EDGE ARTICLE

Check for updates

Greater transferability and accuracy of norm-conserving pseudopotentials using nonlinear core corrections†

Wan-Lu Li, [abcf] Kaixuan Chen, [abc] Elliot Rossomme, [ab] Martin Head-Gordon [abc] and Teresa Head-Gordon [*abcde]

We present an investigation into the transferability of pseudopotentials (PPs) with a nonlinear core correction (NLCC) using the Goedecker, Teter, and Hutter (GTH) protocol across a range of pure GGA, meta-GGA and hybrid functionals, and their impact on thermochemical and non-thermochemical properties. The GTH-NLCC PP for the PBE density functional demonstrates remarkable transferability to the PBE0 and ωB97X-V exchange–correlation functionals, and relative to no NLCC, improves agreement significantly for thermochemical benchmarks compared to all-electron calculations. On the other hand, the B97M-rV meta-GGA functional performs poorly with the PBE-derived GTH-NLCC PP, which is mitigated by reoptimizing the NLCC parameters for this specific functional. The findings reveal that atomization energies exhibit the greatest improvements from use of the NLCC, which thus provides an important correction needed for covalent interactions relevant to applications involving chemical reactivity. Finally we test the NLCC-GTH PPs when combined with medium-size TZV2P molecularly optimized (MOLOPT) basis sets which are typically utilized in condensed phase simulations, and show that they lead to consistently good results when compared to all-electron calculations for atomization energies, ionization potentials, barrier heights, and non-covalent interactions, but lead to somewhat larger errors for electron affinities.

Introduction

Pseudopotentials are mathematical representations of the interactions between explicitly treated valence electrons and effective cores that replace atomic nuclei and inner shell electrons, allowing for a substantial reduction in computational resources required to model large chemical and materials systems, while also accounting for relativistic effects in some cases.[1] The Goedecker, Teter, and Hutter (GTH) formalism and optimization protocol has produced separable, norm-

a Kenneth S. Pitzer Center for Theoretical Chemistry, USA

b Department of Chemistry, USA

c Chemical Sciences Division, Lawrence Berkeley National Laboratory, USA. E-mail: thg@berkeley.edu

d Department of Chemical and Biomolecular Engineering, USA

e Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720, USA

f Department of Nanoengineering and Materials Science and Engineering Program, University of California San Diego, La Jolla, California 92093, USA

† Electronic supplementary information (ESI) available: Electronic configurations of core and valence regions for pseudopotentials of elements studied; comparison of mean absolute deviations for a specific atom type within the G2 dataset; distributions of MADs and all-electron calculations for IP and EA datasets; NLCC performance on non-thermochemical properties. See DOI: https://doi.org/10.1039/d3sc03709f

conserving, Gaussian-type dual-space pseudopotentials (PPs)[2,3] that are widely used in *ab initio* molecular dynamics (AIMD) simulations with a mixed Gaussian-planewave (GPW) strategy.[4] The GTH PP construction and optimization is generally performed utilizing the neutral atomic state as a reference with spherical symmetry. To make an effective PP, it is essential to divide the space into two regions: muffin-tin spheres centered on the atom in a molecule or solid, and the interstitial region that encompasses the rest of the charge density.[5] Knowing the scattering properties on the surface of the muffin-tin spheres then allows for exact solutions of the Schrödinger equation in the interstitial region.[6] By satisfying the norm conservation condition in GTH PPs,[7] the logarithmic derivative function associated with the energy accurately represents the scattering properties of a muffin-tin sphere that contains the charge distribution of the reference configuration. Screening effects give rise to an approximately invariant muffin-tin sphere, wherein the total electronic charge distribution remains largely unaffected by changes in the outer chemical environment.[8] The construction of the traditional GTH[2] and Hartwigsen–Goedecker–Hutter (HGH)[3] PPs followed this approach, resulting in PPs that exhibit excellent transferability for non-spin-polarized systems.

However, in a self-consistent field calculation, the charge distribution undergoes changes when a free atom is inserted

into a molecule or solid, and the potential within the muffin-tin region will generally differ from the potential within a muffin-tin sphere of the same radius around the reference atom. As a result, the scattering properties undergo modifications, and the PP constructed based on the isolated atom might not accurately reproduce the altered scattering properties due to the new chemical environment. In spin-polarized calculations, the concept of an invariant muffin-tin sphere is also not applicable since the spin polarization deviates across different chemical environments, despite the similarity in the total charge within the invariant muffin-tin sphere. In other words, the typical assumption for PPs that assume a linear relationship between the charge distribution and potential leads to inaccuracies because the electron–electron interactions in the core region are highly localized such that they exhibit a nonlinear relationship between the electron density and the associated potential.

One approach to address this is by incorporating nonlinear core corrections (NLCC)[9] into the PP optimization. For the NLCC schemes, the spin and charge densities within the muffin-tin sphere are not solely determined by the valence electrons, but instead encompass the combined contributions of the valence charge and the core charge. The NLCC method allows for a more realistic treatment of the core electron behavior by taking into account the strong electron–electron interactions and the asymmetric charge density distribution near the atomic nucleus. In accordance with the principles underlying the GTH PPs, Willand *et al.* in 2013 put forward a NLCC method in which the core charge density is represented by a single Gaussian function,[10] which captures the essential characteristics of the core charge. The amplitude and width of this Gaussian core charge distribution were subsequently optimized through a rigorous fitting procedure similar to the regular GTH PP optimization[11–14] but including not just ground state but also excited states and ionized electronic configurations. By employing this methodology, the NLCC-PP achieves a high level of precision in describing the atomization energies in molecular systems compared to all-electron calculations and reliability for high-pressure phases of crystalline solids. Together with an adequate treatment of dispersion, such as the empirical Grimme D2 (ref. 15) and D3 (ref. 16) schemes, the inclusion of NLCC corrections allows for a comprehensive description of weakly bound intermolecular interactions, achieving an average error of approximately 0.5 kcal mol^{-1}.

Recently, we presented a study on the optimization and transferability of small and large core GTH PPs for Density Functional Theory (DFT) functionals applied to molecular systems and condensed phase simulations for electrocatalysis.[13] We also have recently conducted a systematic exploration of the extent of PP inconsistency errors (PPIEs)[17] that arise from the transfer of a different PP relative to the chosen level of DFT when evaluating energy differences. To address these errors, we have developed and implemented empirical atom- and density functional approximation specific PP corrections.[17] But in neither case have we considered the role of NLCC PPs and their influence on the accuracy possible with different DFT functionals outside the original parameterization of the PBE functional[10] and when combined with less complete basis sets such

as the molecularly optimized (MOLOPT) basis sets used in AIMD application studies in CP2K.[18]

In this work, we directly use the 2013 NLCC HGH-PP developed for the PBE functional,[10] and investigate its transferability and accuracy across various DFT functionals including the hybrid GGA PBE0,[19] the semi-empirical hybrid ωB97M-rV,[20] and the meta-GGA B97M-rV,[21,22] as measured on performance for thermochemical properties, barrier heights, isomerization energies, and non-covalent interactions. In addition, we also combine the GTH-NLCC method with the medium-sized TZVP MOLOPT basis sets that are used to enhance the computational efficiency in CP2K[23] and compare them to the more complete def2-TZVPPD basis set results. Although the NLCC GTH-PP results are more transferable and yield excellent results for the hybrid functionals regardless of basis set, the B97M-rV functional exhibits a large PPIE. Hence we undertook a reoptimization process to enhance the accuracy of the NLCC parameters for this meta-GGA functional which brings it into line with the other DFT/basis set combinations. As we show below, the use of the NLCC correction improves overall accuracy over the full range of DFT functionals and basis sets considered here, minimizing errors across all data sets but especially for thermochemical properties that are most relevant for applications involving chemical reactivity.

Computational details

GTH PPs and optimization algorithm of NLCC parameters

The norm-conserving, dual-space, and explicitly separable GTH pseudopotential V^{pp} contains local and nonlocal parts,[2,24] where the local term is expressed as

$$V_{loc}^{pp}(r) = -\frac{Z_{ion}}{r} erf\left(\frac{r}{\sqrt{2}r_{loc}^{pp}}\right) + \sum_i^4 C_i^{pp}\left(\frac{r}{r_{loc}^{pp}}\right)^{2i-2} \exp\left[-\frac{1}{2}\left(\frac{r}{r_{loc}^{pp}}\right)^2\right] \quad (1)$$

and erf represents the error function, Z_{ion} is the ionic charge of the atom, r_{loc}^{pp} controls the range of the Gaussian ionic charge distribution, and the C_i^{pp} are the coefficients to be optimized.

The nonlocal part is expressed as

$$V_{nl}^{pp}(r, r') = \sum_{lm}\sum_{ij} \langle r|p_i^{lm}\rangle h_{ij}^l \langle p_j^{lm}|r'\rangle \quad (2)$$

where h_{ij}^l denotes a scattering matrix element and $\langle r|p_i^{lm}\rangle$ is Gaussian-type projector given as

$$\langle r|p_i^{lm}\rangle = N_i^l Y^{lm}(r) r^{l+2i-2} \exp\left[-\frac{1}{2}\left(\frac{r}{r_l}\right)^2\right] \quad (3)$$

where l denotes the angular momentum quantum number of subshells, N_i^l denotes a normalization constant and Y^{lm} is a Laplace spherical harmonic.[2] Therefore, parameters r_{loc} and C_i in the local term, and h_{ij} and r_l in the nonlocal part need to be fitted in the GTH PP optimization.

The nonlinear core correction is given by

$$\rho_c(r) = c_{\text{core}} \frac{Z - Z_{\text{ion}}}{\left(\sqrt{2\pi}\, r_{\text{core}}\right)^3} e^{-\frac{r^2}{2r_{\text{core}}^2}} \quad (4)$$

in which the core charge is described by a single Gaussian function, which is efficient for numerical integration, and is then incorporated into the generalized Kohn–Sham total energy equation:

$$E_{\text{KS}} = \sum_i \left\langle \psi_i \left| \left\{ -\frac{1}{2}\nabla^2 + V_{\text{H}}[\rho] + V_{\text{xc}}[\rho + \rho_c] + V^{\text{pp}} \right\} \right| \psi_i \right\rangle$$
$$- E_{\text{H}}[\rho] + E_{\text{xc}}[\rho + \rho_c] - \int d r \rho(r) V_{\text{xc}}[\rho + \rho_c](r) \quad (5)$$

where E_{xc}, V_{xc} are the exchange–correlation energy and potential, respectively, and V_{H} is the Hartree potential. The amplitude (c_{core}) and width (r_{core}) of the function are optimized based on the usual procedure of modulating initial step size, weight for configurations in different regions such as core, valence and virtual states, and target accuracy for Kohn–Sham eigenvalues, with reference to the all-electron results with the consideration of scalar relativistic effects.[3] However, contrary to the optimization procedure of a traditional GTH PP, which is fitted to a single atomic configuration considered as a symmetric sphere, the optimized NLCC PPs are parameterized not only with the ground state, but also with excited states and ionized electronic configurations. Several atomic properties were chosen in the fitting procedure with respect to the all-electron (AE) calculations: (1) atomic eigenvalues of the occupied and first few unoccupied valence orbitals, (2) charge within the inert region (r_{loc}) of the pseudo atom matching the charge in the same region for all the orbitals; this criterion illustrates that the PP is norm conserving for all the configurations used in the optimization, (3) total energy difference, and (4) spin polarization energy of all-electron calculation is reproduced. This is incorporated into the objective penalty function expressed in eqn (7)

$$S = \sum_{n,l} w_{n,l}^2 \left(p_{n,l}^{\text{AE}} - p_{n,l}^{\text{PP}} \right) \quad (6)$$

where p denotes different atomic properties described above, w is the weight of each property, and n and l are principal and angular momentum quantum numbers, respectively.

Meta-GGA functionals stand out for their ability to enhance the precision of GGAs by incorporating local kinetic energy density considerations. This distinctive feature empowers meta-GGAs to provide a more refined treatment of diverse chemical bonds, including covalent, metallic, and weak bonds, in contrast to the capabilities of LDAs and GGAs.[25,26]

$$E_{\text{xc}}^{\text{mGGA}}[\rho, \tau] = d^3 r f[\rho(r), \nabla\rho(r), \tau(r)] \quad (7)$$

where the kinetic energy density is expressed as the summation over all occupied Kohn–Sham orbitals:

$$\tau(r) = \frac{1}{2} \sum_i^{\text{occ}} [\nabla\varphi_i(r)]^2 \quad (8)$$

However, in the CP2K code,[18] an assumption is made that the kinetic energy density stemming from core orbitals is excluded from consideration, with only valence orbitals taken into account. Consequently, during meta-GGA calculations, the employed variables encompass the core charge augmented density and density gradient, accompanied by the kinetic energy density derived solely from valence orbitals. To address this limitation, we reoptimized the NLCC parameters for the meta-GGA B97M-rV functional.

Computational and theoretical details. The ATOM module implemented in CP2K package[18] was used to optimize the new NLCC GTH PPs at the level of B97M-rV. The NLCC terms optimized on top of Hartwigsen–Goedecker–Hutter (HGH) pseudopotential served as an initial guess.[10] In the calculations for different applications of the new PPs, the energy cutoff was set as 1200 Ry with a box size of 20 Å × 20 Å × 20 Å. During the Self-Consistent-Field (SCF) calculation, a strategy of orbital transform (OT)[27,28] is utilized to accelerate the convergence, of which the threshold is set to be 10^{-6} hartree. The reference results are obtained from all-electron calculations with Gaussian-type def2-TZVPPD basis sets.[29–31]

In order to examine the NLCC correction for different chemical properties, we chose a number of benchmark data sets. We follow the convention outlined by Mardirossian and Head-Gordon[32] to organize datasets into categories: (1) thermochemistry (TC), (2) barrier heights (BH), (3) isomerization energies (I), and (4) non-covalent interactions (NC), ranging from most difficult to easier in regards demands on the PP accuracy. Note however we are not using their MGDB84 dataset here. For the TC category we used the whole series of G2 datasets[33–35] that evaluates the covalent bond formation energy, as well as the TAE140nonMR and TAE140MR[36] for easy and difficult atomization energies, respectively. For the BH category we used HTBH38,[37] NHTBH38,[38] and WCPT27 (ref. 39) datasets representing hydrogen transfer barrier heights, non-hydrogen transfer barrier heights, and barrier heights of water-catalysed proton–transfer reactions, respectively. For the I category we used ID (isomerization energies "difficult") and ISOMERIZATION20 (ref. 36) for isomerization energies; G21EA[40,41] and G21IP[40,41] for adiabatic electron affinities and ionization potentials of atoms and small molecules, respectively. For the NC category we used the NCED (non-covalent "easy" dimers) and S66 (ref. 42 and 43) describing binding energies between organic molecules and biomolecules, and a NCD (non-covalent "difficult") CT20 (ref. 44) dataset for binding energies of charge-transfer complexes.

Results

The current standard in condensed phase simulations for codes such as CP2K is to use GTH PPs that are combined with compact MOLOPT basis sets, such as Table 1 and Fig. 1(a)–(c) shows the resulting errors made on atomization energies with GTH-PP/MOLOPT relative to def2-TZVPPD all-electron calculations, which is the most representative combination of PPs and basis sets used in AIMD applications where chemical transformations are critical. The standard PBE PP and its transferability to the two hybrid functionals for the GTH/MOLOPT combination is overall superior to the larger basis set for

Table 1 Errors in atomization energies predicted by GTH PPs compared to def2-TZVPPD all-electron calculations with or without NLCC corrections and using MOLOPT or def2-TZVPPD basis sets. Mean Absolute Deviations (MADS, in kcal mol^{-1}) of the different DFT functionals and basis set combinations analyzed over the G2 series[33–35] and the TAE140 datasets[36] divided into 124 "easy" molecules without multi-configurational character and 16 multi-reference atomization energies

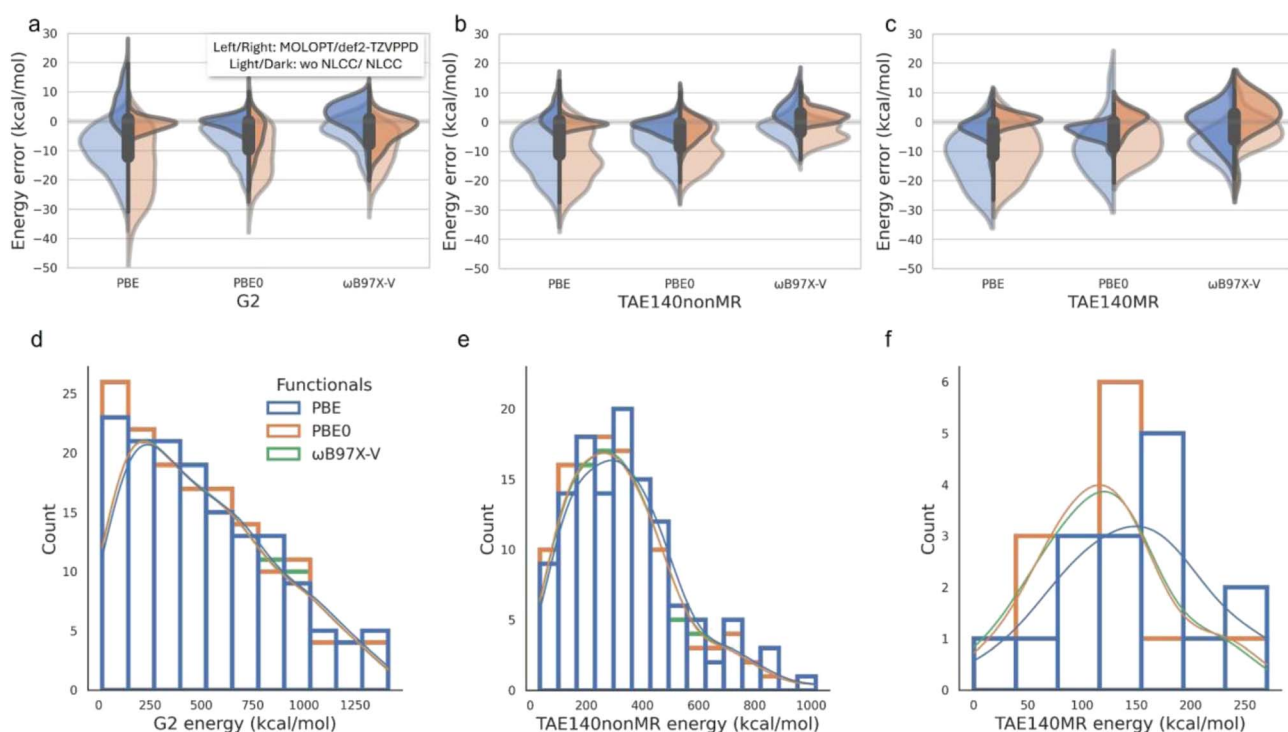| PP		GTH	GTH	GTH-NLCC	GTH-NLCC
Basis set	Dataset	MOLOPT	def2-TZVPPD	MOLOPT	def2-TZVPPD
PBE	G2	10.13	15.55	6.06	1.90
	TAE140nonMR	10.20	12.25	3.36	1.32
	TAE140MR	12.61	9.92	2.61	1.70
PBE0	G2	6.68	11.05	1.80	5.68
	TAE140nonMR	7.69	8.81	2.34	3.53
	TAE140MR	10.65	6.96	3.11	2.76
ωB97X-V	G2	4.98	9.66	4.03	3.99
	TAE140nonMR	4.05	4.16	4.48	2.60
	TAE140MR	8.25	4.35	4.26	4.82

Fig. 1 Nonlinear core corrections and basis set effects on atomization energies for the G2 and TAE140 datasets. We decomposed the total TAE140 dataset into 124 TAE140nonMR and 16 TAE140MR subsets. All the GTH pseudopotentials applied here are at PBE level. (a–c) Violin diagrams of the energy error compared with all-electron def2-TZVPPD results for density functionals of PBE, PBE0 and ωB97X-V applied to G2, TAE140nonMR and TAE140MR data set, respectively. (d–f) Distribution of the all-electron results predicted for different functionals for G2, TAE140nonMR and TAE140MR datasets, respectively.

molecules without multi-reference character, but mostly mean absolute deviations (MADs) are surprisingly large, *i.e.* 5–10% in the mean absolute percentage error (MAPE) normalized by the corresponding average value of each dataset, assessed by the energy scales shown Fig. 1(d)–(f).

To assess the errors incurred for the GTH-NLCC parameters optimized for PBE,[10] Table 1 and Fig. 1(a)–(c) shows a significant reduction in the MADs and MAPE regardless of basis set relative to the all-electron calculations. This correction effectively

addresses the issue of underestimating the covalent interactions arising from the original GTH/MOLOPT parameterization. For the PBE functional, the GTH-NLCC combined with more complete basis set def2-TZVPPD yields the best results, with the MAD generally less than 2 kcal mol^{-1} and MAPE less than 1.2%. Such consistency with the all-electron calculation has been corroborated in the study conducted by Willand *et al.*[10] where they use the standard G2-1 dataset containing only 54 molecules. However the results using GTH-NLCC/MOLOPT, the basis

set relied upon in condensed phase AIMD simulation work, is an impressive reduction in overall error as well.

Often PPs developed for one DFT functional, typically PBE, are transferred for use with other DFT functionals. Table 1 and Fig. 1(a)–(c) show that the underestimation error incurred by PBE can be alleviated by utilizing higher hierarchical density functionals, such as hybrid PBE0 and ωB97X-V functionals that incorporate a fraction of non-local exact exchange typically derived from Hartree–Fock theory, better describing both localized electron–electron interactions such as covalent bonding and delocalized interactions such as dispersion forces and van der Waals interactions.[45–48] And yet the hybrid functionals also benefit from the NLCC PPs transferred from PBE. Notably, the combination of GTH-NLCC and MOLOPT basis sets exhibits a remarkable capacity to faithfully reproduce all-electron outcomes, when compared to the larger basis sets used in electronic structure codes. This underscores the extensive utility of MOLOPT basis in practical computations, as it adeptly strikes a fine balance between precision and computational efficiency.

Table 2 and Fig. 2 depict the errors in atomization energies predicted by various combinations of PPs and basis sets when evaluated across the G2, TAE140nonMR, and TAE140MR datasets for the semi-empirical meta-GGA B97M-rV functional.

Although the GTH-NLCC PP for PBE was highly transferable to the two hybrid functionals, one of which is also a semi-empirical functional developed by the same data-driven procedure,[49] we find the transferability is quite poor for B97M-rV meta-GGA functional,[21,22] yielding large MAD errors (>rbin 50 kcal mol^{-1}) regardless of the completeness of the basis sets when compared to all-electron calculations (Table 2). Notably two clear trends emerge, demonstrating first that the original GTH/MOLOPT approach significantly underestimates the atomically covalent interaction, whereas the existing NLCC correction (GTH-NLCC-2013) exhibits an adverse and now substantial overestimation of the interaction. By comparing the MADs predicted by the original GTH/MOLOPT approach (Table S2†), it is also apparent that under the existing NLCC correction the lighter elements exhibit significantly larger MADs compared to their heavier counterparts. This disparity is likely attributed to the greater sensitivity to the core region for lighter elements, while conversely the stronger screening effect[8] exhibited by heavier elements helps maintain the invariance of the charge density in the core region, resulting in lower MAD values. Nonetheless, we found it imperative in this case to reparameterize the NLCC PPs specifically for the B97M-rV functional. After reoptimization of the NLCC parameters at the level of B97M-rV for only the lighter main-group elements (B, C, N, O,

Table 2 Errors in atomization energies predicted by GTH PPs compared to def2-TZVPPD all-electron calculations with or without NLCC corrections and using MOLOPT or def2-TZVPPD basis sets for the meta-GGA B97M-rV functional. Mean Absolute Deviations (MADS, in kcal mol^{-1}) of the different DFT functionals and basis set combinations analyzed over the G2 series[33–35] and the TAE140 datasets[36] divided into the 124 "easy" molecules without multi-configurational character and 16 multi-reference atomization energies

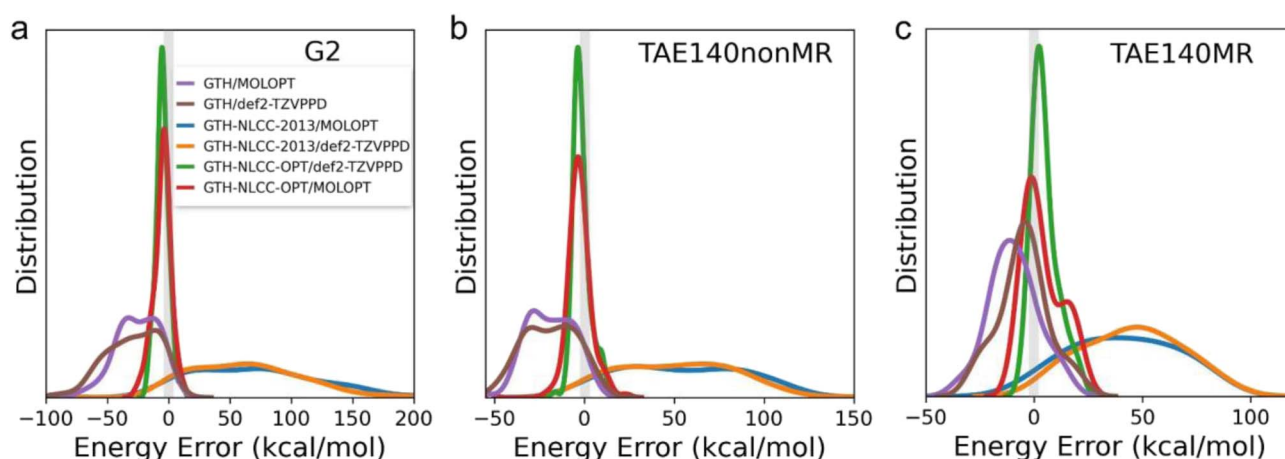| PP	GTH	GTH	GTH-NLCC	GTH-NLCC	GTH-NLCC-OPT	GTH-NLCC-OPT
Basis set	MOLOPT	def2-TZVPPD	MOLOPT	def2-TZVPPD	MOLOPT	def2-TZVPPD
Dataset						
G2	25.42	29.96	68.76	61.39	6.63	5.95
TAE140nonMR	19.38	20.36	55.23	51.93	5.62	4.16
TAE140MR	11.87	9.07	38.71	42.43	6.25	5.24

Fig. 2 Energy error distribution of atomization energies for the (a) G2, (b) TAE140nonMR and (c) TAE140MR datasets predicted by B97M-rV functional. The reference is all-electron calculation with def2-TZVPPD basis sets.

and F), leaving the original GTH parameters in place for the heavier atoms, we get more consistent results with a MAD less than 6 kcal mol⁻¹ using either the MOLOPT or more complete basis sets (Table 2 and Fig. 2).

Additionally, we investigated the transferability and errors for the standard and NLCC PPs and basis set combinations for other thermochemical properties including electron affinities (EA) and ionization potentials (IP) derived from the G21EA and G21IP datasets[41,50] as shown in Fig. 3. As detailed in Fig. S1 and S2,† ionization involves a larger energy scale than that found for electron affinity energies, as it is harder to remove an electron from the highest occupied molecular orbital (HOMO) than to attach an electron into the lowest unoccupied molecular orbital (LUMO) because the former is much deeper than the latter relative to the vacuum level. Overall we find that IPs are fairly insensitive to both PPs and basis set completeness for the pure GGA and hybrid DFT functionals, whereas the B97M-rV meta GGA benefits from NLCC optimization, after which it yields comparable results to the other DFT functionals and regardless of basis set. These figures also highlight a significant trend in which the EA is far more sensitive to the completeness of the basis set, and that the NLCC has a more minor corrective effect on the errors due to overbinding.

Table 3 provides an assessment of PP/basis set combinations for other non-reactive properties including binding energies of non-covalent interactions in organic molecules and biomolecules (S66),[42,43] binding energies of charge–transfer complexes

(CT20),[44] isomerization energies (ISOMERIZATION20),[36] hydrogen transfer barrier heights (HTBH38),[37] non-hydrogen transfer barrier heights (NHTBH38),[38] and barrier heights for water-catalyzed proton–transfer reactions (WCPT27).[39] In Table 3 we consider the accuracy and transferability of the PBE PP to the other functionals. Clearly, the NLCC effect is negligible for these non-thermochemical properties as the chemical environment is not dramatically changed before and after the reaction and the spin polarized state is less involved compared to the atomization process. Similar to the other properties examined previously, the GTH-NLCC pseudopotential at the PBE level exhibits poor transferability to the B97M-rV functional, resulting in significantly worse results compared to the GTH/MOLOPT approach without correction. However, with utilization of optimized NLCC parameters for the B97M-rV functional, substantial improvements can be achieved for non-thermal properties as well.

Discussion and conclusion

Pseudopotentials are widely employed in theoretical chemistry to provide a smooth potential profile within a specified core region, that permits removal of a specified set of core electrons from explicit consideration, thereby reducing the computational overhead and often enabling efficient inclusion of relativistic effects. However, their utilization can introduce non-negligible deviations, particularly in spin-polarized systems,
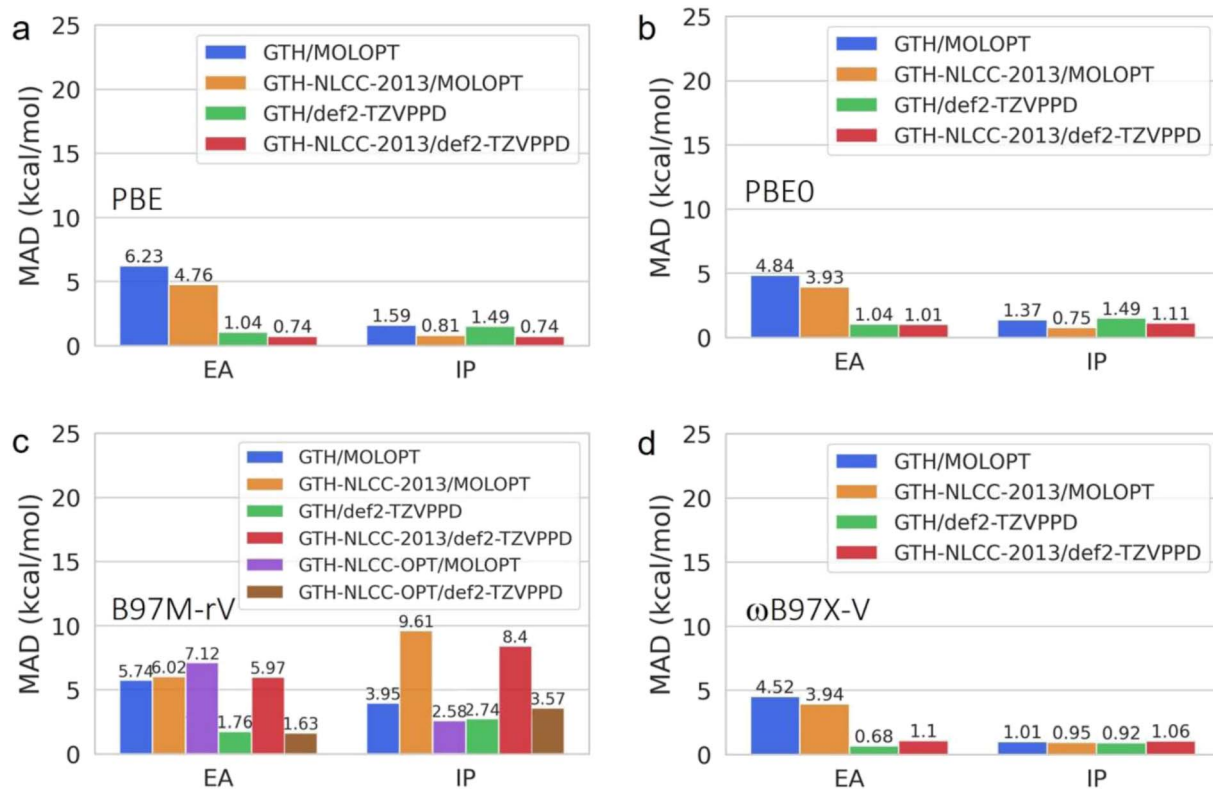

Fig. 3 Sensitivity of electron affinity (EA) and ionization potential (IP) to basis size and NLCC effect at the levels of (a) PBE, (b) PBE0, (c) B97M-rV and (d) ωB97X-V. In the subplot (c) from B97M-rV method, the additional bars are corresponding to the results predicted by optimized NLCC parameters for lighter elements (B, C, N, O, and F) and original GTH for heavier elements (Al, Si, P, S, and Cl).

Table 3 Errors in non-thermal chemistry properties predicted by GTH PPs compared to def2-TZVPPD all-electron calculations with or without NLCC corrections and using TZV2P MOLOPT or def2-TZVPPD basis sets. MAD results (kcal mol^{-1}) predicted by PBE, PBE0 and ωB97M-rV functionals for S66, CT20, ISOMERIZATION20, HTBH38, NHTBH38, and WCPT27

Functional	PP/basis	S66	CT20	ISOMER20	HTBH38	NHTBH38	WCPT27
PBE	GTH/MOLOPT	0.12	0.11	0.63	1.13	2.00	0.43
	GTH/def2-TZVPPD	0.06	0.04	0.99	0.95	1.00	0.90
	GTH-NLCC/MOLOPT	0.14	0.10	1.57	0.38	1.84	0.49
	GTH-NLCC/def2-TZVPPD	0.08	0.05	0.72	0.61	0.88	1.15
PBE0	GTH/MOLOPT	0.22	0.12	0.74	0.83	1.67	0.44
	GTH/def2-TZVPPD	0.07	0.07	0.78	0.71	1.03	1.04
	GTH-NLCC/MOLOPT	0.23	0.12	0.68	0.37	1.87	0.45
	GTH-NLCC/def2-TZVPPD	0.12	0.08	0.78	0.76	1.47	1.31
ωB97M-rV	GTH/MOLOPT	0.25	0.11	0.89	0.52	1.41	0.62
	GTH/def2-TZVPPD	0.10	0.09	0.38	0.51	0.58	1.03
	GTH-NLCC/MOLOPT	0.24	0.11	0.93	0.23	1.63	0.57
	GTH-NLCC/def2-TZVPPD	0.11	0.08	0.37	0.51	1.03	1.28
B97M-rV	GTH/MOLOPT	0.10	0.14	1.56	1.36	2.19	0.50
	GTH/def2-TZVPPD	0.22	0.09	1.46	1.20	1.53	0.80
	GTH-NLCC-2013/MOLOPT	0.30	0.21	5.74	3.96	2.40	2.05
	GTH-NLCC-2013/def2-TZVPPD	0.24	0.17	4.67	4.00	2.06	2.70
B97M-rV	GTH-NLCC-OPT/MOLOPT	0.27	0.16	0.61	0.79	2.58	0.62
	GTH-NLCC-OPT/def2-TZVPPD	0.23	0.11	0.42	0.67	0.75	0.57

as spin polarization exhibits distinct variations in response to different chemical environments, despite the apparent similarity in charge density within the invariant muffin-tin sphere. To overcome this challenge, the NLCC strategy developed for HGH pseudopotentials reported by Willand et al.[10] introduced a single Gaussian function as the core charge density, thereby attaining chemical accuracy when compared to all-electron calculations for the standard G2-1 dataset for the PBE functional as well as for semiempirical models corrected by Grimme dispersion terms.

In this work we tested the transferability and accuracy of the original NLCC parameters to the GTH pseudopotential developed for the PBE GGA functional[51] and associated with the TZV2P MOLOPT basis set,[23] to hybrid DFT functionals including PBE0 (ref. 19) and the range-separated hybrid, meta-GGA ωB9X-V,[20] as well as the B97M-rV meta-GGA functional. Our investigation of transferability and accuracy encompassed a wide range of both thermochemistry and non-thermochemistry datasets, and our findings indicate that the NLCC correction has the largest impact on atomization energies, effectively rectifying the original GTH/MOLOPT method and resulting in an error reduction of less than 1% when compared to all-electron calculations. Importantly, the energy-consistent Stuttgart–Dresden relativistic PPs[52,53] have also been specifically tailored to faithfully reproduce quantities such as atomic excitation and ionization energies, and showcased remarkable congruence between PPs and all-electron outcomes pertaining to dissociation energies across an array of atoms, ions, and dimers[54–59] a principle that is in alignment with the NLCC concept. However, we have previously addressed PP transferability errors, and in that work by Rossomme and co-workers we conducted a comprehensive comparison involving the Stuttgart–Dresden effective core potentials.[17] Notably, Table S8† within that paper unveiled a significant insight: the

performance of the Stuttgart–Dresden PP, even in conjunction with the def2-QZVPPD basis set, is comparable to the GTH results without element-specific "atomic" corrections.[17]

For all properties investigated, the existing NLCC parameters optimized at the level of PBE have good transferability to PBE0 and ωB97M-rV while maintaining satisfactory accuracy using the TZV2P MOLOPT basis sets, although electron affinities were found to be more sensitive to the completeness of basis sets than the NLCC correction. In contrast, the GTH-NLCC PP for PBE shows poor transferability to the meta-GGA B97M-rV functional, with significant deviations of over 50 kcal mol^{-1} in atomization energies when compared to all-electron calculations, thus requiring a reoptimization of the NLCC parameters specifically for this DFT functional. We note that similar PPIEs were found for the SCAN functional as described by Rossomme and co-workers,[17] indicating that meta-GGA functionals may be unusually sensitive to PP replacement of core electrons. Fortunately, our optimization of the GTH-NLCC PP for B97M-rV overcame the drastic errors using the standard GTH/MOLOPT, in which optimization was only necessary for lighter second-row elements while heavier elements did not require any changes, which may be insightful for improving other meta-GGA functionals. We note that the errors in the PPs for the meta-GGA may arise from the assumption made regarding the kinetic energy density, which neglects the contribution of the core charge and focuses solely on the valence density (which is also an assumption in the CP2K program). This highlights the need, and opportunity, for developing an NLCC correction for the kinetic energy density since meta-GGAs may offer better DFT accuracy at an affordable cost.

In addition to transferability, the GTH-NLCC PPs combined with the MOLOPT basis sets performed overall as well as the complete basis set calculations for PBE and hybrid functionals, and comparably for B97M-rV after optimization of its non-

transferable GTH-NLCC PP. Thus our conclusion is that the GTH-NLCC PP/MOLOPT combination can be reliably employed in large-scale systems, effectively reducing computational costs while maintaining good accuracy. In summary, we believe these findings contribute to a better understanding of the sources of error in calculations using PPs and provide a protocol for enhancing the reliability of other DFT functionals when combined with chosen PPs and basis sets.

Data availability

All data generated from this work can be found at **https://github.com/THGLab/Pseudopotential**.

Author contributions

W. L. L., K. C. and T. H. G. designed the project. W. L. L. and K. C. carried out all calculations. W. L. L. and T. H. G. wrote the paper. All authors discussed and analyzed the results and made comments and edits to the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 P. Schwerdtfeger, The Pseudopotential Approximation in Electronic Structure Theory, *ChemPhysChem*, 2011, **12**, 3143–3155.

2 S. Goedecker, M. Teter and J. Hutter, Separable dual-space Gaussian pseudopotentials, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.

3 C. Hartwigsen, S. Goedecker and J. Hutter, Relativistic separable dual-space Gaussian pseudopotentials from H to Rn, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 3641–3662.

4 B. G. Lippert, J. Hutter and M. Parrinello, A hybrid Gaussian and plane wave density functional scheme, *Mol. Phys.*, 1997, **92**, 477–488.

5 R. Martin, *Electronic Structure – Basic Theory and Practical Methods*, Cambridge University Press, West Nyack, NY, 2004.

6 B. J. Austin, V. Heine and L. J. Sham, General Theory of Pseudopotentials, *Phys. Rev.*, 1962, **127**, 276–282.

7 G. B. Bachelet, D. R. Hamann and M. Schlüter, Pseudopotentials that work: from H to Pu, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1982, **26**, 4199–4228.

8 S. Goedecker and K. Maschke, Transferability of pseudopotentials, *Phys. Rev. A*, 1992, **45**, 88–93.

9 S. G. Louie, S. Froyen and M. L. Cohen, Nonlinear ionic pseudopotentials in spin-density-functional calculations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1982, **26**, 1738–1742.

10 A. Willand, Y. O. Kvashnin, L. Genovese, A. Vazquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch and S. Goedecker, Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations, *J. Chem. Phys.*, 2013, **138**, 104109.

11 J.-B. Lu, D. C. Cantu, M.-T. Nguyen, J. Li, V.-A. Glezakou and R. Rousseau, Norm-Conserving Pseudopotentials and Basis Sets To Explore Lanthanide Chemistry in Complex Environments, *J. Chem. Theory Comput.*, 2019, **15**, 5987–5997.

12 J.-B. Lu, D. C. Cantu, C.-Q. Xu, M.-T. Nguyen, H.-S. Hu, V.-A. Glezakou, R. Rousseau and J. Li, Norm-Conserving Pseudopotentials and Basis Sets to Explore Actinide Chemistry in Complex Environments, *J. Chem. Theory Comput.*, 2021, **17**, 3360–3371.

13 W.-L. Li, K. Chen, E. Rossomme, M. Head-Gordon and T. Head-Gordon, Optimized Pseudopotentials and Basis Sets for Semiempirical Density Functional Theory for Electrocatalysis Applications, *J. Phys. Chem. Lett.*, 2021, **12**, 10304–10309.

14 J.-B. Lu, X.-L. Jiang, H.-S. Hu and J. Li, Norm-Conserving 4f-in-Core Pseudopotentials and Basis Sets Optimized for Trivalent Lanthanides (Ln = Ce–Lu), *J. Chem. Theory Comput.*, 2023, **19**, 82–96.

15 S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction, *J. Comput. Chem.*, 2006, **27**, 1787–1799.

16 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu, *J. Chem. Phys.*, 2010, **132**, 154104.

17 E. Rossomme, L. A. Cunha, W. Li, K. Chen, A. R. McIsaac, T. Head-Gordon and M. Head-Gordon, The Good, the Bad, and the Ugly: Pseudopotential Inconsistency Errors in Molecular Applications of Density Functional Theory, *J. Chem. Theory Comput.*, 2023, **19**, 2827–2841.

18 J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, CP2K: atomistic simulations of condensed matter systems, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 15–25.

19 C. Adamo and V. Barone, Toward reliable density functional methods without adjustable parameters: the PBE0 model, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

20 N. Mardirossian and M. Head-Gordon, omegaB97M-V: a combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation, *J. Chem. Phys.*, 2016, **144**, 214110.

21 N. Mardirossian, L. Ruiz Pestana, J. C. Womack, C.-K. Skylaris, T. Head-Gordon and M. Head-Gordon, Use

of the rVV10 Nonlocal Correlation Functional in the B97M-V Density Functional: Defining B97M-rV and Related Functionals, *J. Phys. Chem. Lett.*, 2017, **8**, 35–40.

22 N. Mardirossian and M. Head-Gordon, Mapping the genome of meta-generalized gradient approximation density functionals: the search for B97M-V, *J. Chem. Phys.*, 2015, **142**, 074111.

23 J. VandeVondele and J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases, *J. Chem. Phys.*, 2007, **127**, 114105.

24 M. Krack, Pseudopotentials for H to Kr optimized for gradient-corrected exchange–correlation functionals, *Theor. Chem. Acc.*, 2005, **114**, 145–152.

25 F. Zahariev, S. S. Leang and M. S. Gordon, Functional derivatives of meta-generalized gradient approximation (meta-GGA) type exchange–correlation density functionals, *J. Chem. Phys.*, 2013, **138**, 244108.

26 C. Adamo, M. Ernzerhof and G. E. Scuseria, The meta-GGA functional: thermochemistry with a kinetic energy density dependent exchange–correlation functional, *J. Chem. Phys.*, 2000, **112**, 2643–2649.

27 J. VandeVondele and J. Hutter, An efficient orbital transformation method for electronic structure calculations, *J. Chem. Phys.*, 2003, **118**, 4365–4369.

28 V. Weber, J. VandeVondele, J. Hutter and A. M. N. Niklasson, Direct energy functional minimization under orthogonality constraints, *J. Chem. Phys.*, 2008, **128**, 084113.

29 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.

30 D. Rappoport and F. Furche, Property-optimized Gaussian basis sets for molecular response calculations, *J. Chem. Phys.*, 2010, **133**, 134105.

31 F. Weigend, F. Furche and R. Ahlrichs, Gaussian basis sets of quadruple zeta valence quality for atoms H–Kr, *J. Chem. Phys.*, 2003, **119**, 12753–12762.

32 N. Mardirossian and M. Head-Gordon, Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals, *Mol. Phys.*, 2017, **115**, 2315–2372.

33 L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation, *J. Chem. Phys.*, 1997, **106**, 1063–1079.

34 J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari and L. A. Curtiss, Gaussian-1 theory: a general procedure for prediction of molecular energies, *J. Chem. Phys.*, 1989, **90**, 5622–5629.

35 L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari and J. A. Pople, Gaussian-1 theory of molecular energies for second-row compounds, *J. Chem. Phys.*, 1990, **93**, 2537–2545.

36 A. Karton, S. Daon and J. M. Martin, W4-11: a high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data, *Chem. Phys. Lett.*, 2011, **510**, 165–178.

37 Y. Zhao, B. J. Lynch and D. G. Truhlar, Multi-coefficient extrapolated density functional theory for thermochemistry and thermochemical kinetics, *Phys. Chem. Chem. Phys.*, 2005, **7**, 43–52.

38 Y. Zhao, N. González-García and D. G. Truhlar, Benchmark Database of Barrier Heights for Heavy Atom Transfer, Nucleophilic Substitution, Association, and Unimolecular Reactions and Its Use to Test Theoretical Methods, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.

39 A. Karton, R. J. O'Reilly and L. Radom, Assessment of Theoretical Procedures for Calculating Barrier Heights for a Diverse Set of Water-Catalyzed Proton-Transfer Reactions, *J. Phys. Chem. A*, 2012, **116**, 4211–4221.

40 L. Goerigk and S. Grimme, A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions Assessment of Common and Reparameterized (meta-)GGA Density Functionals, *J. Chem. Theory Comput.*, 2010, **6**, 107–126.

41 L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, Gaussian-2 theory for molecular energies of first- and second-row compounds, *J. Chem. Phys.*, 1991, **94**, 7221–7230.

42 J. Řezáč, K. E. Riley and P. Hobza, Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries, *J. Chem. Theory Comput.*, 2011, **7**, 3466–3470.

43 K. U. Lao and J. M. Herbert, Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The XSAPT Family of Methods, *J. Phys. Chem. A*, 2015, **119**, 235–252.

44 S. N. Steinmann, C. Piemontesi, A. Delachat and C. Corminboeuf, Why are the Interaction Energies of Charge-Transfer Complexes Challenging for DFT?, *J. Chem. Theory Comput.*, 2012, **8**, 1629–1640.

45 M. Stöhr, T. Van Voorhis and A. Tkatchenko, Theory and practice of modeling van der Waals interactions in electronic-structure calculations, *Chem. Soc. Rev.*, 2019, **48**, 4118–4154.

46 J. Sun, B. Xiao, Y. Fang, R. Haunschild, P. Hao, A. Ruzsinszky, G. I. Csonka, G. E. Scuseria and J. P. Perdew, Density Functionals that Recognize Covalent, Metallic, and Weak Bonds, *Phys. Rev. Lett.*, 2013, **111**, 106401.

47 D. Cremer, Density functional theory: coverage of dynamic and non-dynamic electron correlation effects, *Mol. Phys.*, 2001, **99**, 1899–1940.

48 J. N. Harvey, On the accuracy of density functional theory in transition metal chemistry, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, 2006, **102**, 203–226.

49 N. Mardirossian and M. Head-Gordon, omegaB97X-V: a 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9904–9924.

50 L. Goerigk and S. Grimme, Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group

Thermochemistry, Kinetics, and Noncovalent Interactions, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.

51 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

52 W. Küchle, M. Dolg, H. Stoll and H. Preuss, Ab initio pseudopotentials for Hg through Rn, *Mol. Phys.*, 1991, **74**, 1245–1263.

53 A. Bergner, M. Dolg, W. Küchle, H. Stoll and H. Preuß, Ab initio energy-adjusted pseudopotentials for elements of groups 13–17, *Mol. Phys.*, 1993, **80**, 1431–1441.

54 D. Andrae, U. Häußermann, M. Dolg, H. Stoll and H. Preuß, Energy-adjusted ab initio pseudopotentials for the second and third row transition elements, *Theor. Chim. Acta*, 1990, **77**, 123–141.

55 M. Dolg, U. Wedig, H. Stoll and H. Preuss, Energy-adjusted ab initio pseudopotentials for the first row transition elements, *J. Chem. Phys.*, 1987, **86**, 866–872.

56 P. Fuentealba, H. Stoll, L. von Szentpaly, P. Schwerdtfeger and H. Preuss, On the reliability of semi-empirical pseudopotentials: simulation of Hartree–Fock and Dirac–Fock results, *J. Phys. B: At. Mol.*, 1983, **16**, L323.

57 P. Fuentealba, L. von Szentpaly, H. Preuss and H. Stoll, Pseudopotential calculations for alkaline-earth atoms, *J. Phys. B: At. Mol.*, 1985, **18**, 1287.

58 G. Igel-Mann, H. Stoll and H. Preuss, Pseudopotentials for main group elements (IIIa through VIIa), *Mol. Phys.*, 1988, **65**, 1321–1328.

59 M. Dolg, H. Stoll and H. Preuss, Energy-adjusted ab initio pseudopotentials for the rare earth elements, *J. Chem. Phys.*, 1989, **90**, 1730–1734.