On the resolution of identity Coulomb energy approximation in density functional theory[☆]

C.-K. Skylaris[*], L. Gagliardi, N.C. Handy, A.G. Ioannou, S. Spencer, A. Willetts

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Abstract

The Resolution of the Identity approximation for the Coulomb (RI-J) energy in Density Functional Theory improves the computational efficiency of large-scale calculations but requires the use of a second, or "auxiliary" basis set. We examine the performance of some of the existing auxiliary basis sets with a variety of basis sets and molecules. We determine the accuracy of the RI-J approximation for these basis sets and suggest criteria for the selection of combinations of basis set and auxiliary basis set. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: RI-J; DFT; Coulomb; Auxiliary

1. Introduction

1.1. Density functional theory for large systems

Density Functional Theory (DFT) has been established as a practical method for including electron correlation effects in calculations on chemically important molecular systems. While research into finding improved functionals continues, there are a large number of reasonably accurate functionals available. Many studies of chemical properties have been carried out using those functionals. To this end, an efficient computer program is essential, as the amount of computation necessary is large. In the regime of the size of molecules which are of practical interest to many experimental chemists, the number of integrals necessary for a DFT calculation is simply too large to store in core memory or on disk [1]. *Direct SCF* algorithms, first introduced by Almlöf et al. [2], are the only possibility. This breakthrough in computational chemistry was made possible by reducing the prohibitive and unnatural cost of such a calculation, which appeared to be proportional to N_{BF}^4, where N_{BF} is the number of basis functions, a quantity proportional to the size of the molecular system. In the Direct methodology, the integrals are calculated afresh in every SCF cycle and tests are used to skip the evaluation of integrals, which are predicted to make zero contribution to the Fock matrix. As a result, the cost of the calculation scales quadratically for sufficiently large systems. Refinements to this procedure were developed by Ahlrichs and coworkers [3,4]. The direct SCF developments were originally applied in Hartree–Fock calculations. They are however readily transferable to Kohn–Sham theory and in fact easier to implement due to the lack of exchange integrals.

In Kohn–Sham theory, the contributions to the electronic interaction energy are expressed as functionals of the electronic density. These are the exchange correlation term E_{XC} and the Coulomb

---

term E_J

$$E_{XC}[\rho] = \int f(\rho, \nabla\rho, \dots)dv \qquad (1)$$

$$E_J[\rho] = \frac{1}{2} \int \int \rho(\mathbf{r}_1) \frac{1}{r_{12}} \rho(\mathbf{r}_2) dv_1 \, dv_2 \qquad (2)$$

The evaluation of E_{XC} is typically carried out with numerical integration. Its computational cost is a small percent of the whole cost of a DFT calculation for medium sized molecules ($\sim$50 atoms) and it has been shown that it can be made to scale linearly with N_{BF} [5]. It is the evaluation of the Coulomb term which dominates the computational procedure as soon as molecules of reasonable size are encountered. Methods for calculating this term, which scale linearly asymptotically, have been invented [6,7]. However, their breakeven point with conventional methods has been demonstrated to be of the order of hundreds of basis functions and this for cases of molecules of light atoms with extended geometries, such as linear hydrocarbons, graphite sheets and water clusters. It is therefore reasonable to expect that in large complexes containing heavy atoms or three dimensional metal clusters the breakeven point will be located at a prohibitively large number of basis functions. This renders current linear scaling methods unsuitable for practical, every day calculations on such compounds. The complexity of the implementation of such methods, without any optimum "recipe" available in the literature, is another drawback. Quantum chemistry programs such as MAGIC [8] are mainly intended for the study of inorganic compounds containing heavy elements [9]. Methods of reducing the cost of the calculation by having a small prefactor though they still scale quadratically would lead to significant benefits in this case.

1.2. Representation of the density by an auxiliary basis set

A quadratically scaling method for calculating the Coulomb energy, which is faster than the conventional approach, is the fitted density approximation. It was used in different forms in the early work of Baerends [10] and Dunlap [11]. It is based on the observation that the basis we use to describe the molecular orbitals is more flexible than required to describe the density. This can be more clearly understood if we consider that the basis set is required to describe all molecular orbitals. These are usually quite different from each other and contain features not present in the density (e.g. sign changes, different levels of localisation). If $\{\chi_i\}_{i=1}^{N_{BF}}$ is the set of basis functions we use to express the orbitals, the density is represented as a quadratic form in the basis in terms on the density matrix $\mathbf{D}$:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_{BF}} \sum_{j=1}^{N_{BF}} \chi_i(\mathbf{r})\chi_j(\mathbf{r})D_{ji} \qquad (3)$$

When the basis consists of Gaussian functions, then according to the product rule, $\chi_i(\mathbf{r})\chi_j(\mathbf{r}) = \chi_t(\mathbf{r})$ where $\chi_t(\mathbf{r})$ is a linear combination of a few Gaussian functions, centred on the line joining the centres of $\chi_i(\mathbf{r})$ and $\chi_j(\mathbf{r})$ [12]. Consequently, the above expression for the density can equivalently be written as

$$\rho(\mathbf{r}) = \sum_{t=1}^{N_{BF}(N_{BF}+1)/2} \chi_t(\mathbf{r})q_t' \qquad (4)$$

where $q'_{i(i-1)/2+j} = (2 - \delta_{ij})D_{ij}$. As the space of the $\chi_t(\mathbf{r})$ functions is nearly degenerate [13,14], it is desirable to obtain an accurate representation of the density in terms of a new basis set, especially tailored for this purpose. This new basis set $\{f_t\}_{t=1}^{N_{AUX}}$ is commonly called the auxiliary basis set. The density is approximated by the auxiliary basis set as

$$\rho(\mathbf{r}) \simeq \rho_{AUX}(\mathbf{r}) = \sum_{t=1}^{N_{AUX}} f_t(\mathbf{r})q_t \qquad (5)$$

where the expansion coefficients, or "charges" q_t [15,16] are determined in some optimum way. The approximate density is subsequently used to calculate the Coulomb energy

$$E_J[\rho] \simeq \frac{1}{2} \int \int \rho_{AUX}(\mathbf{r}_1) \frac{1}{r_{12}} \rho_{AUX}(\mathbf{r}_2)dv_1 \, dv_2 \qquad (6)$$

The above result would equal the exact Coulomb energy only if the auxiliary set was a complete basis with an infinite number of functions or the set of $N_{BF}(N_{BF}+1)/2$ functions arising from all the distinct products of pairs of the original basis functions. None of these choices would result in any computational savings, however. To have any significant savings, N_{AUX} has to be proportional to N_{BF}. Indeed, Eichkorn

et al. [17,18] have constructed auxiliary basis sets for every element (excluding the lanthanides and the actinides) which satisfy the requirement

$$N_{AUX} < 3N_{BF} \tag{7}$$

and it is claimed that this leads to an increase of the speed of calculation of E_J by an order of magnitude. The absolute error introduced by these auxiliary sets is expected to be smaller than the errors arising from the incompleteness of the basis sets and the inexact treatment of correlation. In fact, the relative error, which is important for chemical properties, is expected to be even less and should not hinder attempts to improve our results in an effort to achieve "chemical accuracy" of 1 mE$_h$ [19].

The MAGIC quantum chemistry program, intended for performing relativistic calculations on large systems containing heavy atoms [20,21], is able to calculate the Coulomb energy either by an auxiliary basis fitted density, or with the "exact" density. In agreement with the above, we have observed that the fitting density approximation is markedly faster than the exact calculation. We find however that there are very few auxiliary basis sets available in the literature [17,18,22]; and the method's accuracy has not been tested extensively. In this paper we are attempting to test the robustness of some of the most widely used auxiliary basis sets with respect to the size of the basis set. We wish to examine their performance for smaller and larger basis sets than the ones they were meant to be used with and see if we can establish criteria for judging the quality of the representation of the fitted density without comparing with the "exact" calculation. In general, every basis set is not accompanied by a corresponding auxiliary set. Therefore, though the auxiliary fit method is efficient, it may not be of much practical value without general-purpose auxiliary sets, accurate for a range of basis sets.

In Section 2 we give a brief comparison of density fitting procedures and the underlying theory including the procedure implemented in MAGIC. In Section 3 we describe the calculations we performed in order to answer our question. Section 4 is a discussion and rationalisation of our results. Finally, in Section 5 we reach some conclusions regarding the transferability of auxiliary basis sets and suggest ways of predicting how well they will perform with a particular basis set and molecule.

2. Density fitting procedures

We summarise briefly and compare the most common density fitting procedures in the literature. We follow the notation of earlier works in the field. Both the basis set and the auxiliary basis set consist of non orthogonal functions. The first step in fitting procedures is to orthonormalise the auxiliary basis set. Symmetric orthonormalisation is used for this purpose. This leads to two broad classes depending on the kind of metric used.

2.1. Inner product: overlap integral

The overlap integral

$$S_{ts} = \int f_t^*(1) f_s(1) \, dv_1 = \langle f_t | f_s \rangle \tag{8}$$

is used as a metric. Then symmetric orthonormalisation

$$\mathbf{f}' = \mathbf{f} \mathbf{S}^{-1/2} \tag{9}$$

yields the auxiliary basis set $\{f_t'\}_{t=1}^{N_{AUX}}$ which is orthonormal in the linear vector space with metric $\mathbf{S}$. The density can be fitted now by simply projecting it onto the orthonormal auxiliary basis.

$$\rho \simeq \tilde{\rho} = \mathbf{f}' \langle \mathbf{f}' | \rho \rangle = \mathbf{f} \mathbf{S}^{-1/2} \langle \mathbf{f} \mathbf{S}^{-1/2} | \rho \rangle = \mathbf{f} \mathbf{S}^{-1} \langle \mathbf{f} | \rho \rangle = \mathbf{f} \tilde{\mathbf{q}} \tag{10}$$

The Coulomb energy can therefore be represented by

$$2E_J \simeq (\tilde{\rho} | \tilde{\rho}) = (\mathbf{f} \mathbf{S}^{-1} \langle \mathbf{f} | \rho \rangle | \mathbf{f} \mathbf{S}^{-1} \langle \mathbf{f} | \rho \rangle)$$

$$= \langle \rho | \mathbf{f} \rangle \mathbf{S}^{-1} (\mathbf{f} | \mathbf{f}) \mathbf{S}^{-1} \langle \mathbf{f} | \rho \rangle = \langle \rho | \mathbf{f} \rangle \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1} \langle \mathbf{f} | \rho \rangle \tag{11}$$

where $\mathbf{V}$ is the matrix of two centre two electron integrals in the auxiliary basis, see Eq. (14) below. Eq. (11) was originally derived by minimising the mean square deviation of the fitted density [10]:

$$\int |\rho(\mathbf{r}) - \tilde{\rho}(\mathbf{r})|^2 dv \tag{12}$$

2.2. Inner product: electrostatic interaction integral

The electrostatic interaction integral

$$V_{ts} = \int \int f_t^*(1) \frac{1}{r_{12}} f_s(2) dv_1 \; dv_2 = (f_t|f_s) \qquad (13)$$

can serve as another definition of inner product and be used as a metric. Symmetric orthonormalisation in a similar fashion as before

$$\mathbf{f}'' = \mathbf{f}\mathbf{V}^{-1/2} \qquad (14)$$

yields the auxiliary basis set $\{f_t''\}_{t=1}^{N_{AUX}}$ which is orthonormal in the linear vector space with metric $\mathbf{V}$. The density is again expanded in the orthonormal auxiliary basis

$$\tilde{\rho} = \mathbf{f}''(\mathbf{f}''|\rho) = \mathbf{f}\mathbf{V}^{-1/2}(\mathbf{f}\mathbf{V}^{-1/2}|\rho) = \mathbf{f}\mathbf{V}^{-1}(\mathbf{f}|\rho) = \mathbf{f}\tilde{\mathbf{q}} \qquad (15)$$

The expression for the Coulomb energy in this case is

$$2E_J \simeq (\tilde{\rho}|\tilde{\rho}) = (\mathbf{f}\mathbf{V}^{-1}(\mathbf{f}|\rho)|\mathbf{f}\mathbf{V}^{-1}(\mathbf{f}|\rho))$$

$$= (\rho|\mathbf{f})\mathbf{V}^{-1}(\mathbf{f}|\mathbf{f})\mathbf{V}^{-1}(\mathbf{f}|\rho) = (\rho|\mathbf{f})\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}(\mathbf{f}|\rho)$$

$$= (\rho|\mathbf{f})\mathbf{V}^{-1}(\mathbf{f}|\rho) \qquad (16)$$

By noting that $(\rho|\tilde{\rho}) = (\tilde{\rho}|\tilde{\rho})$, a variational principle can be written, which shows that the error in Coulomb energy due to auxiliary basis set incompleteness is negative

$$2E_J \geq (\tilde{\rho}|\tilde{\rho}) \qquad (17)$$

The result of Eq. (16) was originally derived by minimising the Coulomb energy of the residual of the fitted density [11]:

$$\frac{1}{2} \int \int \frac{(\rho(\mathbf{r}_1) - \tilde{\rho}(\mathbf{r}_1))(\rho(\mathbf{r}_2) - \tilde{\rho}(\mathbf{r}_2))}{|\mathbf{r}_1 - \mathbf{r}_2|} dv_1 \; dv_2 \qquad (18)$$

2.3. Combination of the two methods

If the two ways of fitting the density are combined the following expression results.

$$2E_J \simeq (\tilde{\rho}|\tilde{\rho}) = \langle\rho|\mathbf{f}\rangle\mathbf{S}^{-1}(\mathbf{f}|\rho) \qquad (19)$$

As will be seen in Section 2.4, this expression is a compromise in computational cost between Eqs. (11) and (16). When used in a direct SCF scheme it requires the evaluation of three-index two-electron integrals and three-index overlap integrals in every SCF cycle.

2.4. Comparison of the three approximations

In order to use Eq. (11) in a direct SCF calculation, one would have first to calculate the $\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}\langle\mathbf{f}|\rho\rangle$ column vector and then contract it with the $\langle\rho|\mathbf{f}\rangle$ row. These two steps require the calculation of three-index overlap integrals twice per SCF cycle. In a similar way we can show that for Eq. (16) it is necessary to calculate three-index two-electron integrals twice per SCF cycle. The three schemes therefore have substantially different computational costs. They also differ substantially in their performance in fitting the density. By intuition, we would expect the use of the electron repulsion integral as a metric to yield better results since the reason we are fitting the density is to approximate such an integral. Or in other words, minimising directly the Coulomb energy of the density residual should lead to better results as far as energy is concerned. This assumption was verified by Vahtras et al. [23] through a series of simple tests. They found that for every auxiliary basis set they tried, Eq. (16) (the "V approximation") was far more accurate than Eq. (11) (the "SVS approximation") while the results owing to Eq. (19) (the "S approximation") were in between. The variational principle of Eq. (17) is an added benefit of the V approximation. It applies strictly for ρ being the density of the RI-J calculation. However under the assumption that the density $\rho_{\text{"exact"}}$ of the non fitted calculation is almost equal to ρ one can deduce that the RI-J error in energy should be negative. When this subtle generalisation is valid, it leads to some cancellation of systematic errors in energy differences. As a consequence, the V approximation has become the most widely used today. It is also referred to as the RI-J approximation (Resolution of the Identity for the Coulomb energy). Some results of accuracy as a function of auxiliary basis set size have been given by Van Alsenoy [24] for the S approximation with the added constraints of preserving charge and dipole moment. However, the auxiliary basis sets used were far from optimised and explicit corrections to the Coulomb energy in terms of four-index electron repulsion integrals had to be included. (We should note at this point that in the case of hybrid

exchange-correlation density functionals, the auxiliary basis set approximation is not as efficient since four index integrals are necessary for the exchange energy and have to be calculated as well.)

MAGIC approximates the Coulomb energy using the V approximation. We chose this method because of its accuracy. Furthermore, calculation of the gradient or higher derivatives is not much more complicated than the conventional approach [25]. One should be careful, however, when determining the charges $\tilde{\mathbf{q}}$ as the $\mathbf{V}$ matrix can very often be nearly singular. Numerical algorithms stable to contamination from nullspace solutions of $\mathbf{V}$ are readily available [17,26] and should be used when solving the system of linear equations $V\tilde{\mathbf{q}} = (\mathbf{f}|\rho)$.

3. Calculations and results

The computational efficiency as well as the quality of the V approximation depend on the auxiliary basis. There is little justification for using RI-J DFT unless the auxiliary basis is optimised for both speed and accuracy. This is achieved by requiring the auxiliary sets to represent the density of the isolated atom in such a way as the error that results in the atomic energy is below a certain threshold. It is important to note however that the optimization procedure of the auxiliary basis is carried out for a particular kind of basis. There is therefore no a priori guarantee that they are going to perform equally well for basis sets other than the one for which they were optimised. On the other hand, the construction of an auxiliary basis is not a trivial task and is certainly not one that can be performed on a daily basis for each basis set we encounter. Consequently, the practice of considering the available auxiliary basis sets more or less as "universal" has prevailed. They are thus used with little regard to the basis set selected. We wish to examine the validity of this assumption for a variety of basis sets and molecules. Furthermore, we wish to gain some insight into the quality of the RI-J approximation for a particular auxiliary basis set as a function of the basis set and the type of molecule.

For this study, we decided to use the auxiliary basis sets of Eichkorn et al. [17,18], which were designed specifically for basis sets of SVP and TZVP quality by the same group [27,28]. These auxiliary sets are

available for almost all the elements, apart from the actinides and lanthanides and are meant to produce an error of less than 0.2 mE_h per atom. We decided to treat these auxiliary basis sets as standard "universal" auxiliary basis sets.

We selected six types of basis sets[1] which we describe in order of increasing size. First, we use the STO3G minimal basis set of Hehre et al. [29,30]. Calculations with a minimal basis are essentially semiempirical as a result of its incompleteness and lack of polarization functions. The second basis set that we use is the double zeta basis of Dunning et al. [31,32]. This is certainly an improvement over the STO3G but still lacks polarisation functions. Next in the series come the SVP and TZVP bases of Ahlrichs et al., the ones for which the auxiliary basis sets were designed. These were therefore expected to perform better than the rest of the basis sets. These sets are fully optimised (exponents + contraction coefficients) and describe well, both the core area with large contractions and especially the valence area. The fifth basis set we tried is the cc-pVTZ basis of Dunning et al. [33,34]. This basis, like the TZVP, deals with the core orbitals with large fixed contractions and puts the emphasis on the valence region which is mainly responsible for most molecular properties, especially the ones related to the energetics of bond breaking and bond forming. It is substantially larger than the TZVP and contains polarisation functions with angular momentum up to and including d for hydrogen and f for second and third row atoms. The sixth and final basis that we employed was the cc-pVQZ, again by Dunning et al. [33,34]. This basis further extends cc-pVTZ with its four zeta description of the valence region and is probably sufficient for the most demanding of DFT calculations. It

---

[1] Basis sets were obtained from the Extensible Computational Chemistry Environment Basis set database, version 1.0, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific North-west Laboratory, P.O. Box 999, Richland, Washington 99352, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multi-programme laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DEAC06-76RLO 1830. Contact David Feller, Karen Schuchardt or Don Jones for further information. The database is accessible via the URL http://www.emsl.pnl.gov:2080/forms/basisform.html.

Table 1
Errors in total energy (E) and atomisation energy (E_A) and WRMSD values (Eq. (20)) for the RI-J method. The number of auxiliary basis functions for each molecule is given in parentheses. Also given are the functional and the method for obtaining the geometry (e.g. BLYP//MM means that the BLYP functional was used and the geometry was obtained by molecular mechanics)

Molecule	Basis set	N_{BF}	$E - E_{RI-J}$ (mE$_h$)	$E_{A,RI-J} - E_A$ (mE$_h$)	WRMSD ($\times 10^7$)
RbCl (74) BLYP//MM	STO3G	39	8.110	0.249	3.33
	DZ	48	0.602	0.533	1.09
	SVP	49	0.567	0.546	0.99
	TZVP	53	0.643	0.611	1.01
	cc-pVTZ	69	0.736	0.705	1.01
	cc-pVQZ	104	0.892	0.801	1.27
K[PtCl$_3$(C$_2$H$_4$)] (439) LDAX//LDAX(SVP)	STO3G	106	55.259	1.304	9.83
	DZ	147	31.943	0.034	8.12
	SVP	172	31.753	0.101	8.09
	TZVP	198	31.856	0.186	8.12
	cc-pVTZ	312	32.018	0.342	8.12
	cc-pVQZ	567	32.276	0.281	8.20
CH$_2$FCOO$^-$ (256) BLYP//MM	STO3G	27	2.288	3.567	2.40
	DZ	54	0.301	0.267	1.72
	SVP	85	0.400	0.374	2.08
	TZVP	112	0.464	0.444	2.06
	cc-pVTZ	205	0.484	0.451	2.12
	cc-pVQZ	420	0.522	0.499	2.21
CH$_3$C$_6$H$_5$ (428) LDAX//MM	STO3G	43	2.155	6.308	1.83
	DZ	86	0.255	0.205	1.56
	SVP	145	0.463	0.428	2.13
	TZVP	188	0.552	0.520	2.18
	cc-pVTZ	365	0.666	0.633	2.38
	cc-pVQZ[a]	525	0.672	0.627	2.39
NO$_2$ (132) BLYP//BLYP(SVP)	STO3G	15	1.494	19.128	2.35
	DZ	30	0.207	0.193	1.58
	SVP	45	0.277	0.256	1.94
	TZVP	60	0.313	0.302	1.80
	cc-pVTZ	105	0.308	0.292	1.81
	cc-pVQZ	201	0.340	0.329	1.86

[a] The cc-pVTZ basis set was used on the C atoms.

contains polarisation functions up to and including f for hydrogen and g for first and second row atoms.

In order to determine the effect of the functional, if any, on the accuracy of the RI-J method, we ran some of the calculations with the Local Density Approximation exchange-only (LDAX) functional [35] and some with the Becke 88 gradient corrected exchange functional [36] plus the correlation contribution by Lee, Yang and Parr [37] (BLYP). The LDAX functional is equivalent to the Xα exchange functional whose adjustable parameter α can be determined according to different criteria [38,39]. We use the value $\alpha = 2/3$, which comes from the uniform electron-gas approximation for the exchange energy [35].

As we are interested in ways of evaluating the quality of the RI-J approximation we assess here the usefulness of the Weighted Root Mean Square deviation of the fitted Density (WRMSD) from the "exact" density for this purpose. The WRMSD test computes the following quantity at the end of an RI-J calculation:

$$\sqrt{\sum_{i}^{N_{GRID}} w_i |\rho(\mathbf{r}_i) - \rho_{AUX}(\mathbf{r}_i)|^2} \tag{20}$$

Table 2
Errors in total energy (E) and atomisation energy (E_A) and WRMSD values (Eq. (20)) for the RI-J method. The number of auxiliary basis functions for each molecule is given in parentheses. Also given are the functional and the method for obtaining the geometry (e.g. BLYP//MM means that the BLYP functional was used and the geometry was obtained by molecular mechanics)

Molecule	Basis set	N_{BF}	$E - E_{RI-J}$ (mE$_h$)	$E_{A,RI-J} - E_A$ (mE$_h$)	WRMSD ($\times 10^7$)
B_2H_6 (178) LDAX//LDAX(SVP)	STO3G	16	0.556	5.507	1.32
	DZ	32	0.049	0.017	0.80
	SVP	60	0.049	0.042	0.78
	TZVP	76	0.058	0.044	0.89
	cc-pVTZ	160	0.074	0.057	1.01
	cc-pVQZ	350	0.088	0.062	1.06
$TiCl_4$ (294) BLYP//BLYP(SVP)	STO3G	85	34.405	−0.519	10.39
	DZ	121	4.082	0.323	6.75
	SVP	125	3.895	0.328	6.55
	TZVP	141	3.944	0.336	6.61
	cc-pVTZ	205	3.949	0.345	6.62
	cc-pVQZ	345	4.228	0.383	6.91
$(H_3PAu)_2S$ (439) BLYP//exp	STO3G	111	25.078	−2.727	7.90
	DZ	144	0.227	0.092	1.13
	SVP	165	0.209	0.139	1.03
	TZVP	183	0.242	0.149	1.14
	cc-pVTZ	285	0.285	0.185	1.14
	cc-pVQZ	510	0.718	0.427	1.54
H_3COSiH_3 (229) LDAX//MM	STO3G	25	10.158	4.754	7.64
	DZ	50	0.333	0.241	1.58
	SVP	79	0.342	0.312	1.82
	TZVP	99	0.385	0.348	1.92
	cc-pVTZ	199	0.402	0.358	1.93
	cc-pVQZ	424	0.590	0.419	2.18
$CH_3NH_3^+$ (178) LDAX//MM	STO3G	16	0.714	0.947	1.55
	DZ	32	0.100	0.067	1.18
	SVP	60	0.152	0.134	1.62
	TZVP	76	0.178	0.161	1.64
	cc-pVTZ	160	0.197	0.179	1.71
	cc-pVQZ	350	0.216	0.187	1.77

where ρ and ρ_{AUX} are given by Eqs. (3) and (5), respectively. Both of them are computed with the density matrix $\mathbf{D}$ and the charge $\mathbf{q}$ of the RI-J calculation. The points $\mathbf{r}_i$ and the weights w_i come from the molecular quadrature grid that MAGIC uses to calculate the DFT exchange-correlation energy and matrix elements [8,40–43]. We note that these grids are extensive with 9060 grid points for every first row element, increasing to 15 100 points per atom for actinide elements. It would be reasonable to expect that the WRMSD test should produce a good indication of the quality of the RI-J approximation for results using the same molecule with different basis sets.

MAGIC is able to perform both "exact" and RI-J (V approximation) DFT calculations. We have selected a wide variety of molecules on which we have run both types of calculations. Some of them are inorganic compounds, and for their metal atoms we have used Effective Core Potentials (ECP) [44] to represent the interaction of the core electrons with the valence. Therefore, for the metal atoms, the valence basis set that accompanies their ECP is always used regardless of what basis is used for the rest of the atoms in the molecule. We use this practice very often in applications because ECPs are a cheap and simple way of taking into account the most important of the relativistic effects. We selected the ECPs of the

Table 3
The RI-J approximation for crown ethers of increasing size. The values of WRMSD (Eq. (20)) and the errors in total energy (E) and atomization energy (E_A) are given. 3c1 is 3-crown-1 (C_2H_4O) and contains 65 auxiliary basis functions

Ether molecules	N_{BF}	$E - E_{RI\text{-}J}$ (mE_h)	$E_{A,RI\text{-}J} - E_A$ (mE_h)	WRMSD ($\times 10^7$)
3c1	65	0.429	0.412	2.33
6c2	130	1.071	1.038	2.95
9c3	195	1.456	1.405	2.98
12c4	260	1.885	1.818	3.08
15c5	325	2.239	2.155	2.98
18c6	390	2.738	2.636	2.97
21c7	455	3.257	3.139	2.81

Stuttgart group [45,46] as they have small cores and large valence basis sets for maximum accuracy and also because our auxiliary sets are optimised especially for them.

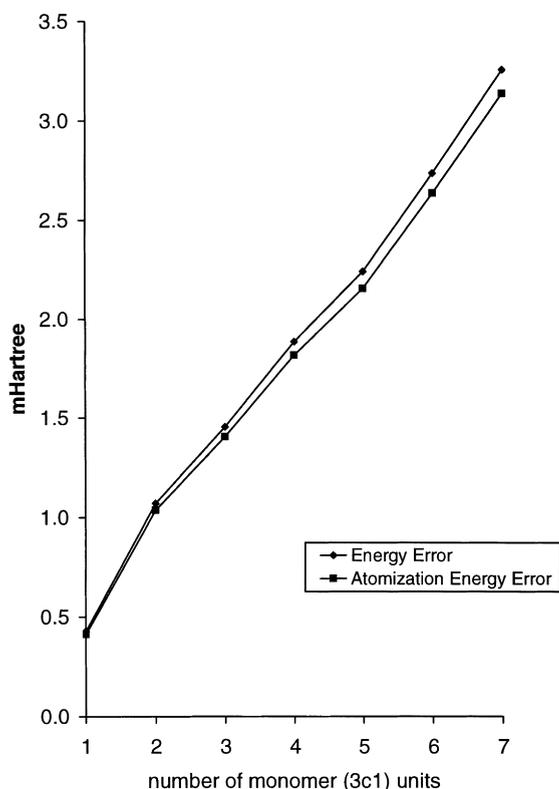The results we obtained for each molecule are given

Fig. 1. Error in energy and in atomization energy for the RI-J approximation as a function of increasing crown ether size.

in Tables 1 and 2. The geometry of RbCl was obtained from a Molecular Mechanics (MM) optimisation and the BLYP functional was used. The geometry of K[PtCl$_3$(C$_2$H$_4$)] was optimised by MAGIC using the SVP basis set and the LDAX functional was used throughout. The geometries of CH$_2$FCOO$^-$ and CH$_3$C$_6$H$_5$ were optimised by MM and the BLYP and LDAX functionals were used, respectively. NO$_2$, B$_2$H$_6$ and TiCl$_4$ were optimised by MAGIC with an SVP basis and the BLYP, LDAX and BLYP functionals were used, respectively. The geometry of (H$_3$PAu)$_2$S was adopted from similar molecules in the literature [47] and the BLYP functional was used. Finally, the LDAX functional was used for both H$_3$COSiH$_3$ and CH$_3$NH$_3^+$ whose geometries were obtained by MM. The number of auxiliary functions and the number of basis functions for each basis set for each molecule are given. For each case we performed an "exact" calculation with four-index two-electron integrals and an RI-J calculation and we present the error in total molecular energy, the error in atomization energy and the WRMSD. Atomization energy is the difference of the total energy of the molecule from the sum of the energies of the neutral isolated atoms comprising the molecule, both calculated either with or without the RI-J approximation.

Having obtained the above-mentioned results on the RI-J approximation for different molecules and basis sets, we performed calculations with the same basis set but on molecules of increasing size. The purpose here was to gain some insight into the effect of the RI-J approximation as the molecules become larger. We therefore ran calculations on crown ethers, starting with 3c1 (C$_2$H$_4$O), to 6c2, all the way up to 21c7. The LDAX functional and the SVP basis set were used and their geometries were obtained by MM. Table 3 contains the results of these calculations, which are again the errors in the energies and atomization energies and the WRMSD. The energy error and atomization energy error are plotted as functions of the number of monomer units (3c1) in Fig. 1.

4. Discussion

Though the list of molecules we have studied is by no means exhaustive, it contains many cases similar to the ones we use in chemical applications with DFT.

We selected equilibrium, or near equilibrium geometries where DFT is most commonly used. We focus on energies, since our auxiliary basis sets are optimised for them and therefore they are a direct indication of how well the RI-J approximation works. Furthermore, atomization energies are useful for comparisons of stability between different molecules and do not depend on errors in the description of the core region. Many molecular properties have this characteristic.

From Tables 1 and 2, we can observe that the RI-J energy is lower than the non fitted energy in accordance with the generalisation of the variational principle of Eq. (17). The atomisation energies of the RI-J calculations are larger than the four-index results, probably because the RI-J energies of the isolated atoms are calculated more accurately than the energies of the molecules. Only in $TiCl_4$ and $(H_3PAu)_2S$, both with the STO3G basis, the four-index atomization energy is higher than the RI-J atomization energy.

Our next observation is that our target accuracy of less than 1 mE_h has been achieved for the atomisation energies for the SVP and TZVP basis sets for which our auxiliary basis functions were designed. In most molecules, this is true for the total energy that gives errors of comparable magnitude. However, in $K[PtCl_3(C_2H_4)]$ and $TiCl_4$, which include third row or lower elements, the errors in energy are much larger and way above our threshold. These errors can be attributed to the density of the core electrons and cancel in the atomization energies, as they are not affected by the chemical environment of the atoms.

We now look at how the rest of our basis sets performed. Surprisingly, STO3G, which is much smaller than SVP and TZV and in some cases has almost an order of magnitude fewer functions than the auxiliary basis, performs rather poorly. It seems that the auxiliary basis does not have the flexibility to follow its inadequate description of the density, which is substantially different from the density of a near-complete basis. This causes errors in the core region description for RbCl, $K[PtCl_3(C_2H_4)]$ and $TiCl_4$, in the valence region for NO_2 and B_2H_6 and in both core and valence for CH_2FCOO^-, $CH_3C_6H_5$, $(H_3PAu)_2S$ and H_3COSiH_3. Only in $CH_3NH_3^+$ it seems to perform well, probably as a result of its positive charge

which contracts its electron density; but it is still worse than the rest of the basis sets.

DZ performs in most cases almost equally well or a little better than the SVP and TZVP bases. This should probably be the general trend for a basis set smaller than the ones for which the auxiliary basis set was optimised but larger than minimal.

The cc-pVTZ and cc-pVQZ basis sets perform surprisingly well. They afford comparable errors to the SVP and TZVP basis sets though they have almost the same number or more (for the cc-pVQZ case) functions than the auxiliary basis. This shows that the auxiliary basis sets are quite robust towards the complete basis set limit and can probably be considered as "universal" auxiliary basis sets as long as the basis set has plenty of flexibility.

In evaluating the usefulness of the WRMSD as a means of deciding the suitability of the auxiliary basis for a particular calculation we observe the following. First of all the WRMSD is not an absolute measure of the quality of a calculation but it has to be compared to some reference, "good" calculation for the molecule in question. For example, in the $(H_3PAu)_2S$ molecule, if our reference result is the one with the SVP basis, we can observe that DZ, TZVP and cc-pVTZ are equally good, with similar values of WRMSD. On the other hand, cc-pVQZ is not as accurate and yields a slightly larger value of WRMSD, while STO3G is much worse and this causes WRMSD to increase substantially. This rationale holds for the rest of the results apart from STO3G $CH_3C_6H_5$. Overall, the WRMSD seems to be fairly sensitive to changes in errors. Its usefulness lies in the fact that it can give (most of the time) some indication about the quality of the RI-J approximation with a particular basis set if no four-index program is available. However, it requires an extra calculation with the reference basis set and gives no a priori indication of the magnitude of the error caused by an unsuitable basis set.

Table 3 presents the variation of the energy error, atomisation energy error and WRMSD with increasing molecular size for the SVP basis set. We observe that the WRMSD is almost constant, as should be expected for the same basis set and class of molecules. The error in energy and atomization energy however grows proportionally with the size of the molecule as shown more clearly in Fig. 1. This was to be expected, since the auxiliary basis sets were expected to produce

errors of less than 0.2 mE$_h$ per atom. It is worrying however to observe that our target of 1 mE$_h$ is missed even for the atomization energy of 6c2. Since the RI-J method is meant for much larger molecules than this, we must hope that this increasing error with molecular size will not render the chemical information we can obtain from a large molecule useless, although this has to be tested. At least we can take advantage of this regular increase of the error with molecule size and use it in a more direct approach to predict the performance of a particular basis set for an auxiliary basis: RI-J and four-index calculations on a molecule much smaller than the one of interest should provide an error from which we can estimate the error on the large molecule by extrapolation.

In all the RI-J calculations, we observed that the convergence of the SCF procedure was either as good as the four-index case or better for some cases, in the sense that it required less iterations. There seems to be no observable trend in RI-J accuracy related to the type of functional that we used. BLYP which is a gradient corrected functional is not more demanding in terms of auxiliary basis set requirements than LDAX. Our calculations also raise the question of the accuracy of the RI-J approximation for molecules containing transition metals. The answer should be probably postponed to a future paper, as the examples presented here are not sufficient for this purpose. We can note however that complexes with quasi-spherical metals such as Au and Rb are more accurately described than complexes containing transition metals with partly filled d orbitals such as Ti and Pt.

5. Conclusions

We have tested extensively the accuracy of the existing auxiliary basis sets for a variety of molecules and basis sets. We have found that when used in conjunction with smaller basis sets, than the ones they were designed for, they perform equally well except in the case of minimal basis sets. Furthermore, they performed very well in conjunction with increasingly larger basis sets, producing errors of the same order of magnitude as their standard basis sets. The robustness of the auxiliary basis sets towards the complete basis set limit is a particularly interesting result. Increasing size of basis set leads to very steep

increase in the cost of a calculation and an accurate RI-J approximation with current, readily available auxiliary sets makes possible much larger calculations.

We have introduced the WRMSD test for predicting how well a combination of basis set and auxiliary set for a particular molecule will perform. It is useful when there is no capability for an "exact" four-index calculation, however it requires a reference RI-J result. For the auxiliary bases we tested, this reference result is a calculation with the SVP or TZVP basis set. When four-index calculations are feasible, we suggest a much more direct test for the suitablity of an auxiliary basis set. An RI-J and a four-index calculation on a molecule similar to the one in question but much smaller give an error which can reliably predict the error of the RI-J calculation for the large molecule. As Table 3 shows, the error of an RI-J calculation grows linearly with the size of the molecule. This should always be taken into account.

The RI-J method appears to be the only practical way for DFT calculations on the type of large inorganic molecules we are interested in. We will continue to use it and based on the results of this work we will ensure that we achieve the accuracy we desire for each calculation. Furthermore, we wish to fill in the gap in the literature by designing auxiliary basis sets suitable for the valence basis sets of effective core potentials for lanthanide and actinide elements. Finally, the option of an algorithm for optimising an auxiliary basis set on the fly for each new basis set we encounter is to be explored.

Acknowledgements

References

[1] R. Gáspár, Int. J. Quantum Chem. 29 (1986) 1043.
[2] J. Almlöf, K. Korsel, K. Faegri Jr., J. Comput. Chem. 3 (1982) 385.
[3] M. Häser, R. Ahlrichs, J. Comput. Chem. 10 (1988) 104.
[4] H. Horn, H. Weiss, M. Häser, M. Ehring, R. Ahlrichs, J. Comput. Chem. 12 (1991) 1058.

[5] R.E. Stratmann, G.E. Scuseria, M.J. Frisch, Chem. Phys. Lett. 257 (1996) 213.

[6] C.A. White, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, Chem. Phys. Lett. 230 (1994) 8.

[7] M.C. Strain, G.E. Scuseria, M.J. Frisch, Science 271 (1996) 51.

[8] A. Willetts, L. Gagliardi, A.G. Ioannou, A.M. Simper, C.-K. Skylaris, S. Spencer, N.C. Handy, Int. Rev. Phys. Chem. (2000) in press.

[9] L. Gagliardi, A. Willetts, C.-K. Skylaris, N.C. Handy, S. Spencer, A.G. Ioannou, A.M. Simper, J. Am. Chem. Soc. 120 (1998) 11727.

[10] E.J. Baerends, D.E. Ellis, P. Ros, Chem. Phys. 2 (1973) 41.

[11] B.I. Dunlap, J.W.D. Connolly, J.R. Sabin, J. Chem. Phys. 71 (1979) 3396.

[12] L.E. McMurchie, E.R. Davidson, J. Comput. Phys. 26 (1978) 218.

[13] N.H.F. Beebe, J. Linderberg, Int. J. Quantum Chem. 12 (1977) 683.

[14] D.W. O'Neal, J. Simons, Int. J. Quantum Chem. 36 (1989) 673.

[15] G.G. Hall, C.M. Smith, Int. J. Quantum Chem. XXV (1984) 881.

[16] C.M. Smith, G.G. Hall, Theor. Chim. Acta 69 (1986) 63.

[17] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, Chem. Phys. Lett. 240 (1995) 283.

[18] K. Eichkorn, F. Weingend, O. Treutler, R. Ahlrichs, Theor. Chem. Acc. 240 (1995) 283.

[19] M.W. Schmidt, K. Ruedenberg, J. Chem. Phys. 71 (1979) 3951.

[20] L. Gagliardi, N.C. Handy, A.G. Ioannou, C.-K. Skylaris, S. Spencer, A. Willetts, A.M. Simper, Chem. Phys. Lett. 283 (1998) 187.

[21] S. Spencer, L. Gagliardi, N.C. Handy, A.G. Ioannou, C.-K. Skylaris, A. Willets, A.M. Simper, J. Chem. Phys. A 103 (1999) 1831.

[22] N. Godbout, D.R. Salahub, J. Andzelm, E. Wimmer, Can. J. Chem. 70 (1992) 560.

[23] O. Vahtras, J. Almlöf, M.W. Feyereisen, Chem. Phys. Lett. 213 (1993) 514.

[24] C. Van Alsenoy, J. Comput. Chem. 9 (1988) 1398.

[25] A. Komornicki, G. Fitzgerald, J. Chem. Phys. 98 (1993) 1398.

[26] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, 2, Cambridge University Press, Cambridge, UK, 1994.

[27] A. Schäfer, H. Horn, R. Ahlriches, J. Chem. Phys. 97 (1992) 2571.

[28] A. Schäfer, C. Huber, R. Ahlrichs, J. Chem. Phys. 100 (1994) 5839.

[29] W.J. Hehre, R.F. Stewart, J.A. Pople, J. Chem. Phys. 51 (1969) 2657.

[30] W.J. Hehre, R. Ditchfield, R.F. Stewart, J.A. Pople, J. Chem. Phys. (1970) 2769.

[31] T.H. Dunning, J. Chem. Phys. 53 (1970) 2823.

[32] T.H. Dunning, P.J. Hay, Methods of Electronic Structure Theory, 3, Plenum Press, New York, 1977, pp. 1–27 chap. 1.

[33] T.H. Dunning, J. Chem. Phys. 90 (1989) 1007.

[34] D.E. Woon, T.H. Dunning, J. Chem. Phys. 98 (1993) 1358.

[35] R.G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules, Oxford University Press, Oxford, 1989.

[36] A.D. Becke, Phys. Rev. A 38 (1988) 3098.

[37] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B (1988) 32.

[38] R. Gáspár, Acta. Phys. Hung. 35 (1974) 213.

[39] R. Gáspár, A. Nagy, J. Phys. B 20 (1987) 3631.

[40] V.I. Lebedev, Zh. Vychlisl. Mat. Mat. Fiz. 15 (1975) 48.

[41] V.I. Lebedev, Zh. Vychlisl. Mat. Mat. Fiz. 16 (1976) 293.

[42] V.I. Lebedev, Sibirsk. Mat. Zh. 18 (1977) 132.

[43] M.E. Mura, P.J. Knowles, J. Chem. Phys. 104 (1996) 9848.

[44] C.-K. Skylaris, L. Gagliardi, N.C. Handy, A. Ioannou, S. Spencer, A. Willetts, A.M. Simper, Chem. Phys. Lett. 296 (1998) 445.

[45] M. Dolg, U. Wedig, H. Stoll, H. Preuss, J. Chem. Phys. 86 (1987) 866.

[46] W. Küchle, M. Dolg, H. Stoll, H. Preuss, J. Chem. Phys. 100 (1994) 7535.

[47] P.G. Jones, C. Thöne, Chem. Ber. 124 (1991) 2725.