



Cite this: DOI: 10.1039/d5cp04944j

Error-consistent basis sets for large-core ECPs for elements La–Lu

 Marcel Lukowski and Florian Weigend *

We present the extension of the system of error-consistent segmented contracted Gaussian basis sets (Karlsruhe def2-bases [Weigend and Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305]) for the lanthanide large core effective core potentials (lcECPs) designed by Dolg, Stoll, Savin and Preuss, *Theor. Chim. Acta*, 1989, **75**, 173–194. For La–Lu, sets of double zeta (“split”, S), triple zeta (TZ), and quadruple zeta (QZ) valence (V) quality were optimized in atomic Hartree–Fock calculations for each of the different lcECPs that model the occupations f^{n-k} , $k = 0, 1, 2$, and n being the (typical) ground state f shell occupation; e.g. for Pr, $n = 3$, for each of the occupations f^3 , f^2 , and f^1 , an SV, TZV, and QZV basis were optimized and termed lcecp- k -XV ($k = 0, 1, 2$, X = S, TZ, QZ). Polarization functions for the quadruple zeta valence bases were taken from Weigand, Cao, Yang, and Dolg, *Theor. Chem. Acc.*, 2010, **126**, 117–127, for the smaller basis sets they were appropriately reduced. The conformity with the def2-series in regards to error-consistency was assessed for a set of 120 molecules by comparing distances, bond angles, vibration frequencies and exchange reaction energies in regards to the basis set limit and also to all-electron scalar relativistic calculations.

 Received 18th December 2025,
 Accepted 30th January 2026

DOI: 10.1039/d5cp04944j

rsc.li/pccp

Introduction

Rare-earth elements are indispensable for a range of modern technologies such as wind turbine generators, low-energy lighting, fuel cells, rechargeable batteries, magnetic refrigeration and hydrogen storage. Interest in these elements is reflected in lively research activity, both on solid phases and on molecules, and both experimentally and in calculations. The latter bear complexities arising from the partially occupied f shells that make quantum chemical calculations often tedious if not impossible. The explicit treatment of the f shell of course is inevitable for some of the properties that lanthanides are famous for, e.g. magnetic properties or electronic transitions from, to or within the f shells, but for properties being more related to the energy surface, like e.g. equilibrium structures, IR frequencies or reaction energies or enthalpies, the explicit consideration of f electrons is often not necessary, as the 4f shell is spatially much less extended than the 6s/5d shells and thus does much less overlap with neighbouring atoms. This motivated the development of so-called large core effective core potentials (lcECPs) for lanthanides that cover not only the core-electrons but also the f electrons although the corresponding shells are only partially filled. A consistently optimized system of relativistically corrected lcECPs, which is complete in the

sense of chemical relevance, was provided by Dolg, Stoll, Savin and Preuss (DSSP) already in 1989.¹ These Wood–Boring type ECPs cover the completely filled inner shells 1s–4d as well as n , $n - 1$ or $n - 2$ electrons of the f shell, where we label $n(\text{La}) = 1$, $n(\text{Ce}) = 2, \dots, n(\text{Lu}) = 15$ in this work. ECPs with $n - 1$ f electrons (f^{n-1} ECPs, in the following termed “lcecp-1”), which is the usual f occupation in $\text{Ln}(\text{III})$ compounds, are available for La–Lu, such with n electrons (f^n ECPs, in the following termed “lcecp-0”), which is the ground state occupation of the f shell for most of the lanthanides, for La–Yb. After several years of use in quantum chemistry practice, the f-parts of these ECPs for La–Yb were adjusted² to model non-integer fillings of the f shells, $4f^{n+x}$, with x being small and resulting from a small charge transfer from the ligand to the lanthanide. The basis sets presented in this work are designed for the use with these revised ECPs. Later, also ECPs with $n - 2$ f electrons (f^{n-2} ECPs, in the following termed “lcecp-2”) were provided³ for Ce, Pr, Nd, Tb and Dy where this occupation can be adopted if the lanthanide is surrounded by four or more highly electronegative elements. Together with the ECPs, DSSP also provided $(7s6p5d/[5s4p3d]\{31111/3111/311\}$ basis sets, i.e. mid-sized sets without polarization functions. In 2005, the sets for lcecp-1 were re-optimized for solid-state treatments avoiding too diffuse functions that can be critical for this purpose,⁴ and expanded by three polarizing f and two g functions, which are available within a study of lanthanide and actinide trifluorides by Weigand, Cao, Yang and Dolg (WCYD).⁵ At about the same time, in our group a system of polarized basis sets of double,

Institute for Quantum Materials and Technologies, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany.
 E-mail: florian.weigend@kit.edu



triple and quadruple zeta quality was designed, which targets consistent errors in molecular (ground-state) properties for the entire periodic table, the so-called def2-basis sets,⁶ with a (mainly) correlation-consistent polarization set and, where possible, a smaller polarization set for DFT calculations. For H–Kr, these are non-relativistic all-electron basis sets, for heavier elements they were designed to be used together with Wood–Boring ECPs⁷ (for the s and d elements) or Dirac–Fock ECPs (for the p elements).⁸ At that time, lanthanides were excluded, but corresponding sets for the usage in connection with small-core Wood–Boring ECPs,⁹ *i.e.* for the explicit treatment of f electrons, were delivered later.¹⁰ Moreover, the bases of all s/p/d elements were re-optimized for the usage in connection with Dirac–Hartree–Fock ECPs¹¹ and further equipped with extensions for two-component procedures (termed “dhf-bases”).¹² Furthermore, also the corresponding system of basis sets for all-electron relativistic calculations optimized for the one-electron exact two-component method (X2C)^{13,14} meanwhile is available.^{15,16} For lanthanides, the latter show usually less convergence problems in SCF procedures than the small-core ECP bases; nevertheless, the treatment of open f shells remains tedious, in particular, if more than one Ln atom is present in the molecule and/or if the geometric structure is not close to a minimum structure, as it is often the case at the beginning of a structure optimization. We thus decided to extend the system of error-consistent basis sets to sets to be used with the f^n , f^{n-1} and f^{n-2} lceCPs and labelled them lcecp-*k*-XVP (*k* = 0, 1, 2; X = S, TZ, QZ). In the following we describe the main ideas behind the development and discuss the results of molecular test calculations. The bases are provided as SI (file ln-lcecp-bases.txt). All calculations were done with TURBOMOLE,¹⁷ energies were converged to $10^{-9} E_h$, for DFT fine grids¹⁸ (5 for ECP calculations, 5a¹⁹ for all-electron calculations) were employed.

Atomic occupations and valence basis sets

Atomic ground states of lanthanides typically show f^n occupation, except for Ce, Gd, La and Lu, which show f^{n-1} occupation; the latter is also the usual occupation in the typical oxidation state +III. Further, for a given f occupation, there are several possible s/d occupations of similar energy, which thus need to be described with similar accuracy by a given basis set: s^1 and d^1 for cations with f^n occupation (for brevity termed f^n cation), s^2 , d^2 and s^1d^1 for the f^n atom and the f^{n-1} cation, s^2d^1 , s^1d^2 and d^3 for the f^{n-1} atom and the f^{n-2} cation and s^2d^2 , s^1d^3 and d^4 for the f^{n-2} atom. For an overview of the energetic situation, we calculated restricted open-shell Hartree–Fock energies of all these states employing the lcecp-*k* ECPs, determining the f occupation to $n - k$, and a 22s18p12d even-tempered reference basis set (ratio of subsequent exponents amounting to $10^{1/4} = 1.778$), which was optimized at the s^1d^1 state for f^n , at s^2d^1 for f^{n-1} , and at s^2d^2 for f^{n-2} . The Roothaan parameters²⁰ are listed in Table S1, the lowest s/p/d exponents in Table S2. The

individual energies relative to the state that is most favourable for the heavier lanthanides are given in Table S3. Their ranges, which are also listed in the fourth column of Table 2 (see below), are briefly summarized as follows. Energy differences between different occupations change monotonically from left to right in the series of lanthanides. The most preferred occupation for a given ECP and a given charge is often not the same for the lighter and the heavier lanthanides. Differences between the most favourable and the second most favourable occupation are always below 100 mE_h. This may be compared *e.g.* to the energy difference between the d^9s^2 and the (preferred) $d^{10}s^1$ state of Cu/Ag/Au, which amount to 14/141/69 mE_h. These numbers for Cu/Ag/Au are met by errors less than 0.15/1.7/3.8 mE_h by previously optimized dhf-QZVP/TZVP/SVP sets.¹² These errors served as a guideline for the development of our lanthanide basis sets.

The DSSP basis sets served as an excellent starting point for our development of double-, triple- and quadruple zeta valence bases designed to meet the above error ranges. The finally resulting contraction schemes and the total number of contracted spherical spd functions are listed in Table 1, together with that for the DSSP sets and the more recent WCYD sets⁵ (available only for lcecp-1). For all bases of a specific type, the contraction schemes are the same for the different types of ECPs, only for f^n ECP bases the second s contraction is released, and for f^{n-2} ECP bases one additional primitive d function is added to the first contraction. Optimizations of exponents and coefficients for the new bases for the lcecp-0 ECPs were done at the s^1d^1 state, those for the lcecp-1 ECPs at the s^2d^1 state, and that for the lcecp-2 ECPs for the s^2d^2 state. We started with designing and optimizing the QZV sets, from which we derived the TZV sets by reducing the three most diffuse s and d functions to two, and by changing the p contraction scheme from 41111 to 4211, followed by a re-optimization of all changed parts. Finally, SV sets were derived from TZV sets by changing the contraction of s from 411111 to 5211, and that of d from 311 to 41 followed by a re-optimization of all changed parts in the s and a full re-optimization of the d sets. The

Table 1 Contraction patterns of s-, p- and d and the polarization sets for the newly developed bases SV, TZV and QZV as well as for the bases originally presented together with the ECPs by Dolg, Stoll, Savin and Preuss (DSSP) and for the more recent ones by Weigand, Cao, Yang and Dolg (WCYD, available only for f^{n-1} ECPs). The data refer to the bases for the f^{n-1} ECPs; small modifications for the other sets were necessary to ensure error consistency, see text for details. For QZV, P/PP denote a smaller and a larger polarization set added to the valence set to get QZVP and QZVPP sets, SV(P) and SVP bases are identical as well as TZVP and TZVPP. n_{spd} denotes the number of spherical harmonic basis functions for the valence basis sets, n_{pol} that for the polarization sets

	n_{spd}	Contraction pattern	P/PP	n_{pol}
SV	26	{5211/4211/41}	(2f)[1f]/(2f)[1f]	7/7
TZV	33	{411111/4211/311}	2f/2f	14/14
QZV	45	{4111111/411111/3111}	2f/3f2g	14/39
DSSP	32	{31111/3111/311}	—	0
WCYD	46	{311111/31111/21111}	3f2g	39



contraction of p was changed from 4211 to 431 in case of lcecp-0 and lcecp-2, but kept as 4211 for lcecp-1 after exploratory molecular calculations, because – like in previous work^{6,10,12,15,16,19} – the focus was on error consistency rather than on rigid contraction schemes. With this in mind, the following minor changes seemed reasonable to us: in the SV, TZV and QZV bases for the f^{n-2} ECPs the first contracted d function contains one additional primitive function, and in the SV bases for the f^n ECPs, the second s contraction is released. For f^n and f^{n-2} , a somewhat higher contracted p set for the SV bases, {431}, turned out to be of sufficient flexibility.

In Table 2, column 3, the ranges of energy differences, ΔE between the occupations specified in column 2 for atoms and cations for the lcecp- k ECPs (column 1) are listed for the reference basis. In the subsequent columns, the statistical values (mean value and standard deviations) of differences to the reference values for the new bases as well as for the DSSP and WCYD bases are listed. Individual results are sampled in Tables S3–S8. For simplifying the discussion, here and in the following discussions we summarize the absolute mean values plus standard deviations as “typical errors”. For the QZV bases these typical errors in the energy differences between different states amount to about 0.1 mE_h, for TZV about 0.4 mE_h, and for SV about 2 mE_h. This fits very well to the corresponding errors for the coinage metals, see above. The DSSP sets, which are of similar size as TZV, yield typical errors of 0.9 mE_h, while the WCYD sets, which are of similar size as QZV, perform less well in this regard, showing typical errors of 2.3 mE_h (Table S8). The reason for this is their somewhat incomplete coverage of the 6s-space, which means that the s² state is significantly less well described than the d² state. It is noted that, for the QZV and TZV bases, the errors of the d³ occupation in the f^{n-1} ECP are substantially larger (0.5 and 1.5 mE_h) than for the other occupations, but we consider this to be tolerable given the low relevance of this state for chemistry, as it is typically more than 100 mE_h (~ 3 eV) higher than the ground state.

Polarization functions

The 6s shell is sufficiently polarized by the p and d functions that are already part of the valence basis. For instance, in the case of the QZV basis, the largest contribution to the 6s orbital stems from the second flattest s function whose exponent e.g. in case of Gd amounts to 0.053. This is just between the lowest and the second-lowest exponent of both the p and the d set. The polarization of the 5d shell requires f functions at least. The WCYD bases contain 3f2g sets, optimized by maximizing the atomic correlation energy obtained by a suited post-Hartree-Fock procedure. Concerning both the size and the optimized quantity, this is very similar to the polarization of the def2-QZVPP sets for the subsequent 5d elements. However, such large polarization sets are not urgently needed for HF/DFT treatments and further lead to imbalance for double or triple zeta valence bases. We thus carried out calculations of equilibrium structures and vibration frequencies for GdO, GdH₃, GdF₃ and GdN at levels DFT(PBE0)²¹ and MP2²² with QZV bases combined with the original and with modified polarization sets. The results are collected in Table 3, together with that for all-electron (scalar) relativistic X2C treatments with x2c-QZVPPall bases.¹⁶

We focus on the differences in distances, angles and vibration frequencies to the 3f2g set and, as above, we summarize the absolute mean values plus standard deviations as “typical errors”. Apparently, a 2f set generated from the 3f set by forming the (geometric) mean of the exponents, reproduces the data of the 3f2g set very well. Distances change by ~ 0.2 pm, angles by $\sim 0.1^\circ$ and frequencies by ~ 1 cm⁻¹ at DFT level, while changes at MP2 level are not much larger. These changes are small compared to the differences to the X2C results, amounting overall to ~ 3 pm, $\sim 1^\circ$ and ~ 20 cm⁻¹. Further, when contracting these two f functions with coefficients determined at GdN, the differences to the 3f2g set are still tolerable, in particular compared to the difference between the ECP and the X2C calculations. We decided to employ this 1f set for the

Table 2 Range of restricted open-shell Hartree–Fock energy differences ΔE between different s/d occupations for atoms ($q = 0$) and mono-cations ($q = 1$) obtained with a large reference basis set (see Table S2), as well as errors of different basis sets (QZV, TZV, SV, DSSP, WCYD) for these quantities in mE_h for the f^{n-k} ECPs (lcecp- k) specified in the first column. Listed are mean values ($k = 0$: La–Yb, $k = 1$: La–Lu, $k = 2$: Ce, Pr, Nd, Tb, Dy) as well as their standard deviations. As an example, for the neutral atom ($q = 0$) with lcecp-0, the average energy difference between $f^n s^1 d^1$ and $f^n s^2$ covers a range from 34 mE_h (for La) to 96 mE_h (for Yb) and the average error for this energy difference with the QZV basis amounts to 0.03 mE_h with a standard deviation of 0.03 mE_h. Individual values for all elements are given in Tables S4–S8

k, q	ΔE	QZV	TZV	SV	DSSP ^a	
0,0	$s^1 d^1 - s^2$	34...96	0.03 ± 0.03	0.08 ± 0.09	-0.45 ± 0.19	0.02 ± 0.23
0,0	$d^2 - s^2$	99...221	0.11 ± 0.05	0.39 ± 0.12	-0.17 ± 0.12	1.06 ± 0.67
0,1	$d^1 - s^1$	31...109	0.06 ± 0.05	0.02 ± 0.28	-0.11 ± 0.79	-0.40 ± 0.51
1,0	$s^1 d^2 - s^2 d^1$	-3...70	0.07 ± 0.03	0.25 ± 0.08	0.40 ± 0.18	0.28 ± 0.10
1,0	$d^3 - s^1 d^1$	46...194	0.36 ± 0.16	1.28 ± 0.31	0.78 ± 0.37	1.41 ± 0.45
1,1	$s^1 d^1 - s^2$	-40...41	0.04 ± 0.02	-0.30 ± 0.08	-0.53 ± 0.18	-1.37 ± 0.28
1,1	$d^2 - s^2$	-36...133	0.09 ± 0.03	-0.21 ± 0.11	-0.83 ± 0.22	-1.29 ± 0.29
2,0	$s^2 d^2 - s^1 d^3$	43...4	-0.02 ± 0.01	-0.08 ± 0.03	-1.50 ± 0.16	0.02 ± 0.01
2,0	$d^4 - s^1 d^3$	29...74	0.12 ± 0.09	0.47 ± 0.16	2.35 ± 0.41	0.24 ± 0.02
2,1	$s^2 d^1 - s^1 d^2$	90...49	0.07 ± 0.01	0.33 ± 0.02	2.91 ± 0.60	0.62 ± 0.22
2,1	$d^3 - s^1 d^2$	-26...22	-0.01 ± 0.01	-0.04 ± 0.01	1.14 ± 0.22	-0.33 ± 0.06

^a The data of Lu were not included due to non-typically large errors of up to 8 mE_h, probably arising from incorrect contraction coefficients of the s functions in ref. 1.



Table 3 Distances (d , in pm), angles (α , in $^\circ$) and vibration frequencies (f , in cm^{-1}) for selected Gd compounds obtained with lcecp-1-QZVPP sets (column 3f2g) at levels DFT(PBE0) and MP2 and differences to these values for the smaller polarization sets 2f and 1f as well as differences to all-electron scalar relativistic X2C calculations with x2c-QZVPPall basis sets (for PBE0 only). For the four-atomic compounds, f_1 and f_2 refer to the lowest frequencies, corresponding to the a_1 - and the e -type bend vibration

	DFT(PBE0)				MP2		
	3f2g	2f	1f	ECP vs. X2C	3f2g	2f	1f
GdH₃							
d	201.65	0.03	0.05	-2.06	201.10	0.20	-0.70
α	112.08	0.07	0.15	-0.15	114.19	0.46	0.00
f_1	296.12	-0.33	-1.34	9.27	295.81	-0.41	-1.14
f_2	547.51	0.58	0.63	-0.58	547.00	0.37	1.13
GdF₃							
d	203.65	0.05	0.00	-0.99	204.46	0.31	-0.10
α	117.18	0.00	0.04	-0.55	118.75	0.02	-0.56
f_1	51.15	0.74	0.57	4.16	51.17	0.74	2.43
f_2	137.78	0.68	0.42	-2.06	137.71	0.66	-0.61
GdN							
d	179.03	0.22	0.30	-4.12	185.91	0.14	0.07
f	843.40	0.40	2.06	29.98	667.53	-0.35	10.93
GdO							
d	181.16	0.22	0.31	-2.08	182.27	0.33	0.50
f	846.60	0.64	2.32	13.45	828.46	-1.90	1.23
Δd	0.13 ± 0.10	0.17 ± 0.16	-2.31 ± 1.31		-0.25 ± 0.09	-0.06 ± 0.50	
Δf	0.45 ± 0.40	0.78 ± 1.32	9.03 ± 11.81		-0.15 ± 0.99	2.33 ± 4.41	
$\Delta \alpha$	0.04 ± 0.05	0.10 ± 0.08	-0.35 ± 0.28		0.24 ± 0.31	-0.28 ± 0.40	

SV bases, to specify the large-core ECP basis sets lcecp- k -SVP and lcecp- k -SV(P) ($k = 0, 1, 2$), the 2f set for the TZV and for the QZV bases to specify lcecp- k -TZVP, lcecp- k -TZVPP and lcecp- k -QZVP. For the lcecp- k -QZVPP set, the WYCD 3f2g sets were combined with the QZV bases.

We note that one has to be careful with adding in particular steeper f functions, as this can easily lead to non-physical occupations, in particular for open-shell cases like GdO. The (Mulliken) occupation obtained with an all-electron relativistic calculation at DFT(PBE0) level and x2c-QZVPPall bases for GdO is $6s^{0.86}6p^{0.12}5d^{1.21}4f^{7.14}$. Thus, when employing the f^{n-1} ECP ($n = 8$ for Gd), one would expect an f occupation of about 0.14 electrons. With the polarizing 3f2g set of the WCYD bases one gets $f^{0.07}$, which is a bit too small, but still quite reasonable, but when adding three further steep f functions (by repeated multiplication of the steepest f exponent by 2.5) one gets $f^{0.91}$, thus an overall f^8 occupation instead of f^7 . For this, it is not possible to determine a saturated polarization set that could be used as reference for smaller polarization, so we use the 3f2g set also for the reference basis. We note in passing that, at least for GdO, this problem is specific to DFT. At HF level the f occupation amounts to 0.12, very well meeting the expectations of the all-electron treatment.

Molecular tests

Following the above considerations, we tested the lcecp- k -XVP/PP ($k = 0, 1, 2$, X = S, TZ, QZ) sets at a set of 120 Ln compounds which represent the oxidation states +II, +III and +IV (fluorides, chlorides, hydrides and oxides) at DFT (PBE0) level; a hybrid-DFT functional was chosen because of the intended

comparison with all-electron calculations, as with hybrid functionals the relative energetic position of the f shell is reproduced more realistically than with pure DFT functionals. We determined the errors of lcecp- k -QZVPP, lcecp- k -QZVP, lcecp- k -TZVP, lcecp- k -SV(P) and lcecp- k -DSSP sets with respect to an even-tempered reference basis set (see above), added by the 3f2g polarization functions of the WYCD sets (identical to those of the lcecp-QZVPP bases). Further, we calculated the differences of lcecp- k -QZVPP sets to all-electron scalar relativistic calculations (X2C) with x2c-QZVPPall bases. Individual results are listed in Tables S9–S14, mean values for the deviations in distances, (lowest) vibration frequencies and angles in Table 4 for $k = 0–2$. It is of important note that the X2C calculations sometimes lead to f occupations between two integer numbers which makes it difficult to compare them with lcecp calculations, where the f occupation is fixed to an integer by definition. The compounds affected by this in a way that the X2C f occupation differs from f^{n-2} by more than 0.5 electrons, PrH₄, NdCl₄, NdH₄, TbH₄, DyCl₄ and DyH₄, were thus not included in the statistics shown in Table 4.

To start with the most relevant case, f^{n-1} , the typical differences (absolute of the mean value plus standard deviation) of the ECP technique to the probably more reliable all-electron X2C method are 3.5 pm for distances, 40 cm^{-1} for frequencies and 2° for angles. The basis set errors within the ECP technique are smaller throughout: 0.7 pm/9 cm^{-1} /1.2° for SV(P), 0.4 pm/6 cm^{-1} /0.3° for TZVP, 0.3 pm/1.5 cm^{-1} /0.2° for QZVP and 0.06 pm/1.4 cm^{-1} /0.1° for QZVPP. The DSSP sets are slightly worse than the SV(P) sets, 2.6 pm/12 cm^{-1} /1.3°, but also here the basis set errors are smaller than the differences of the ECP to X2C themselves. Similar holds for f^n , with some



Table 4 Mean values and standard deviations of errors in distances (in pm), angles (in degree) and lowest vibration frequencies (in cm^{-1}) for lcecp- k -bases with respect to the reference basis, both used in combination with the corresponding lcecp- k effective core potentials at level DFT(PBE0). The last column contains the difference between lcecp- k -QZVPP and scalar relativistic all-electron X2C treatment with x2c-QZVPPall bases. Individual data are listed in Tables S9–S14

k	QZVPP	QZVP	TZVP	SV(P)	DSSP	ECP vs. X2C
0	Δd 0.001 \pm 0.067	0.030 \pm 0.103	-0.234 \pm 0.234	0.691 \pm 1.308	0.997 \pm 0.921	-6.228 \pm 2.339
	Δf -0.172 \pm 0.804	-0.061 \pm 0.931	1.491 \pm 2.280	-2.061 \pm 8.168	0.268 \pm 5.317	21.13 \pm 25.97
	$\Delta \alpha$ 0.152 \pm 0.092	0.193 \pm 0.103	-0.143 \pm 0.234	1.996 \pm 4.416	-0.801 \pm 0.617	-7.008 \pm 3.211
1	Δd 0.031 \pm 0.030	0.123 \pm 0.179	-0.093 \pm 0.298	0.242 \pm 0.503	1.410 \pm 1.211	-2.225 \pm 1.298
	Δf -0.380 \pm 1.142	-0.250 \pm 1.272	1.061 \pm 5.123	0.704 \pm 8.567	0.383 \pm 11.34	11.99 \pm 18.25
	$\Delta \alpha$ 0.056 \pm 0.053	0.121 \pm 0.112	0.106 \pm 0.262	0.322 \pm 0.850	-0.815 \pm 0.548	-0.064 \pm 2.144
2	Δd -0.083 \pm 0.103	0.004 \pm 0.135	-0.154 \pm 0.182	-0.551 \pm 0.397	0.935 \pm 0.803	-0.903 \pm 0.666
	Δf 0.197 \pm 0.541	0.986 \pm 1.090	0.966 \pm 0.933	-3.777 \pm 3.870	-1.159 \pm 2.895	-30.13 \pm 12.00

exceptions: the differences to X2C for bonds and angles are larger, 9 pm/47 $\text{cm}^{-1}/10^\circ$, and the quality of DSSP is between SV(P) and TZVP; nevertheless, already for SV(P) the basis set errors are smaller than the differences of ECP to X2C, 2 pm/10 $\text{cm}^{-1}/6^\circ$. For f^{n-2} , results for basis set errors are in the same range; this does also hold for the agreement with X2C, when excluding compounds with X2C f occupations that deviate by more than 0.5 electrons from f^{n-2} . For clarity we note that the X2C data not necessarily are always more reliable than the ECP data, in particular at DFT level, due to sometimes non-physical over-filling of the f shell, as reported *e.g.* for trivalent Eu compounds with soft ligands.²³

For assessing error consistency with the subsequent d element basis sets we give the errors for dihydrides, difluorides, dichlorides and monooxides of Hf (individual data in Table S15). Taking def2-QZVPP as reference, the typical errors for def2-SV(P) amount to 3.4 pm/19 $\text{cm}^{-1}/9.8^\circ$, for def2-TZVP to 0.4 pm/4.5 $\text{cm}^{-1}/1.2^\circ$, and for def2-QZVP amount to 0.08 pm/1.0 $\text{cm}^{-1}/0.7^\circ$. This is quite similar as for the lanthanides in case of TZVP and QZVP, but somewhat larger for Hf than for the lanthanides in case of the SV(P) bases, partly due to the fact, that the SV(P) basis for Ln contains a polarizing f function, while this is not the case for the SV(P) basis for the 6d elements (see also below).

Finally, we assess the errors in energies of an exchange reaction of the type $\text{Ln}^A\text{H}_3 + \text{Ln}^B\text{Cl}_3 \rightarrow \text{Ln}^A\text{Cl}_3 + \text{Ln}^B\text{H}_3$ ($\text{Ln}^A, \text{Ln}^B = \text{La-Lu}$) within ECP-based treatments with lcecp-1-QZVP/PP, lcecp-1-TZVP, lcecp-1-SV(P) and DSSP bases with respect to the reference basis sets as well as the differences of ECP-based treatments with lcecp-1-QZVPP bases and X2C treatments with x2c-QZVPPall bases. The energies for these reactions range from very few kJ mol^{-1} for Ln^A and Ln^B being neighboured to up to $\sim 30 \text{ kJ mol}^{-1}$ for $\text{Ln}^A = \text{La}$ and $\text{Ln}^B = \text{Lu}$.

Table 5 Mean values and standard deviations of errors in DFT (PBE0) energies of reactions of the type $\text{Ln}^A\text{H}_3 + \text{Ln}^B\text{Cl}_3 \rightarrow \text{Ln}^A\text{Cl}_3 + \text{Ln}^B\text{H}_3$ ($\text{Ln}^A, \text{Ln}^B = \text{La-Lu}$) with respect to reference basis sets for lcecp-1 ECPs and for X2C with respect to lcecp-QZVPP basis sets in kJ mol^{-1} . Individual data are listed in Tables S16–S22

QZVPP	QZVP	TZVP	SV(P)	DSSP	ECP vs. X2C
-0.08 \pm 0.06	0.49 \pm 0.39	1.12 \pm 0.99	3.01 \pm 2.69	5.21 \pm 3.58	-1.87 \pm 9.07

The individual numbers are listed in Tables S16–S22, mean values and standard deviations are summarized in Table 5. Again, all basis set typical errors are smaller than the typical differences between ECP and X2C techniques (11 kJ mol^{-1}). For QZVP/PP sets, they are below 1 kJ mol^{-1} , for TZVP around 2 kJ mol^{-1} , for SV(P) 5 kJ mol^{-1} and for the DSSP sets 9 kJ mol^{-1} . Nevertheless, in view of the comparably small reaction energies of only up to 30 kJ mol^{-1} , the differences between ECP and X2C techniques as well as the errors of DSSP are quite large. The comparably large differences between X2C and ECP treatments are not too surprising as the ECPs average over all configurations of a given f occupation, whereas the all-electron calculations describe the ground state configuration. A good agreement in energy differences thus results only in case of a good transferability of the open shell core, *i.e.* if the specific configurations of the f shell of the Ln atoms are similar for educts and products. Geometries are less critical with respect to this approximation.

We note in passing that exploratory calculations employing an SV(P) set with removed f function yield errors in the range of DSSP. We thus decided to keep the f function also for the SV(P) set, although this is not consistent with the subsequent d elements, where such a function is present only in the SVP but not in the SV(P) set. We further note that the lcecp-1-SV(P) and lcecp-1-TZVP sets were already successfully employed for the structure clarification of cluster anions of the type $\text{Ln}_n\text{X}_{3n+1}^-$ ($\text{Ln} = \text{La-Lu}$; $\text{X} = \text{Cl, Br}$; $n = 1-6$)^{24,25} with a DFT-based genetic algorithm procedure complementing ion mobility studies.

Conclusions

For all lanthanides, we designed, optimized and tested polarized segmented contracted Gaussian basis sets of double, triple, and quadruple zeta valence quality for the use together with the f-in-core effective core potentials provided by Dolg *et al.* In this way we extended the system of error-consistent basis sets (“def2”) to these elements. For each element, we developed such a series for each of the ECPs modelling a specific f occupation: f^n , the (usual) ground state occupation, f^{n-1} , which is present in most Ln(m) compounds, and f^{n-2} , for highly oxidized states in case of Ce, Pr, Nd, Tb and Dy. They are termed lcecp- k -XVP ($\text{X} = \text{S, TZ, QZ}$, $k = 0, 1, 2$). Their quality was assessed through DFT(PBE0) for a set of 120 molecules,

covering mainly fluorides, chlorides, hydrides and oxides in oxidation states +II, +III and +IV. Further, for the same set, the differences between the ECP technique and the scalar relativistic all-electron X2C technique were compared for QZVPP bases, *i.e.* very close to the basis set limit. For the most relevant case, f^{n-1} , the typical differences (absolute of the mean value plus standard deviation) between the two techniques are 3.5 pm for distances, 40 cm⁻¹ for frequencies and 2° for angles. The basis set errors within the ECP technique are smaller throughout: 0.7 pm/9 cm⁻¹/1.2° for SV(P), 0.4 pm/6 cm⁻¹/0.3° for TZVP, 0.3 pm/1.5 cm⁻¹/0.2° for QZVP and 0.06 pm/1.4 cm⁻¹/0.1° for QZVPP. The basis sets originally developed by Dolg *et al.* are roughly in the range of SV(P) sets, 2.6 pm/12 cm⁻¹/1.3°; also for them, the basis set errors are smaller than the errors of the ECP method itself. For f^{n-2} , comparisons in most cases are not easily feasible, as occupations close to f^{n-2} in all-electron calculations are achieved mainly for LnF₄ compounds. Generally, for QZVP and TZVP bases, the errors are similar to the subsequent d elements (tested for HfH₂, HfF₂, HfCl₂, HfO); for SV(P) bases, in contrast, the errors for the lanthanides are somewhat smaller than for the d elements. The newly optimized basis sets will be helpful in particular for calculations of properties directly related to the energy surface, *e.g.* geometry optimizations or IR spectra.

Author contributions

F. W.: conceptualization, design and optimization of basis sets, writing. M. L.: molecular calculations and their evaluation, data curation, review and editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

The lcecp-bases presented in this work are stored in the file In-lcecp-bases.txt together with the corresponding effective core potentials [M. Dolg, H. Stoll, A. Savin, H. Preuss, Energy-Adjusted Pseudopotentials for the Rare-Earth Elements, *Theor. Chim. Acta*, 1989, **75**, 173–194 and M. Dolg, H. Stoll, H. Preuss, A combination of quasirelativistic pseudopotential and ligand field calculations for lanthanoid compounds, *Theor. Chim. Acta*, 1993, **85**, 441–450.] and are additionally available at <https://basissets.turbomole.org/> and to <https://www.basissetexchange.org/>. All other data (Roothaan parameters, reference basis sets, atomic and molecular energies, distances, bond angles and frequencies) are available within the supplementary information (SI). All calculations were done with TURBOMOLE [TURBOMOLE V7.8 2023, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <https://www.turbomole.org>]. From this website, the program suite can be obtained for educational and reviewing tasks free of charge.

Supplementary information contains computational details, specification of the reference basis sets, and individual results used for the statistics in the main text. See DOI: <https://doi.org/10.1039/d5cp04944j>.

Acknowledgements

We gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Collaborative Research Centre 1573 “4f-for-future” in project Q.

Notes and references

- 1 M. Dolg, H. Stoll, A. Savin and H. Preuss, Energy-Adjusted Pseudopotentials for the Rare-Earth Elements, *Theor. Chim. Acta*, 1989, **75**, 173–194.
- 2 M. Dolg, H. Stoll and H. Preuss, A combination of quasi-relativistic pseudopotential and ligand field calculations for lanthanoid compounds, *Theor. Chim. Acta*, 1993, **85**, 441–450.
- 3 M. Huelsen, A. Weigand and M. Dolg, Quasirelativistic energy-consistent 4f-in-core pseudopotentials for tetravalent lanthanide elements, *Theor. Chem. Acc.*, 2009, **122**, 23–29.
- 4 J. Yang and M. Dolg, Valence basis sets for lanthanide 4f-in-core pseudopotentials adapted for crystal orbital ab initio calculations, *Theor. Chem. Acc.*, 2005, **113**, 212–224.
- 5 A. Weigand, X. Cao, J. Yang and M. Dolg, Quasirelativistic f-in-core pseudopotentials and core-polarization potentials for trivalent actinides and lanthanides: molecular test for trifluorides, *Theor. Chem. Acc.*, 2010, **126**, 117–127.
- 6 F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 7 D. Andrae, U. Häussermann, M. Dolg, H. Stoll and H. Preuss, Energy-adjusted ab initio pseudopotentials for the 2nd and 3rd row transition-elements, *Theor. Chim. Acta*, 1990, **77**, 123–141.
- 8 B. Metz, H. Stoll and M. Dolg, Small-core multiconfiguration-Dirac–Hartree–Fock-adjusted pseudopotentials for post-*d* main group elements: Application to PbH and PbO, *J. Chem. Phys.*, 2000, **113**, 2563–2569.
- 9 M. Dolg, H. Stoll and H. Preuss, Energy-Adjusted Ab initio Pseudopotentials for the Rare Earth Elements, *J. Chem. Phys.*, 1989, **90**, 1730–1734.
- 10 R. Gulde, P. Pollak and F. Weigend, Error-Balanced Segmented Contracted Basis Sets of Double- ζ to Quadruple- ζ Valence Quality for the Lanthanides, *J. Chem. Theory Comput.*, 2012, **8**, 4062–4068.
- 11 D. Figgen, K. A. Peterson, M. Dolg and H. Stoll, Energy-consistent pseudopotentials and correlation consistent basis sets for the elements Hf–Pt, *J. Chem. Phys.*, 2009, **130**, 164108.



12 F. Weigend and A. Baldes, Segmented contracted basis sets for one- and two-component Dirac–Fock effective core potentials, *J. Chem. Phys.*, 2010, **133**, 174102.

13 W. Kutzelnigg and W. Liu, Quasirelativistic theory equivalent to fully relativistic theory, *J. Chem. Phys.*, 2005, **123**, 2421102.

14 Y. J. Franzke, N. Middendorf and F. Weigend, Efficient implementation of one- and two-component analytical energy gradients in exact two-component theory, *J. Chem. Phys.*, 2018, **148**, 104110.

15 P. Pollak and F. Weigend, Segmented Contracted Error-Consistent Basis Sets of Double- and Triple- ζ Valence Quality for One- and Two-Component Relativistic All-Electron Calculations, *J. Chem. Theory Comput.*, 2017, **13**, 3696–3705.

16 Y. J. Franzke, L. Spiske, P. Pollak and F. Weigend, Segmented Contracted Error-Consistent Basis Sets of Quadruple- ζ Valence Quality for One- and Two-Component Relativistic All-Electron Calculations, *J. Chem. Theory Comput.*, 2020, **16**, 5658–5674.

17 TURBOMOLE V7.8 2023, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from <https://www.turbomole.com>.

18 O. Treutler and R. Ahlrichs, Efficient molecular numerical integration schemes, *J. Chem. Phys.*, 1995, **102**, 346–354.

19 Y. J. Franzke, R. Tress, T. Pazdera and F. Weigend, Error-consistent segmented contracted all-electron relativistic basis sets of double- and triple-zeta quality for NMR shielding constants, *Phys. Chem. Chem. Phys.*, 2019, **30**, 16658–16664.

20 C. C. J. Roothaan, Self-Consistent-Field Theory for Open Shells of Electronic Systems, *Rev. Mod. Phys.*, 1960, **32**, 179–185.

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

22 F. Haase and R. Ahlrichs, Semidirect mp2 gradient evaluation on workstation computers – the mpgrad program, *J. Comput. Chem.*, 1993, **14**, 907–912.

23 M. Dolg, X. Cao and J. Ciupka, Misleading evidence for covalent bonding from Eu^{II}X and Am^{II}X density functional theory bond lengths, *J. Electron Spectrosc. Relat. Phenom.*, 2014, **194**, 8–13.

24 Y. Nakajima, P. Weis, F. Weigend, M. Lukanowski, F. Misaizu and M. M. Kappes, Lanthanide chloride clusters, $\text{Ln}_x\text{Cl}_{3x+1}^-$, $x = 1–6$: an ion mobility and DFT study of isomeric structures and interconversion timescales, *Phys. Chem. Chem. Phys.*, 2025, **27**, 1017–1030.

25 Y. Nakajima, P. Weis, F. Weigend, M. Lukanowski, F. Misaizu and M. M. Kappes, A cyclic ion mobility and DFT study of the structures, isomer space and isomer interconversion of lanthanide bromide clusters, $\text{Ln}_x\text{Br}_{3x+1}^-$, $x = 1–6$, *Phys. Chem. Chem. Phys.*, 2026, **28**, 596–612.

