RSC Publishing PCCP

Auxiliary basis sets for density-fitted correlated wavefunction calculations: Weighted core-valence and ECP basis sets for post-d elements

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID:	CP-ART-02-2012-040400.R1
Article Type:	Paper
Date Submitted by the Author:	07-Mar-2012
Complete List of Authors:	Hattig, Christof; Ruhr University Bochum, Theoretical Chemistry Department Kossmann, Joerg; Ruhr University Bochum, Theoretical Chemistry Schmitz, Gunnar; Ruhr University Bochum, Theoretical Chemistry

SCHOLARONE™
Manuscripts

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

Auxiliary basis sets for density-fitted correlated wavefunction calculations: Weighted core-valence and ECP basis sets for post-*d* elements †

Christof Hättig,[*a‡] **Gunnar Schmitz,**[a§] **and Jörg Koßmann**[a¶]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

We report optimised auxiliary basis sets for the resolution-of-the-identity (or density-fitting) approximation of two-electron integrals in second-order Møller-Plesset perturbation theory (MP2) and similar electronic structure calculations with correlation-consistent basis sets for the post-*d* elements Ga–Kr, In–Xe, and Tl–Rn. The auxiliary basis sets are optimised such that the density-fitting error is negligible compared to the one-electron basis set error. To check to which extend this criterion is fulfilled we estimated for a test set of 80 molecules the basis set limit of the correlation energy at the MP2 level and evaluated the remaining density-fitting and the one-electron basis set errors. The resulting auxiliary basis sets are only 2–6 times larger than the corresponding one-electron basis sets and lead in MP2 calculations to speed-ups of the integral evaluation by one to three orders of magnitude. The density-fitting errors in the correlation energy are at least hundred times smaller than the one-electron basis set error, i.e. in the order of only 1–100 μH per atom.

1 Introduction

The resolution-of-the-identity (RI) approximation—also known as density-fitting (DF) approximation—for two-electron integrals is used since the 1970's in the context of density-functional theory to reduce the costs for the computation of the Kohn-Sham matrix[1,2] and as alternative to numerical quadrature[3]. During the last two decades it has also become an important tool to reduce the computational costs for several correlated wavefunction methods. For second-order Møller-Plesset perturbation theory (MP2) the RI approximation reduces the computational costs by one to three orders of magnitude and, thereby, significantly extends the applicability of MP2 to larger molecules. RI- or DF-MP2 codes are nowadays available in several widely used quantum chemistry packages.[4–10] Similar savings are obtained for other second-order methods e.g. CC2[11,12], or CIS(D)[13] or their spin-scaled variants[14–16]. The application of the RI approximation to additional two-electron integrals needed for explicitly correlated so-called F12 methods[17] has been an important cornerstone in turning these methods into highly efficient tools for electronic structure calculations[18,19]. In this context it has been essential that the RI or DF approximation is applicable to arbitrary two-electron operators which, for example, do not need to be positive definite.[17] This generality and its higher efficiency are two main advantages of the RI approximation compared to the Cholesky decomposition of the two-electron integral matrices[20], which is an alternative inner projection approximation for two-electron integrals. The latter has however the advantage that it does not require auxiliary basis sets and has an accuracy controlled by a single threshold. The RI approximation has also been instrumental in increasing the efficiency of local correlation methods[21] and recently also in reviving the pair natural orbital (PNO) approach to local methods[22].

A prerequisite for the efficiency and accuracy of the RI approximation are auxiliary basis sets optimised for the type of one-electron basis used in a calculation. To enhance further the efficiency, the optimisation of the auxiliary basis is usually also tuned to the class of methods or, more precisely, the class of integrals or intermediates which are evaluated within this approximation.[2,23–27] For correlated wavefunction methods including the domain- and PNO-based local methods[22] and the F12 methods[18] it has during the last decade become standard to employ auxiliary basis sets optimised at the MP2 level according to a recipe proposed by Weigend *et al.*[23] Such optimised auxiliary basis sets for density-fitted correlated wavefunction calculations are today available[23–25,28]

† Electronic Supplementary Information (ESI) available: optimised auxiliary basis sets for Ga–Kr (cc-pV5Z, (aug-)cc-pwCV*n*Z with *n*=D,T,Q,5, and (aug-)cc-p(wC)V*n*Z-PP with *n*=D,T,Q,5) and for In–Xe and Tl–Rn ((aug-)cc-p(wC)V*n*Z-PP with *n*=D,T,Q,5). See DOI: 10.1039/b000000x/

[a] *Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany.*
‡ E-mail: christof.haettig@rub.de
§ E-mail: gunnar.schmitz@rub.de
¶ E-mail: joerg.kossmann@rub.de

for the atoms H–Rn and the segmented contracted orbital basis sets SVP[29], TZVP[30], TZVPP[31], and QZVPP[32], including the respective def[2,31] and revised def2[33] effective core potential (ECP) versions for heavy elements, developed by Ahlrichs and Weigend and coworkers. For the correlation consistent basis set families developed by Dunning and coworkers[34–42] and Peterson and coworkers[43–50] optimised auxiliary basis sets for correlated calculations have been published for cc-pVnZ, cc-pV(n+d)Z, aug-cc-pVnZ, aug-cc-pV(n+d)Z with n = D, T, Q, 5, 6 for H, He, Li–Ne, Na–Ar[25,26] (for n= D, T, Q also for Ga–Kr[26] and for n = T additionally for Sc–Zn[51]), for the weighted core-valence (aug-)cc-pwCVnZ with n = D, T, Q, 5 for B–Ne, Al–Ar[25] and for the pseudo-potential basis sets (aug-)cc-pVnZ-PP for Y–Pd[52,53].

In the present paper we report about the extension of our previous work on the auxiliary basis set towards the correlation-consistent all-electron and pseudo-potential valence (cc-pVnZ and cc-pVnZ-PP) and weighted core valence (cc-pwCVnZ and cc-pwCVnZ-PP) basis sets for the post-d elements Ga–Kr, In–Xe, and Tl–Rn, including the versions augmented with diffuse functions, aug-cc-p(wC)VnZ(-PP). An advantage of the general contracted correlation consistent basis set family compared to the segmented contracted (def2)-SVP, (def2)-TZVP, (def2)-TZVPP and (def2)-QZVPP basis sets reported by Ahlrichs, Weigend and coworkers is the strictly systematic building principle which allows for a simple but quite reliable extrapolation of correlation energies towards the basis set limit[54], which is not to the same extend possible with the Ahlrichs/Weigend basis sets. They are therefore often preferred for highly accurate calculations, even though the computational costs for the integral evaluation are with these basis sets somewhat larger than with segmented contracted basis sets. Anyway, this disadvantage is diminished if the two-electron integrals are evaluated through the RI approximation. The current work thus paves the way for a more efficient calculation of correlation energies close the basis set limit also for molecules containing post-d elements.

The auxiliary valence basis sets (aug-)cc-pVnZ-PP are already available in the Basis Set Exchange database[28] and distributed as part of the basis set library of the TURBOMOLE program[10] package. The remaining auxiliary basis sets will be made available through the same databases in the near future. In addition all basis sets reported in this work are available as Electronic Supplementary Information.[†]

2 Optimisation of the auxiliary basis sets

As in our previous work[24–26] we optimised the auxiliary basis sets for correlation consistent one-electron orbital basis sets for the post-d elements by minimising the positive semi-definite functional

$$\delta_{\mathrm{RI}} = \frac{1}{4}\sum_{aibj}\frac{\left(\langle ab||ij\rangle_{\mathrm{RI}} - \langle ab||ij\rangle\right)^2}{\varepsilon_a - \varepsilon_i + \varepsilon_b - \varepsilon_j}\,, \qquad (1)$$

where $\langle ab||ij\rangle = (ai|bj) - (aj|bi)$ is an antisymmetrized two-electron repulsion integral, $\langle ab||ij\rangle_{\mathrm{RI}}$ the density-fitting approximation for $\langle ab||ij\rangle$ and the ε_p canonical Hartree-Fock orbital energies. Here and in the following we use the convention that indices i, j refer to active occupied and a, b to unoccupied (virtual) canonical spin-orbitals.

The functional δ_{RI} is related to the error introduced by the approximate resolution of the identity for the two-electron integrals in the MP2 correlation energy[25,26]. It vanishes only if the RI approximation becomes exact for all integrals $\langle ab||ij\rangle$ included in the summation in Eq. (1). The orbital energy denominators give to each integral a weight related to its importance for correlation energies and thereby tunes the auxiliary basis for an optimal convergence of energies, potential energy surfaces and other energy-related properties. The auxiliary basis sets obtained by minizing δ_{RI} perform very well in the approximation of two-electron integrals of the form $\langle ip|\hat{o}_{12}|ap\rangle$ with at least one occupied/virtual product (ai) and arbitrary functions p and q and general two-electron operators $\hat{o}_{12}$ (which is important for there use in F12 methods).

To keep the auxiliary basis sets for the post-d elements consistent with those for elements H–Kr, we use uncontracted Gaussian type orbitals (GTOs). A summary of the composition of the auxiliary basis sets, i.e. the number of GTOs for each angular momentum, is given in Table 1. The minimisation of δ_{RI} with respect to the exponents of the auxiliary GTOs was done the with relax module of the TURBOMOLE package[10] in combination with the analytic gradient for the functional δ_{RI} reported in Ref. 25. All optimisations were carried out for the atoms using spin-unrestricted Hartree-Fock molecular orbitals and orbital energies. No symmetry restrictions were applied in the Hartree-Fock calculations.

2.1 All-electron quintuple-ζ basis sets for Ga–Kr

In Refs. 25,26 no auxiliary cc-pV5Z basis sets were reported for the atoms Ga–Kr, because cc-pVnZ orbital basis sets are constructed to correlate only the electrons in the valence 4s4p shell. Since for accurate calculations also the correlation of the outer-core 3d-orbitals has to be included, calculations in the cc-pV5Z basis for Ga–Kr are uncomely. The cc-pV5Z orbital basis sets are, however, also the headstones for the core-valence basis sets cc-pCV5Z and cc-pwCV5Z, which are derived from the parent basis by augmentation with steep correlation functions (and some minor changes in the contraction of the s and p functions). A similar building principle is used for the corresponding auxiliary basis sets[25] which motivated

This journal is © The Royal Society of Chemistry [year]

Table 1 Composition of the auxiliary basis sets. The ratio N_{Aux}/N_{Orb} between the size of auxiliary and the corresponding orbitals basis is given in parentheses behind the basis set composition. For the core-valence basis sets the number of functions replaced in the underlying auxiliary valence basis is listed in a second row in parentheses.

	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z
Ga–Kr	9s8p7d4f2g (4.2)	10s9p8d6f4g1h (3.9)	11s10p9d7f5g3h1i (3.3)	15s13p12d10f8g5h3i1k (3.5)
	wC-DZ	wC-TZ	wC-QZ	wC-5Z
Ga–Kr	+4s4p4d3f3g3h (5.4) (1f2g)	+6s5p5d4f4g4h2i (4.1) (3p2f1h)	+5s6p6d5f4g3h3i2k (3.1) (7s1p3f2g2h1i)	+5s6p5d6f5g5h3i3k2l (3.0) (1g5h3i1k)
	aug-DZ	aug-TZ	aug-QZ	aug-5Z
Ga–Kr	+1s1p1d1f (4.7)	+1s1p1d1f1g (3.7)	+1s1p1d1f1g1h (2.9)	+1s1p1d1f1g1h1i1k (2.8)
	pseudo-potential basis sets			
	cc-pVDZ-PP	cc-pVTZ-PP	cc-pVQZ-PP	cc-pV5Z-PP
Ga–Kr	8s7p6d3f2g (4.3)	8s8p7d5f4g1h (3.8)	10s10p9d8f6g3h1i (3.7)	13s12p11d9f8g5h3i1k (3.5)
In–Xe	7s6p5d3f2g (3.9)	8s8p7d5f3g1h (3.6)	11s11p9d7f5g3h1i (3.4)	13s12p11d9f7g5h3i1k (3.3)
Tl–Rn	7s6p5d3f2g (3.9)	8s8p7d5f3g2h (3.9)	11s11p9d7f5g3h2i (3.6)	13s12p11d9f7g5h3i1k (3.3)
	wC-DZ-PP	wC-TZ-PP	wC-QZ-PP	wC-5Z-PP
Ga–Kr	+1s2p2d2f3g2h (4.6) (3f2g)	+2s2p2d3f3g3h2i (3.4) (3f1g)	+3s3p3d3f3g3h3i2k (3.0) (4g1h)	+2s3p3d3f4g3h3i3k1l (2.5) (1f1h)
In–Xe	+1s3p3d2f2g2h (4.3) (1f)	+3s3p3d2f3g2h2i (3.2) (3f1g)	+2s1p2d3f3g2h3i2k (2.7) (2p4d2f2h)	+3s3p3d3f3g3h3i3k1l (2.4) (1f1g2h1i)
Tl–Rn	+2s2p2d3f2g2h (4.3) —	+3s3p2d2f3g1h2i (3.1) (3d)	+2s2p2d3f3g2h2i2k (2.7) (3f2g1h)	+3s3p3d3f3g3h2i3k1l (2.3) (1s1p3f4g3h2i)
	aug-DZ-PP	aug-TZ-PP	aug-QZ-PP	aug-5Z-PP
Ga–Kr	+1s1p1d1f1g (4.2)	+1s1p1d1f1g1h (3.2)	+1s1p1d1f1g1h1i (2.8)	+1s1p1d1f1g1h1i1k (2.4)
In–Xe	+1s1p1d1f1g (4.0)	+1s1p1d1f1g1h (3.0)	+1s1p1d1f1g1h1i (2.6)	+1s1p1d1f1g1h1i1k (2.3)
Tl–Rn	+1s1p1d1f1g (4.0)	+1s1p1d1f1g1h (3.0)	+1s1p1d1f1g1h1i (2.6)	+1s1p1d1f1g1h1i1k (2.3)

the optimisation of auxiliary cc-pV5Z and aug-cc-pV5Z sets also for Ga–Kr.

Following the same conservative strategy as for cc-pVnZ with n = D, T, Q as described in Ref. 26, we kept also for the optimisation of the auxiliary cc-pV5Z sets only the $1s^2 2s^2 2p^6$ shells frozen, even though the orbital basis set has only been optimised to correlate electrons in the valence $4s4p$ shell. The inclusion of the outer-valence $3d$ electrons lead to the requirement that the auxiliary basis should contain up to k functions. (The general requirement for fitting occupied/virtual products $\phi_i\phi_a$ is that the auxiliary basis contains functions up to $l_{max} = l_{occ} + l_{orb}$, where l_{occ} and l_{orb} are the maximum angular momenta of, respectively, an occupied and an unoccupied or general molecular orbital.)

For the auxiliary aug-cc-pV5Z basis sets we added for each angular momentum s–k one primitive function which was optimised for the electronic ground state of the anions of the atoms.

2.2 All-electron weighted core-valence basis sets for Ga–Kr

The auxiliary basis set for the all-electron weighted core-valence sets cc-pwCVnZ for the atoms Ga–Kr were constructed by augmenting the corresponding valence auxiliary basis sets with additional tight functions and—in accord with the rule $l_{max} = l_{occ} + l_{orb}$—with functions with angular momenta of $l = n + 3$ (where n is the cardinal number). The number of additional primitive GTOs was chosen such that in calculations for the electronic ground states of the atoms with the cc-pwCVnZ basis sets the error functional $\delta_{\rm RI}$ in Eq. (1) was at least as small as for the corresponding valence cc-pVnZ auxiliary and orbital basis set combination. For the cc-pwCVnZ basis set we correlated, however, also the electrons in the $2s^2 2p^6$ shell and kept only the $1s^2$ shell frozen to ensure that the cc-pwCVnZ auxiliary basis sets can also be applied in calculations with unusual small frozen cores.

The exponents of the additional tight and higher angular momentum functions have been determined by minimising $\delta_{\rm RI}$ or by even-tempered interpolation between exponents of the cc-pVnZ auxiliary sets. To avoid near linear dependencies and unnecessary large auxiliary sets, we also replaced some exponents of the cc-pVnZ auxiliary sets by new exponents optimised together with the other added basis functions for the cc-pwCVnZ orbital basis. The ratio N_{Aux}/N_{Orb} between the number functions in the auxiliary and the orbital basis sets is for the cc-pwCVDZ and cc-pwCVTZ basis sets somewhat larger, but for the cc-pwCVQZ and cc-pwCV5Z basis sets somewhat smaller than for respective valence basis sets. A similar trend is observed below for the pseudo-potential basis sets.

2.3 Pseudo-potential valence basis sets for post-d elements

Similar as the all-electron cc-pVnZ basis sets for Ga-Kr, also the pseudo-potential orbital basis sets cc-pVnZ-PP of Refs. 43,44,50 are optimised for correlation of the electrons in the outer nsp valence shell. To correlate also efficiently the $(n-1)d$ semi-core, which is essential for accurate calculations, these basis sets must be augmented with tight correlation functions to cc-pwCVnZ basis sets.

For this reason we designed the cc-pVnZ-PP auxiliary basis sets such that they serve as a good starting point for cc-pwCVnZ-PP auxiliary basis sets and included in the calculation of the error functional $\delta_{\rm RI}$ during the auxiliary basis set optimisation all electrons not replaced by the pseudo-potential. We thus applied for the construction of the cc-pVnZ-PP sets the same strategy as used in Ref. 26 and above for the all-electron basis sets for the third row atoms Ga–Kr. Compared to the orbital basis sets, the replacement of the innermost core electrons by the pseudo-potentials leads for the auxiliary basis sets only to a moderate reduction of the number of exponents needed to achieve a desired accuracy. The composition of the cc-pVnZ-PP auxiliary basis sets for Ga–Kr differs thus from the composition of the cc-pVnZ all-electron auxiliary basis sets only by a small reduction in the number of (tight) s, p, d and f functions. An exception are the quadruple-ζ sets for Ga–Kr, where we included for the pseudo-potential basis sets an additional g function, which led to a significant reduction of the RI error.

The auxiliary basis sets for the fourth and fifth row atoms In–Xe and Tl–Rn have a similar composition as those for Ga–Kr, which is of cause a consequence of the similarity between the respective orbital basis sets. Thus also the ratios N_{bas}/N_{orb} are similar for third, fourth, and fifth row atoms.

The aug-cc-pVnZ basis set for Ga–Kr for n = 2, 3, and 4 reported in Ref. 26 did not include a diffuse function for the highest angular momentum $l_{max} = 2 + n$ as such functions are only needed for double excitations of two electrons out of the semi-core $(n-1)d$ shell into the diffuse functions with highest angular momentum in the orbital basis set, which contribute usually only little to the correlation energy. In the present work we have been somewhat more conservative and included such diffuse functions for all aug-cc-pVnZ-PP basis sets (with n = D, T, Q, 5) similar as above for the all-electron aug-cc-pV5Z basis sets. The reason for this is that most molecules studied by ab initio methods, which contain post-d elements, contain actually only a few of these atoms, but they play often a central role for the investigated problem and we want to avoid that the overall accuracy might be limited by the RI error for these atoms.

This journal is © The Royal Society of Chemistry [year]

Table 2 Test sets of molecules used to evaluate the density-fitting and the one-electron basis set errors. The structures—optimised at the DFT/B-P86/SV(P) level—have been taken from Refs. 23,24. The test sets are identical to the subsets of molecules used in Ref. 24 for the post-*d* elements, with the exception that IK and AtNb are omitted since for K and Nb not all required (auxiliary) basis sets are yet available.

Ga–Kr	As_4, $AsCl_3$, $AsCl_6^-$, AsH_3, As_4S_4, Br_2, BrCl, BrO^-, BrO_2^-, BrO_3^-, BrO_4^-, GaCl, $GaCl_3$, GaF, GaF_2, GaH_3, GaO, $GeCl_4$, GeF_3, GeF_4, GeH_4, GeO, GeO_2, Kr_2, Se_8, SeH_2, SeO, SeO_2, HBr, $HCBr_3$
In–Xe	InCl, $InCl_3$, InH, InH_3, InO, SnF_3, SnH_4, SnO, SnO_2, $SbCl_6^-$, SbF, SbF_3, SbH_3, SbO_2, TeF_3, TeH_2, TeO, TeO_2, I_2, IH, IO_4^-, ICl, IO_3^-, XeF_2, XeF_4, $XeOF_4$
Tl–Rn	TlCl, $TlCl_3$, TlH, TlH_3, TlO, PbF_3, PbH_4, PbO, PbO_2, $BiCl_6^-$, BiF, BiF_3, BiH_3, BiO_2, $PoBr_2$, $PoCl_4$, PoF_2, PoF_6, PoO_2, At_2, AtCl, AtF_3, Rn_2, RnF_2

2.4 Pseudo-potential weighted core-valence basis sets for post-*d* elements

The auxiliary basis for the weighted core-valence basis sets cc-pwCV*n*Z-PP have been built similar to the all-electron cc-pwCV*n*Z sets for Ga–Kr by adding to the valence auxiliary sets cc-pV*n*Z-PP additional tight functions and a set of higher angular momentum functions with $l = n+3$ until the value of δ_{RI} for the cc-pwCV*n*Z-PP basis sets was as small or smaller than for the respective valence cc-pV*n*Z-PP basis. In addition we replaced some functions of the cc-pV*n*Z-PP auxiliary sets as indicated in Table 1 to avoid near linear dependencies and to obtain balanced auxiliary basis sets. The exponents of the new functions were either optimised by minimising δ_{RI} or interpolated as even-tempered series between exponents of the cc-pV*n*Z-PP auxiliary basis, which were kept fix.

We found for all of the post-*d* atoms and for all four ζ-levels D, T, Q, and 5 that the number of tight functions that had to be added to the cc-pV*n*Z-PP auxiliary sets to achieve the required accuracy is somewhat smaller than it has been for the all-electron basis sets of Ga–Kr with the same cardinal number. As consequence, the resulting auxiliary basis sets have also somewhat smaller ratios N_{Aux}/N_{Orb} than the all-electron basis sets: 4.3–4.6 for cc-pwCVDZ-PP (5.4 for cc-pwCVDZ of Ga–Kr) and 2.3–2.5 for cc-pwcV5Z-PP (3.0 for cc-pwCV5Z of Ga–Kr).

3 Performance in calculations on molecules

To evaluate the performance of the new auxiliary basis sets we employed a test set of molecules (listed in Table 2) taken from Refs. 23,24. It was designed such that it covers all important oxidation states for the post-*d* elements Ga–Kr, In–Xe, and Tl–Rn. For this test set we computed the one-electron and auxiliary basis set errors in the MP2 valence plus outer core correlation energies with the (aug-)cc-pwCV*n*Z and (aug-)cc-pwCV*n*Z-PP basis sets for the post-*d* elements combined with the (aug-)cc-pwCV*n*Z sets[49] for the main group elements C–F, S and Cl and the (aug-)cc-pV*n*Z sets[35] for H and the corresponding optimised auxiliary basis sets from the present work and Ref. 25. The one-electron basis set errors were calculated with respect to estimated basis set limits obtained by the two-point extrapolation[54]

$$E_{lim}^{corr} \approx \frac{E_X^{corr} \cdot X^3 - E_Y^{corr} \cdot Y^3}{X^3 - Y^3}$$

from the QZ and 5Z results of the respective basis set series.

The results for the cc-pwCV*n*Z and cc-pwCV*n*Z-PP basis sets are listed in Tables 3 – 7. In Tables 3 and 4 we have included for comparison the results for the cc-pwCV*n*Z basis sets for the lighter main group elements B–Ne and Al–Ar. With the weighted core-valence sets cc-pwCV*n*Z the relative one-electron basis set errors in the correlation energies for Ga–Kr are still about 50% larger than for the lighter main group elements with the same basis set families, but converge much more smoothly to the extrapolated basis set limits as with just the valence basis sets cc-pV*n*Z, where the convergence is much slower[26]. The on average 50% larger one-electron basis set errors for the post-*d* elements have their origin in the fact that the core correlation functions of the cc-pwCV*n*Z sets are optimized for the core-valence correlation energy without the core-core correlation contribution while we evaluated the errors in the total (valence plus core) correlation energy. We did not exclude in test calculations the $3s3p3d$ intra-shell correlation energy of Ga-Kr from the error analysis to check whether the new auxiliary basis sets for these atoms have the flexibility to approximate also accurately integrals $\langle ab||ij\rangle$ with both i and j from a $3s3p3d$ shell of a post-*d* element.

The auxiliary basis set errors, also known as resolution-of-the-identity (RI) or density fitting (DF) errors, are with the cc-pwCV*n*Z basis sets for the atoms Ga-Kr between three and four orders of magnitude smaller than the one-electron basis set errors, similar as for the auxiliary basis sets reported in our earlier work. This high accuracy guarantees that the RI errors are also negligible for energy differences as e.g. reaction or excitation energies or properties or for other two-electron integrals than the electron repulsion integrals for which they have been optimised, which is important when they are applied e.g. for CC2 response calculations or in the explicitly correlated R12 and F12 methods.

The results for the pseudo-potential basis set families cc-pwCV*n*Z-PP for the row 4, 5 and 6 post-*d* elements Ga–Kr, In–Xe, and Tl–Rn (listed in Tables 5 and 6) parallel very much the performance of the all-electron cc-pwCV*n*Z basis sets for Ga-Kr. Again, we included the $(n-1)s(n-1)p$ shells in the

Table 3 One-electron basis set errors in MP2 valence plus core correlation energies with the cc-pwCVnZ basis sets. cc-pwCVnZ orbital basis sets are used for all atoms apart form H, where the cc-pVnZ basis sets are used. In these calculations only the electrons in the inner-core orbitals ($1s^2$ for S, Cl and $1s^2 2s^2 2p^6$ for Ga–Kr) were kept frozen. Listed are the mean deviation $\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Delta_i$, the standard deviation $\Delta_{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Delta_i - \bar{\Delta})^2/(n-1)}$ and the maximum deviation $|\Delta|_{max} = max_i|\Delta_i|$, with $\Delta_i = (E^{corr} - E^{corr}_{lim})/E^{corr}_{lim} \times 100\%$. For comparison also the results for the atoms B–Ne, Al–Ar, taken over the test sets used in Ref. 25 (without H_2 and He_2), are given.

	D	T	Q	5
Ga–Kr				
$\bar{\Delta}$	-54.22	-23.84	-10.75	-5.50
Δ_{std}	3.91	4.01	1.29	0.66
$\|\Delta\|_{max}$	60.58	41.79	15.10	7.73
B–Ne, Al–Ar				
$\bar{\Delta}$	-38.28	-14.56	-6.44	-3.30
Δ_{std}	3.71	1.83	0.88	0.45
$\|\Delta\|_{max}$	54.16	21.95	10.02	5.13

Table 4 Relative RI errors in the MP2 valence plus core correlation energies with the (auxiliary) cc-pwCVnZ basis sets. The auxiliary basis sets for the atoms H–Ar have been taken from Refs. 25,26. For further explanations see Tab. 3 The relative RI errors have been calculated as $\Delta_i = (E^{corr}_{RI} - E^{corr})/E^{corr}_{lim} \times 100\%$. In addition to $\bar{\Delta}$, Δ_{std}, and $|\Delta|_{max}$, also the mean absolute errors $\Delta_{mad} = \frac{1}{n}\sum_{i=1}^{n}|\Delta_i|$ are listed.

	D	T	Q	5
Ga–Kr				
$\bar{\Delta}$	0.0025	0.0010	0.0003	-0.0002
Δ_{std}	0.0028	0.0013	0.0006	0.0002
Δ_{mad}	0.0031	0.0013	0.0005	0.0002
$\|\Delta\|_{max}$	0.0096	0.0041	0.0015	0.0006
B–Ne, Al–Ar				
$\bar{\Delta}$	0.0020	-0.0012	-0.0022	-0.0015
Δ_{std}	0.0115	0.0047	0.0022	0.0009
Δ_{mad}	0.0099	0.0046	0.0025	0.0015
$\|\Delta\|_{max}$	0.0269	0.0182	0.0092	0.0046

Table 5 One-electron basis set errors (in %) in MP2 valence plus core correlation energies with the cc-pwCVnZ-PP basis sets for post-d elements. The cc-pwCVnZ orbital basis sets are used O, F, S, and Cl, and the cc-pVnZ basis sets for H. In addition to the electrons replaced by ECPs, only the $1s^2$ cores at S and Cl were kept frozen.

	D	T	Q	5
Ga–Kr				
$\bar{\Delta}$	-54.75	-23.36	-10.71	-5.48
Δ_{std}	4.05	2.09	1.03	0.53
$\|\Delta\|_{max}$	62.32	26.31	12.24	6.27
In–Xe				
$\bar{\Delta}$	-52.61	-23.23	-10.97	-5.62
Δ_{std}	4.87	2.75	1.35	0.69
$\|\Delta\|_{max}$	61.15	27.72	13.06	6.69
Tl–Rn				
$\bar{\Delta}$	-49.91	-22.65	-10.79	-5.52
Δ_{std}	4.05	2.47	1.29	0.66
$\|\Delta\|_{max}$	56.65	26.06	12.77	6.54

correlation treatment when we evaluated the one-electron and auxiliary basis errors for the MP2 correlation energies in order to check that the cc-pwCVXZ-PP auxiliary basis sets are flexible enough to approximate also integrals which involve these sub-valence shells. We also evaluated the RI errors for the augmented aug-cc-pwCVnZ-PP basis sets (see Tab. 6). They have for all three rows Ga–Kr, In–Xe, and Tl–Rn a similar magnitude as the RI errors for the corresponding non-augmented basis sets.

As pointed out in the introduction, the pseudo-potential valence cc-pVnZ-PP basis sets are usually only used as part of the core-valence (aug-)cc-pwCVnZ-PP sets, since they are, similar as the all-electron valence basis sets cc-pVnZ, for post-d elements not well suited for situations where the sub-valence $(n-1)d$ orbitals contribute significantly to binding energies, which is the case for several of the molecules in our test sets. Therefore, we evaluated for the cc-pVnZ-PP basis sets no one-electron basis set limits and basis errors but only the RI errors which are given in Tab. 7. They are for all cardinal numbers of similar magnitude as the RI errors for the all-electron cc-pVnZ basis sets for the main group elements B–Ne, Al–Ar, and Ga–Kr.

To demonstrate how the RI approximation in combination with the new auxiliary basis sets extends the applicability of MP2 and related wavefunctions methods for compounds with post-d elements, we studied four of the $(GaAs)_n$ clusters proposed in Ref. 55 shown in Fig. 1. The structures of the clusters have been optimised at the RI-DFT/PBE level with TZVP basis sets.[2,30] We then carried out single-point RI-MP2 and MP2 calculations with the all-electron triple-ζ weighted core-

This journal is © The Royal Society of Chemistry [year]

Table 6 Relative RI errors in the MP2 valence plus core correlation energies with the (auxiliary) cc-pwCV*n*Z-PP basis sets for the post-*d* elements. In parentheses are the results for the aug-cc-pwCV*n*Z-PP basis sets given. For further explanations see caption of Tab. 5.

	D	T	Q	5
Ga–Kr				
$\bar{\Delta}$	0.0023	0.0016	0.0004	-0.0001
	(0.0023)	(0.0004)	(0.0001)	(-0.0002)
Δ_{std}	0.0026	0.0017	0.0005	0.0002
	(0.0021)	(0.0007)	(0.0002)	(0.0002)
Δ_{mad}	0.0028	0.0019	0.0006	0.0002
	(0.0025)	(0.0007)	(0.0002)	(0.0002)
$\lvert\Delta\rvert_{max}$	0.0089	0.0062	0.0014	0.0006
	(0.0077)	(0.0018)	(0.0007)	(0.0007)
In–Xe				
$\bar{\Delta}$	0.0010	0.0020	0.0000	-0.0001
	(0.0020)	(0.0006)	(-0.0004)	(-0.0003)
Δ_{std}	0.0028	0.0021	0.0009	0.0003
	(0.0025)	(0.0008)	(0.0003)	(0.0002)
Δ_{mad}	0.0026	0.0025	0.0007	0.0002
	(0.0027)	(0.0007)	(0.0004)	(0.0003)
$\lvert\Delta\rvert_{max}$	0.0056	0.0062	0.0019	0.0006
	(0.0070)	(0.0021)	(0.0012)	(0.0007)
Tl–Rn				
$\bar{\Delta}$	0.0014	0.0017	0.0002	-0.0002
	(0.0025)	(0.0006)	(0.0000)	(-0.0004)
Δ_{std}	0.0020	0.0013	0.0005	0.0003
	(0.0016)	(0.0006)	(0.0002)	(0.0003)
Δ_{mad}	0.0022	0.0020	0.0004	0.0003
	(0.0025)	(0.0007)	(0.0002)	(0.0004)
$\lvert\Delta\rvert_{max}$	0.0048	0.0036	0.0012	0.0010
	(0.0053)	(0.0016)	(0.0007)	(0.0007)

Table 7 Relative RI errors in the MP2 valence plus outer-core correlation energies with the (auxiliary) cc-pV*n*Z-PP basis sets for the post-*d* elements. In the test calculations electrons in the 1*s* core orbitals of O and F, and 1*s*2*s*2*p* core orbitals of S and Cl, and the $(n-1)s(n-1)p$ core orbitals of the post-*d* elements were kept frozen.

	D	T	Q	5
Ga–Kr				
$\bar{\Delta}$	0.0062	0.0026	0.0000	-0.0001
Δ_{std}	0.0077	0.0040	0.0013	0.0005
Δ_{mad}	0.0082	0.0040	0.0010	0.0004
$\lvert\Delta\rvert_{max}$	0.0210	0.0107	0.0031	0.0011
In–Xe				
$\bar{\Delta}$	0.0039	0.0047	0.0008	0.0002
Δ_{std}	0.0101	0.0047	0.0018	0.0007
Δ_{mad}	0.0096	0.0059	0.0018	0.0006
$\lvert\Delta\rvert_{max}$	0.0215	0.0123	0.0034	0.0014
Tl–Rn				
$\bar{\Delta}$	0.0030	0.0027	0.0008	0.0006
Δ_{std}	0.0071	0.0030	0.0011	0.0005
Δ_{mad}	0.0064	0.0037	0.0013	0.0007
$\lvert\Delta\rvert_{max}$	0.0151	0.0081	0.0023	0.0012

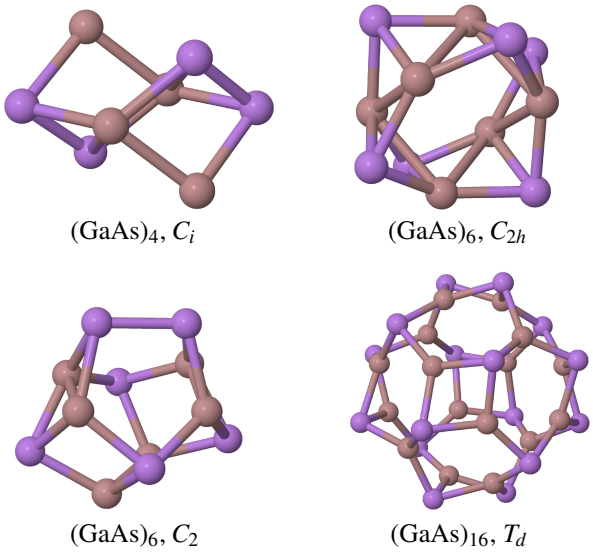


Fig. 1 Structures of the four gallium arsenide cluster $(GaAs)_4$ (C_i), $(GaAs)_6$ (C_{2h}), $(GaAs)_6$ (C_2), and $(GaAs)_{16}$ (T_d).

Table 8 MPb_{12}^+ cationic clusters. Wall clock times on an Intel Xeon E5430 2.66 GHz CPU for MP2 energy and gradient calculations with (ricc2) and without RI (mpgrad) approximation and results for equilibrium bond distances. Numbers in parenthesis are relative RI errors. The B3-LYP results have been taken from Ref. 56 and were obtained with the cc-pVDZ basis for Al and cc-pVDZ-PP basis sets for all other atoms.

	MP2 energy, wall clock time in min and errors		
	DZ	TZ	QZ
$AlPb_{12}^+$	1.0/5.1 (0.06‰)	4.6/109 (0.05‰)	21.4/1744
$GaPb_{12}^+$	1.2/6.8 (0.05‰)	6.0/224 (0.04‰)	47.5/—
$InPb_{12}^+$	1.2/6.8 (0.05‰)	5.9/260 (0.04‰)	41.5/—
$TlPb_{12}^+$	1.2/6.8 (0.05‰)	5.9/258 (0.04‰)	42.8/—
	MP2 gradient, wall clock time in min		
	DZ	TZ	
$AlPb_{12}^+$	5.7/35.9	41.8/—	
$GaPb_{12}^+$	9.7/41.3	72.3/—	
$InPb_{12}^+$	9.4/43.7	70.5/—	
$TlPb_{12}^+$	9.6/44.2	74.8/—	
	M-Pb and Pb-Pb bond distances in Å		
	MP2/DZ	MP2/TZ	B3-LYP/DZ
$AlPb_{12}^+$	3.11/3.27	3.06/3.22	3.19/3.36
$GaPb_{12}^+$	3.10/3.26	3.05/3.21	3.19/3.36
$InPb_{12}^+$	3.14/3.30	3.09/3.25	3.24/3.40
$TlPb_{12}^+$	3.15/3.31	3.10/3.26	3.25/3.42

valence basis sets cc-pwCVTZ. The wall clock timings on a AMD Opteron 152 CPU with TURBOMOLE V6.2 are for RI-MP2 calculations (using the ricc2 program): 9, 24, and 62 min for, respectively, $(GaAs)_4$, $(GaAs)_6$ (C_{2h}), and $(GaAs)_6$ (C_2). The conventional MP2 calculations (using mpgrad) took 1244, 3733, and 14814 min, respectively. This corresponds to a speed-up between 135 and 240. The calculation for $(GaAs)_{16}$ could only with RI approximation be done sequentially and took 55.2 h. Without RI approximation it took about 14.5 days on eight Intel Xeon E5345 2.33 GHz cores. The errors in the MP2 correlation energies due to the RI approximation have been 0.0006% for $(GaAs)_4$, 0.0001% for $(GaAs)_6$ (C_{2h}), 0.0002% for $(GaAs)_6$ (C_{2h}), and 0.0001% for $(GaAs)_{16}$, which is well within the mean absolute devation for the test set in Tab. 4 with the cc-pwCVTZ basis.

As example for the application of cc-pwCV*n*Z-PP auxiliary basis sets we studied the MPb_{12}^+ cationic clusters with M = Al, Ga, In, and Tl. Previous experimental[57] and theoretical studies[56,57] at the DFT level have shown that these clusters are remarkably stable with icosahedral ground state geometry. In our calculations we correlated the outer-valence electrons and for Pb, Ga, In, and Tl also the sub-valence $(n-1)d$ shell electrons. However, the timings for gradient calculations with mpgrad were, due to technical limitations of the code, obtained without frozen-core approximation. The reported MP2 geometries have been obtained with RI and frozen-core approximations. The results are collected in Table 8. Already at the double-ζ level RI approximation speeds the calculation up by almost a factor of five. With the triple-ζ basis sets the speed-up is already 22–44 and with the quadruple-ζ basis set for $AlPb_{12}^+$ more than 80. The large computational savings obtained from the RI approximation allowed to optimise geometries of the cationic clusters at the MP2/cc-pwCVDZ-PP and MP2/cc-pwCVTZ-PP levels within about half a day of CPU time for each compound even without employing any parallelisation (note that for cc-pwCVDZ-PP the calculations comprised 1040 basis functions). The bond lengths obtained at the MP2/cc-pwCVDZ-PP level are about 10 pm shorter than obtained in Ref. 56 with B3-LYP/cc-pVDZ. With the cc-pwCVTZ-PP basis sets the bond lengths shrink further by about 5 pm. We attribute this effect mainly to the dispersive interaction between the $5d^{10}$ sub-valence shells of the Pb atoms, which is not accounted for with the hybrid density functional B3-LYP. The same effect is most likely responsible for the contraction of the bond lengths by 1 pm from M = Al, which has no sub-valence *d*-shell, to $GaPb_{12}^+$, which is also not observed with B3-LYP.

4 Summary and Conclusions

We optimised and evaluated auxiliary basis sets for the post-*d* elements Ga–Kr, In–Xe, and Tl–Rn for use in density-fitted MP2, CC2, MP2-F12 and related correlated calculations with the correlation-consistent weighted core-valence and ECP basis set families (aug-)cc-pwCV*n*Z, (aug-)cc-pV*n*Z-PP, and (aug-)cc-pwCV*n*Z-PP with *n* = D, T, Q, and 5. With these auxiliary basis sets the RI (or density-fitting) errors are at between three and four orders of magnitude smaller than the remaining one-electron basis set errors in the correlation energy and thus completely negligible for any pratical purposes. Due to the large size of weighted core-valence one-electron basis sets, the RI approximation reduces already for the cc-pwCVDZ-PP basis sets the wall clock time for the calculations by about a factor of five and this factor increases rapidly with the cardinal number in the series DZ, TZ, QZ, and 5Z. For the all-electron basis sets cc-pwCV*n*Z for Ga–Kr the computational savings are even larger. The new auxiliary basis sets allow to apply the efficient implementations of MP2, CC2, L-DF-MP2, MP2-F12, L-PNO-CCSD, and others, that have been developed during the last fifteen years, now also for calculations on molecules with post-*d* elements with accurate account of correlation effects from the $(n-1)d$-shell electrons.

References

1 E. J. Baerends, D. E. Ellis and P. Ros, *Chem. Phys.*, 1973, **2**, 41–51.

This journal is © The Royal Society of Chemistry [year]

2 K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **242**, 652–660.
3 K. R. Glaesemann and M. S. Gordon, *J. Chem. Phys.*, 2000, **112**, 10738–10745.
4 F. Weigend and M. Häser, *Theor. Chem. Acc.*, 1997, **97**, 331–340.
5 C. Hättig, A. Hellweg and A. Köhn, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1159–1169.
6 *Q-CHEM, A Quantum Leap Into the Future of Chemistry, see http://www.q-chem.com*.
7 *ORCA, An ab initio, DFT and semiempirical SCF-MO package, see http://www.thch.uni-bonn.de/tc/orca/*.
8 *NWChem, High-Performance Computation Chemistry Software, see http://www.nwchem-sw.org*.
9 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz *et al.*, *MOLPRO, version 2010.1, a package of ab initio programs*, 2010, see http://www.molpro.net.
10 *TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.*
11 C. Hättig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154–5161.
12 D. Kats, T. Korona and M. Schütz, *J. Chem. Phys.*, 2006, **125**, 104106.
13 M. Head-Gordon, R. J. Rico, M. Oumi and T. J. Lee, *cpl*, 1994, **219**, 21–29.
14 A. Hellweg, S. Grün and C. Hättig, *Phys. Chem. Chem. Phys.*, 2008, **10**, 1159–1169.
15 Y. Jung, R. C. Lochan, A. D. Dutoi and M. Head-Gordon, *The Journal of Chemical Physics*, 2004, **121**, 9793–9802.
16 N. O. C. Winter and C. Hättig, *J. Chem. Phys.*, 2011, **134**, 184101.
17 F. R. Manby, *J. Chem. Phys.*, 2003, **119**, 4607–4613.
18 C. Hättig, W. Klopper, A. Köhn and D. P. Tew, *Chem. Rev.*, 2012, **112**, 4–74.
19 R. A. Bachorz, F. A. Bischoff, A. Glöß, C. Hättig, S. Höfener, W. Klopper and D. P. Tew, *J. Comput. Chem.*, 2011, **32**, 2492–2513.
20 H. Koch, A. S. de Meras and T. B. Pedersen, *J. Chem. Phys.*, 2003, **118**, 9481–9484.
21 H.-J. Werner, F. R. Manby and P. J. Knowles, *J. Chem. Phys.*, 2003, **118**, 8149–8160.
22 F. Neese, F. Wennmohs and A. Hansen, *jcp*, 2009, **130**, 114108.
23 F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Letters*, 1998, **294**, 143–152.
24 A. Hellweg, C. Hättig, S. Höfener and W. Klopper, *Theor. Chem. Acc.*, 2007, **117**, 587–597.
25 C. Hättig, *Phys. Chem. Chem. Phys.*, 2005, **7**, 59–66.
26 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2001, **116**, 3175–3183.
27 F. Weigend, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4285–4291.
28 *Basis Set Exchange: A Community Database for Computational Science, see https://bse.pnl.gov/bse/portal*.
29 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.
30 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829–5835.
31 K. Eichkorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119–124.
32 F. Weigend, F. Furche and R. Ahlrichs, *J. Chem. Phys.*, 2003, **119**, 12753–12762.
33 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
34 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
35 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
36 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1993, **98**, 1358–1371.
37 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1994, **100**, 2975–2988.
38 A. K. Wilson, D. E. Woon, K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 1999, **110**, 7667–7676.
39 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1995, **103**, 4572–4585.
40 A. K. Wilson, T. van Mourik and T. H. Dunning, *J. Mol. Struct. (THEOCHEM)*, 1996, **308**, 339–349.
41 T. H. Dunning, K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244–9253.
42 T. van Mourik and T. H. Dunning, *Int. J. Quantum Chem.*, 2000, **76**, 205–221.
43 K. A. Peterson, *J. Chem. Phys.*, 2003, **119**, 11099.
44 K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2003, **119**, 11113.
45 K. A. Peterson and K. E. Yousaf, *J. Chem. Phys.*, 2010, **133**, 174116.
46 K. A. Peterson and C. Puzzarini, *Theor. Chem. Acc.*, 2005, **114**, 283.
47 K. A. Peterson, D. Figgen, M. Dolg and H. Stoll, *J. Chem. Phys.*, 2007, **126**, 124101.
48 K. A. Peterson, M. Dolg and H. Stoll, *J. Chem. Phys.*, 2009, **130**, 164108.
49 K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548–10560.
50 K. A. Peterson, B. C. Shepler, D. Figgen and H. Stoll, *J. Phys. Chem. A*, 2006, **110**, 13877.
51 J. G. Hill and J. A. Platts, *J. Chem. Phys.*, 2008, **128**, 044104.
52 J. G. Hill and J. A. Platts, *J. Chem. Theory Comput.*, 2009, **5**, 500.
53 J. G. Hill and J. A. Platts, *J. Chem. Phys.*, 2009, **129**, 134101.
54 T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639.
55 G. L. Gutsev, R. H. O'Neal, Jr., B. C. Saha, M. D. Mochene, E. Johnson and C. W. Bauschlicher, Jr., *J. Phys. Chem. A*, 2008, **112**, 10728–10735.
56 D.-L. Chen, W. Q. Tian, W.-C. Lu and C.-C. Sun, *J. Chem. Phys.*, 2006, **124**, 154313.
57 S. Neukermans, E. Janssens, Z. F. Chen, R. E. Silverans, P. v. R. Schleyer and P. Lievens, *Phys. Rev. Lett.*, 2004, **92**, 163401.