ground state[16] motivated us to examine the initial and final state energetics as a function of CO bond length and the local electronic charge on the C atom. Our findings indicate that the origin of the 0.4 eV error in CEBE may not necessarily derive from possible deficiencies in our theoretical method (B3LYP ΔSCF/MOM). Higher level calculations (CCSD, below) do not improve the error. Instead, the numerical results in **Table S5** indicate that we can effectively reduce the error to within 0.1 eV when including the effect of CO bond shrinkage in the excited state. By contrast, CF_4 indicates no significant change in either the local charge on C or the C 1s binding energy as a function of the C–F bond length. It is also interesting to note that our new core-excited state basis is able to accurately locate the shift of equilibrium bond length, or the equilibrium distance for the excited state, in comparison with the experiments.

Molecule	Bond Length	$E_{initial}$	E_{final}	Charge (initial)	Charge (final)	BE	Error	Comment
CO	1.0562	-113.3441	-102.4677	-0.115	0.395	295.96	-0.1	
CO	1.0662	-113.3479	-102.4681	-0.102	0.404	296.05	0.0	Equilibrium Distance for excited state (Calculated)
CO	1.0762	-113.3509	-102.4679	-0.089	0.413	296.14	0.0	Equilibrium Distance for excited state (Experimental)[15]
CO	1.0862	-113.3533	-102.4671	-0.076	0.423	296.23	0.1	
CO	1.0962	-113.3551	-102.4657	-0.064	0.432	296.31	0.2	
CO	1.1062	-113.3563	-102.4639	-0.051	0.441	296.40	0.3	
CO	1.1162	-113.3570	-102.4615	-0.039	0.451	296.48	0.4	
CO	1.1262	-113.3573	-102.4588	-0.027	0.460	296.56	0.5	Equilibrium Distance for ground state (Calculated)
CO	1.1362	-113.3570	-102.4556	-0.015	0.469	296.64	0.5	
CO	1.1462	-113.3564	-102.4521	-0.003	0.478	296.72	0.6	
CF_4	1.2958	-437.6609	-426.5711	0.534	0.641	301.77	0.0	
CF_4	1.3058	-437.6613	-426.5712	0.538	0.643	301.78	0.0	
CF_4	1.3158	-437.6615	-426.5712	0.541	0.646	301.78	0.0	Equilibrium Distance for excited state (Calculated)
CF_4	1.3258	-437.6615	-426.5710	0.545	0.648	301.79	0.0	Equilibrium Distance for excited state (Experimental)[15]
CF_4	1.3358	-437.6615	-426.5708	0.548	0.649	301.79	0.0	Equilibrium Distance for ground state (Calculated)
CF_4	1.3458	-437.6613	-426.5704	0.551	0.651	301.80	0.0	
CF_4	1.3558	-437.6610	-426.5699	0.554	0.652	301.80	0.0	

Table S4 Initial (ground-state) and final (core-excited) energies (in Ha, computed using B3LYP and MOM), local C charge (electrons), and CEBE and associated experimental error (eV) of CO and CF_4 for various bond lengths (Å) – plotted in **Figure 4** of the main manuscript.

2.4. Comparison of results obtained through CCSD vs. B3LYP

As discussed in the manuscript, it is tempting in these model calculations of small molecules to try a higher level of theory such as coupled cluster	**basis**	**molecule**	**Ei(Ha)**	**Ef(Ha)**	**CalBE**	**expBE**	**error**

including single and double excitations (CCSD) for describing the more exact correlation energy. In the case of the CO CEBE error, applying a higher level of theory does not solve the problem. We find that CCSD does not resolve the error, nor does it consistently improve the CEBE estimates (but rather makes them worse – which warrants further investigation beyond this work). Even if we consider a systematic offset in the energy estimates, the so-called chemical shift							

between CO and CH_4 is measured to be +5.4 eV, with B3LYP producing a chemical shift of +5.7643 eV, while CCSD estimates a shift of +5.9209 eV, that is –0.1566 eV worse than B3LYP. Instead, our proposed explanation for the CO outlier is related to a geometry change in the final state, as discussed below in **Table S5** and in the main text.**functional**							
B3LYP	cc-PVTZ	CH_4	-40.538242	-29.84068	291.0955	290.8	-0.2955
B3LYP	cc-PVTZ	CH_3SiH_3	-331.261966	-320.582089	290.6143	290.3	-0.3143
B3LYP	cc-PVTZ	CH_3SH	-438.755803	-428.033956	291.7563	291.4	-0.3563
B3LYP	cc-PVTZ	HCN	-93.461381	-82.663036	293.8379	293.5	-0.3379
B3LYP	cc-PVTZ	CH_3OH	-115.772247	-105.017493	292.6518	292.4	-0.2518
B3LYP	cc-PVTZ	H_2NCN	-148.845466	-138.050236	293.7532	293.5	-0.2532
B3LYP	cc-PVTZ	CO	-113.357258	-102.44786	296.8598	296.2	**-0.6598**
CCSD	cc-PVTZ	CH_4	-40.213385	-29.526221	290.8126	290.8	-0.0126
CCSD	cc-PVTZ	CH_3SiH_3	-330.322478	-319.655244	290.2702	290.3	0.0298
CCSD	cc-PVTZ	CH_3SH	-437.756551	-427.044351	291.4938	291.4	-0.0938
CCSD	cc-PVTZ	HCN	-92.908119	-82.13483	293.1561	293.5	**0.3439**
CCSD	cc-PVTZ	CH_3OH	-115.089457	-104.341733	292.4605	292.4	-0.0605
CCSD	cc-PVTZ	H_2NCN	-147.966987	-137.18283	293.4519	293.5	0.0481

CCSD	cc-PVTZ	CO	-112.780287	-101.875534	296.7335	296.2	**-0.5335**

Table S5 Performance comparison of B3LYP and CCSD with cc-pVTZ basis sets when applied to estimates of the CEBE for CO and CH_4. For other molecules such as CH_3SiH_3, CH_3SH, etc. where B3LYP sufficiently describes the relative CEBE, the improvement of CCSD on the accuracy is not very pronounced. Note that the geometries in B3LYP and CCSD are updated accordingly and self-consistently. The cases with big relative CEBE errors are outlined using bold numbers.

2.5. Ground and core-excited state orbitals for B, N, and O

To test the transferability of our core-excited state basis construction beyond carbon, we provide results for other light atoms: boron, nitrogen, and oxygen. Comparing the B atomic orbitals constructed using various basis sets (**Figure S2)**, we find that rescaling exponents alone based on Slater's Rules is sufficiently accurate to reproduce our benchmark, grid-based, all-electron (AE) orbitals in the B $1s^1$ core-excited state.

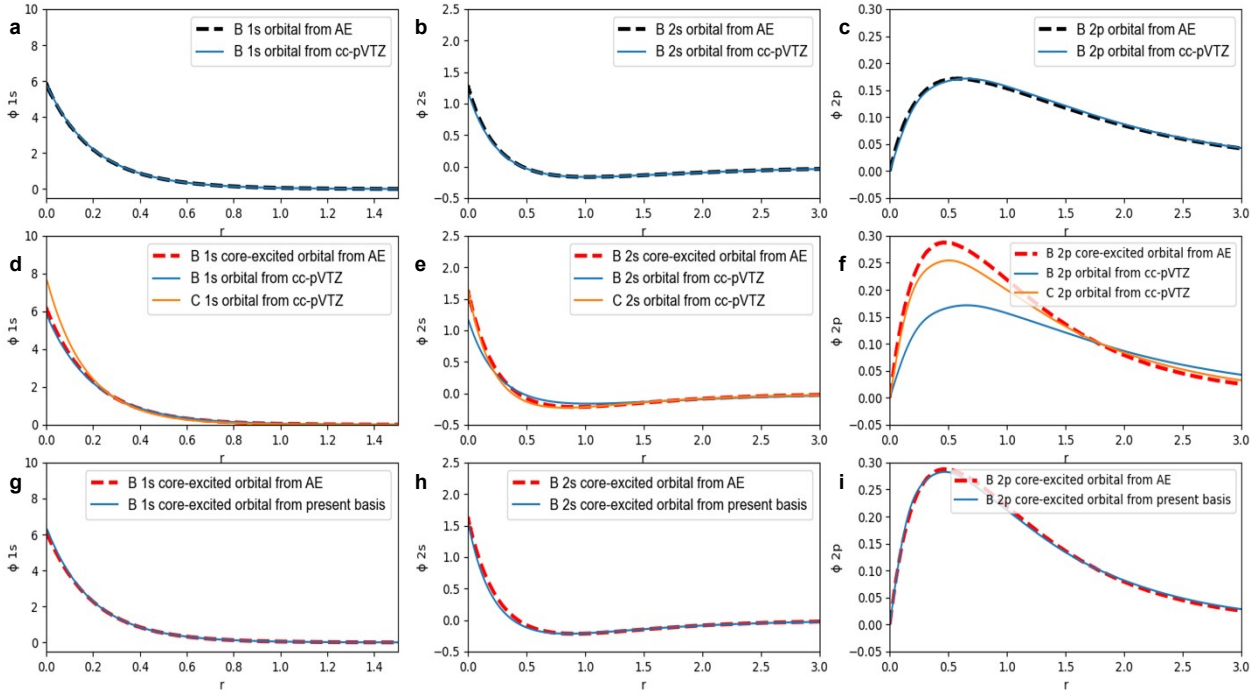

Figure S2 B atomic orbitals generated using contractions from various basis sets (B ground state, C ground state as Z+1, and Slater rescaled) in comparison with benchmark all-electron (AE) grid-based orbitals generated with the atomic code of Vanderbilt[5].

For nitrogen, we find that although rescaling according to Slater provides a good starting point to reproduce the core-excited orbitals, it is no longer sufficiently accurate, especially for the 2p orbital. Empirically, we also tried fixing the exponents according to Slater's Rule (**Table S1**) and used the contraction coefficients of oxygen GTOs to achieve better agreement. Ultimately, as long as the core-excited orbitals are reproduced to a sufficient degree by the GTO contractions (**Figure S3**) we would hope that they reduce systematic errors in describing the core-excited state. Our optimized basis set is provided in Section **2.9** and plotted in **Figure S3.**

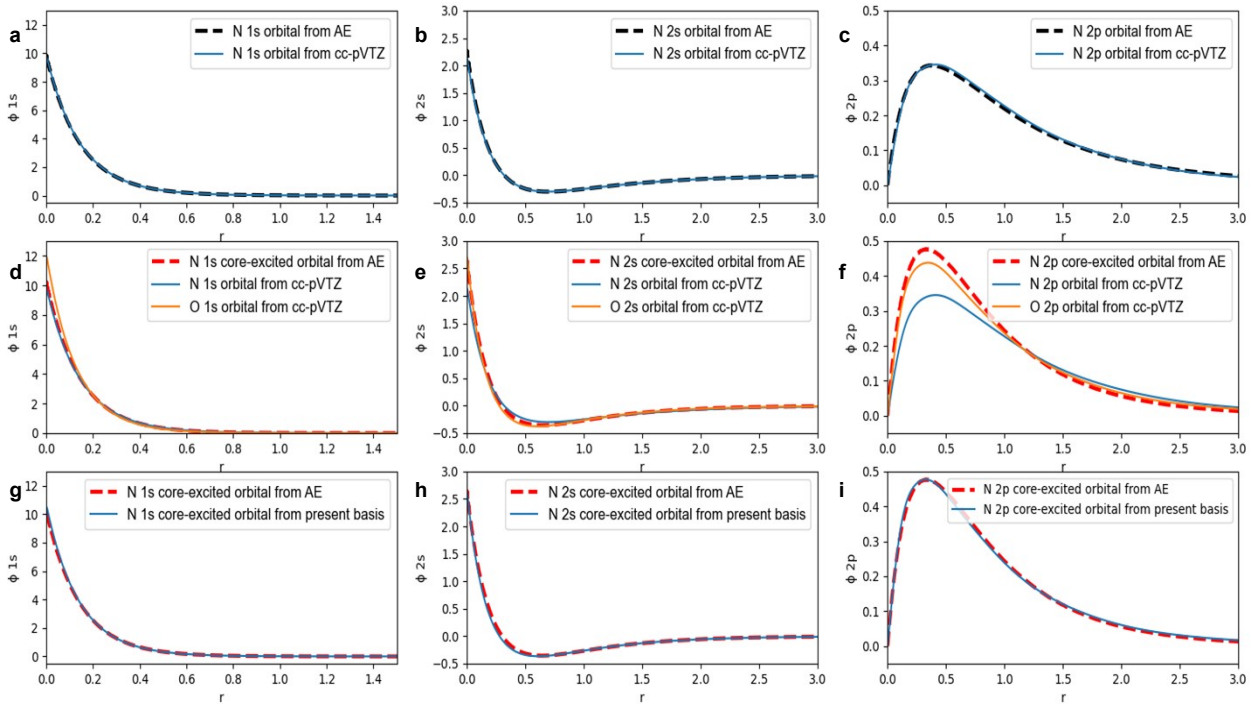

Figure S3 N atomic orbitals generated using contractions from various basis sets (N ground state, O ground state as Z+1, and Slater rescaled) in comparison with benchmark all-electron (AE) grid-based orbitals generated with the atomic code of Vanderbilt[5].

As expected, when we move further into the periodic table, to oxygen, we find that the scaling factor based on Slater's Rules provides a good starting point, but it is no longer sufficiently accurate in comparison to the benchmark grid-based AE orbitals for the oxygen core-excited state. In this case, optimizing exponents first and then optimizing coefficients achieve good agreement with the grid based AE orbitals (**Figure S4**). All our optimized basis sets are provided below in Section **2.7**.

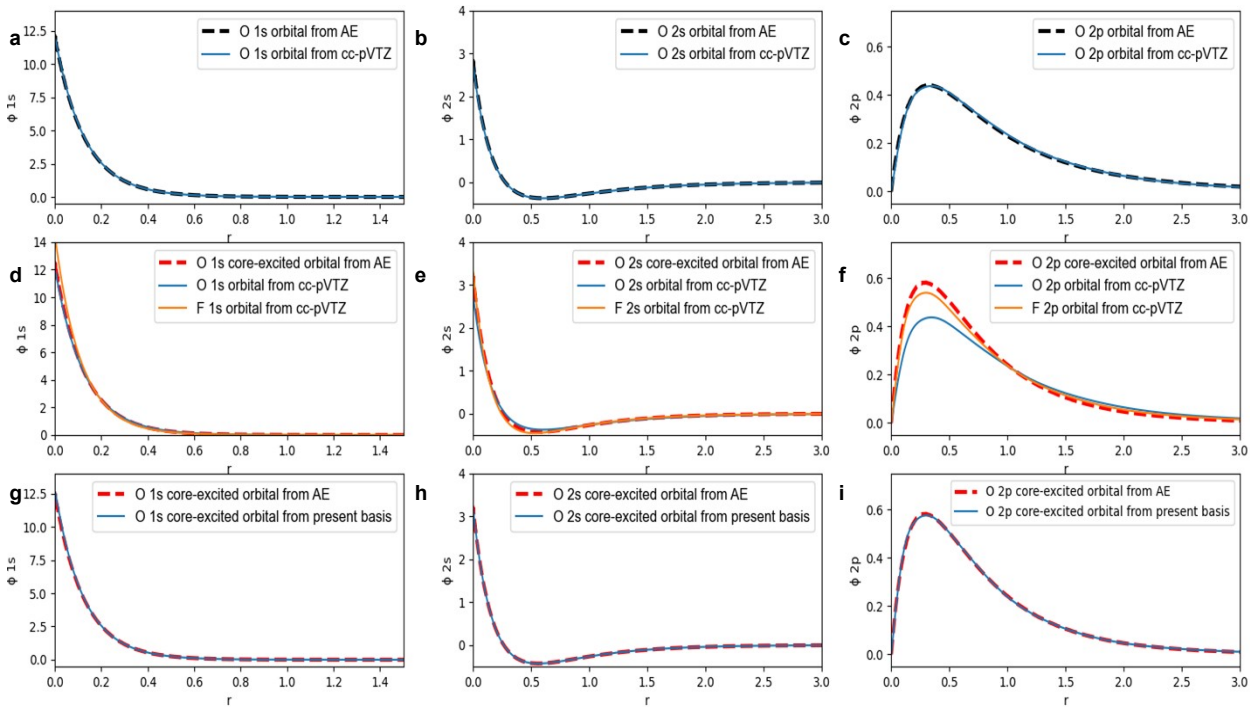

Figure S4 O atomic orbitals generated using contractions from various basis sets (N ground state, O ground state as Z+1, and Slater rescaled) in comparison with benchmark all-electron (AE) grid-based orbitals generated with the atomic code of Vanderbilt[5].

2.6. Energies and charges for B, N, and O

We observe that the variational principle is consistently followed using our computational and basis set strategies (**Secs 1.1, 1.2, 1.3**) for these additional elements: boron, nitrogen and oxygen. Discussion of the accuracy of CEBE estimates for a broad range of elements will be explored in a future publication.

Molecule	E(Initial State)	E(Final State) (ground state)	E(Final State) (N+1 ground state)	E(Final State) (Slater's Rule)	Charge on excited atom
BH3CO	-140.0277	-132.8485	-132.8266	-132.8542	-0.3960
B(OCH3)3	-370.5255	-363.2451	-363.2133	-363.2511	0.2121
BF3	-324.6964	-317.2373	-317.1997	-317.2446	0.5209
CH3**N**H2	-95.9000	-81.0076	-80.9610	-81.0130	-0.3323
NH3	-56.5847	-41.6758	-41.6277	-41.6796	-0.4596
(CH3)C**N**	-132.8068	-117.9014	-117.8606	-117.9058	-0.0393
NH2CHO	-169.9655	-155.0230	-154.9793	-155.0282	-0.2028
HC**N**	-93.4614	-78.5060	-78.4626	-78.5109	-0.0631
H2C**O**	-114.5494	-94.7296	-94.6740	-94.7452	-0.2396
H2**O**	-76.4598	-56.6245	-56.5686	-56.6390	-0.4355
C**O**F2	-313.1411	-293.2780	-293.2240	-293.2939	-0.2134
C**O**	-113.3573	-93.4182	-93.3625	-93.4343	0.0271

Table S6 B3LYP total energies in Hartree for B, N, and O-containing molecules in the initial (ground) state, employing the cc-pVTZ basis, and B, N, O $1s^1$ core-excited state MOM total energies computed using various cc-pVTZ (or similar) basis sets for the excited atom (in bold): ground state, ground state of Z+1, and our new basis inspired by Slater's Rules. The new basis set produces variationally improved total energies for the excited state. The Mulliken local charge on the core-excited atom is also provided.

2.7. Core-excited Basis Sets

We provide the final contracted core-excited basis sets that we constructed for boron, carbon, nitrogen, and oxygen, consistent with the cc-pVTZ family. These basis sets are provided in the Q-Chem[1] format, but they can be easily converted for use in any other quantum chemistry code.

#start B

B 0

S 10 1.00

6.18E+03 5.55E-04

9.28E+02 4.29E-03

2.11E+02 2.19E-02

5.97E+01 8.44E-02

1.93E+01 2.39E-01

6.78E+00 4.35E-01

2.50E+00 3.42E-01

6.64E-01 3.69E-02

2.73E-01 -9.55E-03

9.73E-02 2.37E-03

S 10 1.00

6.18E+03 -1.12E-04

9.28E+02 -8.68E-04

2.11E+02 -4.48E-03

5.97E+01 -1.77E-02

1.93E+01 -5.36E-02

6.78E+00 -1.19E-01

2.50E+00 -1.66E-01

6.64E-01 1.20E-01

2.73E-01 5.96E-01

9.73E-02 4.11E-01

S 1 1.00

6.64E-01 1.00E+00

S 1 1.00

9.73E-02 1.00E+00

P 5 1.00

2.12E+01 1.31E-02

4.60E+00 7.99E-02

1.32E+00 2.77E-01

4.20E-01 5.04E-01

1.35E-01 3.54E-01

P 1 1.00

4.20E-01 1.00E+00

P 1 1.00

1.35E-01 1.00E+00

D 1 1.00

1.16E+00 1.00E+00

D 1 1.00

3.50E-01 1.00E+00

F 1 1.00

8.62E-01 1.00E+00

****

#end B

#start C

C 0

S 10 1.00

9.142E+03 5.310E-04

1.371E+03 4.108E-03

3.117E+02 2.109E-02

8.799E+01 8.185E-02

2.840E+01 2.348E-01

9.987E+00 4.344E-01

3.684E+00 3.461E-01

1.006E+00 3.938E-02

4.044E-01 -8.983E-03

1.426E-01 2.385E-03

S 10 1.00

9.142E+03 -1.130E-04

1.371E+03 -8.780E-04

3.117E+02 -4.540E-03

8.799E+01 -1.813E-02

2.840E+01 -5.576E-02

9.987E+00 -1.269E-01

3.684E+00 -1.704E-01

1.006E+00 1.404E-01

4.044E-01 5.987E-01

1.426E-01 3.954E-01

S 1 1.00

1.006E+00 1.000E+00

S 1 1.00

1.426E-01 1.000E+00

P 5 1.00

2.975E+01 1.403E-02

6.571E+00 8.687E-02

1.908E+00 2.902E-01

6.085E-01 5.010E-01

1.922E-01 3.434E-01

P 1 1.00

6.085E-01 1.000E+00

P 1 1.00

1.922E-01 1.000E+00

D 1 1.00

1.744E+00 1.000E+00

D 1 1.00

5.056E-01 1.000E+00

F 1 1.00

1.210E+00 1.000E+00

****

#end C

#start N

N 0

S 10 1.00

1.24E+04 5.23E-04

1.87E+03 4.05E-03

4.24E+02 2.08E-02

1.20E+02 8.07E-02

3.88E+01 2.33E-01

1.37E+01 4.34E-01

5.06E+00 3.47E-01

1.41E+00 4.13E-02

5.58E-01 -8.51E-03

1.95E-01 2.38E-03

S 10 1.00

1.50E+04 -1.15E-04

2.24E+03 -8.95E-04

5.10E+02 -4.62E-03

1.44E+02 -1.85E-02

4.66E+01 -5.73E-02

1.64E+01 -1.32E-01

6.08E+00 -1.73E-01

1.69E+00 1.52E-01

6.70E-01 6.00E-01

2.34E-01 3.87E-01

S 1 1.00

1.69E+00 1.00E+00

S 1 1.00

2.34E-01 1.00E+00

P 5 1.00

3.94E+01 1.59E-02

8.80E+00 9.97E-02

2.58E+00 3.10E-01

8.21E-01 4.91E-01

2.55E-01 3.36E-01

P 1 1.00

8.21E-01 1.00E+00

P 1 1.00

2.55E-01 1.00E+00

D 1 1.00

2.63E+00 1.00E+00

D 1 1.00

6.94E-01 1.00E+00

F 1 1.00

1.62E+00 1.00E+00

****

#end N

#start O

O 0

S 10 1.00

1.61E+04 5.08E-04

2.41E+03 3.93E-03

5.49E+02 2.02E-02

1.55E+02 7.92E-02

4.99E+01 2.31E-01

1.76E+01 4.33E-01

6.52E+00 3.50E-01

1.84E+00 4.27E-02

7.23E-01 -8.15E-03

2.50E-01 2.38E-03

S 10 1.00

1.61E+04 -1.87E-04

2.41E+03 -7.52E-04

5.49E+02 -5.74E-03

1.55E+02 -1.95E-02

4.99E+01 -7.02E-02

1.76E+01 -1.55E-01

6.52E+00 -1.90E-01

1.84E+00 2.59E-01

7.23E-01 6.41E-01

2.50E-01 2.53E-01

S 1 1.00

2.52E+00 1.00E+00

S 1 1.00

3.43E-01 1.00E+00

P 5 1.00

4.96E+01 1.67E-02

1.12E+01 8.97E-02

3.28E+00 3.35E-01

1.03E+00 4.99E-01

3.08E-01 2.96E-01

P 1 1.00

1.03E+00 1.00E+00

P 1 1.00

3.08E-01 1.00E+00

D 1 1.00

3.33E+00 1.00E+00

D 1 1.00

9.29E-01 1.00E+00

F 1 1.00

2.06E+00 1.00E+00

****

#end O

2.8. Example Q-chem input files for deltaSCF

All structures were first optimized using B3LYP functionals with cc-pVTZ basis, following the example input file in Q-chem shown below as the first run:

$molecule

0 1

C 1.0582 0.9352 0.8103

H 1.4165 1.5700 -0.0048

H 1.2073 1.4482 1.7644

H -0.0062 0.7282 0.6702

H 1.6154 -0.0056 0.8114

$end

$rem

JOB_TYPE OPT

method B3LYP

basis cc-pVTZ

MAX_CIS_CYCLES = 103

PRINT_ORBITALS = TRUE

$end

For the second run, we use the optimized geometry from the first run, and perform a delta-SCF calculation with two subtasks: a) ground state single point SCF calculation with cc-pVTZ basis,

or the so-called initial state calculation b) core-excited state single point SCF calculation with the present Slater's rule motivated basis, or the so-called final state calculation. Example input file is provided below for clarity.

```
$molecule
0,1
C
H  1 1.088311
H  1 1.088320 2 109.474233
H  1 1.088330 2 109.474115 3 -120.003199 0
H  1 1.088336 2 109.470119 3 119.999047 0
$end

$rem
JOB_TYPE SP
method B3LYP
MAX_CIS_CYCLES  = 100
PRINT_ORBITALS = TRUE
basis cc-pVTZ
$end

@@@
$molecule
+1,2
C
H  1 1.088311
H  1 1.088320 2 109.474233
H  1 1.088330 2 109.474115 3 -120.003199 0
H  1 1.088336 2 109.470119 3 119.999047 0
$end
```

$rem

JOB_TYPE SP

UNRESTRICTED TRUE

method B3LYP

MAX_CIS_CYCLES = 100

SCF_GUESS read

MOM_START 1

BASIS GEN

PURECART 11

$end

$basis

C   0

S   10   1.00

9.142E+03   5.310E-04

1.371E+03   4.108E-03

3.117E+02   2.109E-02

8.799E+01   8.185E-02

2.840E+01   2.348E-01

9.987E+00   4.344E-01

3.684E+00   3.461E-01

1.006E+00   3.938E-02

4.044E-01   -8.983E-03

1.426E-01   2.385E-03

S   10   1.00

9.142E+03   -1.130E-04

1.371E+03   -8.780E-04

3.117E+02 -4.540E-03

8.799E+01 -1.813E-02

2.840E+01 -5.576E-02

9.987E+00 -1.269E-01

3.684E+00 -1.704E-01

1.006E+00 1.404E-01

4.044E-01 5.987E-01

1.426E-01 3.954E-01

S 1 1.00

1.006E+00 1.000E+00

S 1 1.00

1.426E-01 1.000E+00

P 5 1.00

2.975E+01 1.403E-02

6.571E+00 8.687E-02

1.908E+00 2.902E-01

6.085E-01 5.010E-01

1.922E-01 3.434E-01

P 1 1.00

6.085E-01 1.000E+00

P 1 1.00

1.922E-01 1.000E+00

D 1 1.00

1.744E+00 1.000E+00

D 1 1.00

5.056E-01 1.000E+00

F 1 1.00

1.210E+00 1.000E+00

```
****
H     0
S   5   1.00
      3.387000D+01          6.068000D-03
      5.095000D+00          4.530800D-02
      1.159000D+00          2.028220D-01
      3.258000D-01          5.039030D-01
      1.027000D-01          3.834210D-01
S   1   1.00
      3.258000D-01          1.000000D+00
S   1   1.00
      1.027000D-01          1.000000D+00
P   1   1.00
      1.407000D+00          1.000000D+00
P   1   1.00
      3.880000D-01          1.000000D+00
D   1   1.00
      1.057000D+00          1.0000000
****
$end


$occupied
1:5
1:0 2:5
$end
```

There are several calculation setups and physical facts that are worth pointing out:

First of all, the structure for the deltaSCF in initial and final state should be kept consistent, as the timescale for core-excitation in the calculation of CEBE is too short for nuclear motion of the molecule.

Secondly, only the core-excited atom needs the present Slater's rule motivated basis to describe the screening effect of the core-hole. Other atoms in the molecule should still adopt the cc-pVTZ basis. With the current proposed scheme of creating basis from grid-based AE solution, the interaction energy of core-excited atom and its surrounding atoms is already, in principle, reflected through the construction of the grid-based AE solution modified from Vanderbilt [5] self-consistent pseudopotentials.

3. Additional Information

Scripts for this article can be provided upon request.

4. References

(1) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; III, H. L. W.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; Jr, R. A. D.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; III, W. A. G.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; III, H. F. S.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Molecular Physics* **2015**, *113* (2), 184–215. https://doi.org/10.1080/00268976.2014.952696.

(2) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.

(3) 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* (50), 13164–13171. https://doi.org/10.1021/jp801738f.

(4) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(5) Vanderbilt, D. Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B* **1990**, *41* (11), 7892–7895. https://doi.org/10.1103/PhysRevB.41.7892.

(6) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72* (1), 650–654. https://doi.org/10.1063/1.438955.

(7) Dunning, T. H. Gaussian Basis Functions for Use in Molecular Calculations. I. Contraction of (9s5p) Atomic Basis Sets for the First-Row Atoms. *J. Chem. Phys.* **1970**, *53* (7), 2823–2833. https://doi.org/10.1063/1.1674408.

(8) 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* (2), 1007–1023. https://doi.org/10.1063/1.456153.

(9) Almlöf, J.; Taylor, P. R. General Contraction of Gaussian Basis Sets. I. Atomic Natural Orbitals for First- and Second-row Atoms. *J. Chem. Phys.* **1987**, *86* (7), 4070–4077. https://doi.org/10.1063/1.451917.

(10) Slater, J. C. Atomic Shielding Constants. *Phys. Rev.* **1930**, *36* (1), 57–64. https://doi.org/10.1103/PhysRev.36.57.

(11) Clementi, E.; Raimondi, D. L. Atomic Screening Constants from SCF Functions. *J. Chem. Phys.* **1963**, *38* (11), 2686–2689. https://doi.org/10.1063/1.1733573.

(12) Chong, D. P.; Hu, C.-H.; Duffy, P. Accurate Density-Functional Calculation of Core-Electron Binding Energies with a Scaled Polarized Triple-Zeta Basis Set. Twelve Test Cases and Application to Three C2H4O2 Isomers. *Chemical Physics Letters* **1996**, *249* (5), 491–495. https://doi.org/10.1016/0009-2614(95)01442-X.

(13) Carniato, S.; Millié, P. Accurate Core Electron Binding Energy Calculations Using Small 6-31G and TZV Core Hole Optimized Basis Sets. *J. Chem. Phys.* **2002**, *116* (9), 3521–3532. https://doi.org/10.1063/1.1446025.

(14) Ehara, M.; Kuramoto, K.; Nakatsuji, H.; Hoshino, M.; Tanaka, T.; Kitajima, M.; Tanaka, H.; De Fanis, A.; Tamenori, Y.; Ueda, K. C1s and O1s Photoelectron Satellite Spectra of CO with Symmetry-Dependent Vibrational Excitations. *J Chem Phys* **2006**, *125* (11), 114304. https://doi.org/10.1063/1.2346683.

(15) Carroll, T. X.; Børve, K. J.; Sæthre, L. J.; Bozek, J. D.; Kukk, E.; Hahne, J. A.; Thomas, T. D. Carbon 1s Photoelectron Spectroscopy of CF4 and CO: Search for Chemical Effects on the Carbon 1s Hole-State Lifetime. *J. Chem. Phys.* **2002**, *116* (23), 10221–10228. https://doi.org/10.1063/1.1476933.

(16) Peterson, K. A.; Dunning, T. H. The CO Molecule: The Role of Basis Set and Correlation Treatment in the Calculation of Molecular Properties. *Journal of Molecular Structure: THEOCHEM* **1997**, *400*, 93–117. https://doi.org/10.1016/S0166-1280(97)90271-2.