

## Regular article

# A scheme for the economical use of numerical basis sets in calculations with SIESTA

R. Q. Zhang, Q. Z. Zhang, M. W. Zhao

Center of Super-Diamond and Advanced Films & Department of Physics and Materials Science, City University of Hong Kong, Hong Kong SAR, China

Received: 16 September 2003 / Accepted: 5 January 2004 / Published online: 16 June 2004  
© Springer-Verlag 2004

**Abstract.** We demonstrate the effectiveness of an economical scheme that uses numerical basis sets in computations with SIESTA. The economical basis sets demonstrated, in which high-level double-zeta basis plus polarization orbitals (DZP) are applied only for atoms of strong electronegativity and metal atoms while a double-zeta basis is applied to the rest of the atoms of small proton-bound carboxylic acid clusters and sodium–organic compounds, predict correct geometric structures very close to those obtained using DZP for all atoms. The use of economical basis sets can save about 30–50% of the CPU time that is used for calculations with large basis sets. This study provides a general guideline for basis set selection in SIESTA computations of large systems.

**Keywords:** SIESTA – Economical basis sets – Proton-bound carboxylic acid cluster – Metal-organic interaction

## Introduction

The development of the density functional theory (DFT) [1], in particular by incorporating formalisms as does a tight-binding scheme, allows increasingly complex systems to be treated quantum mechanically with moderate accuracy [2, 3, 4, 5]. The tools developed are particularly useful for cases in which a smaller model system would inadequately describe the proper physical environment. Large systems such as nucleic acid and enzymes can now be treated with reasonable accuracy, which is computationally unattainable using first-principles methods. Because of the increase in computational power and the further progress in methodology (i.e., linear scaling), even large systems with increased complexity can now be treated computationally, as recently shown with a

SIESTA method [6, 7]. It is also an efficient approximation that the active site of the system and the remainder are treated differently. In particular, the hybrid quantum mechanics/molecular mechanics method, in which the active site is calculated with quantum mechanics and the rest with molecular mechanics, is very useful in computations of large molecules [8].

The computational cost using SIESTA is also related to the basis set used. The more complicated the basis set, the higher the computational resources required. However, for complicated systems such as those that involve hydrogen bonds or weak interactions between metal atoms and organic molecules, high-level basis sets are necessary to give reliable geometric structures and binding energies. Therefore, it is desirable to develop an economical scheme for selecting basis sets, in which the reliability of the result is maintained and the computational costs are minimized. Recently, the principle for applying such a scheme to various systems in calculations that use Hartree-Fock theory and DFT has been studied [9, 10, 11, 12]. However, there is a need to devise a similar scheme for the use of numerical basis sets in SIESTA calculations, as SIESTA is becoming an increasingly important tool in the computation of large systems in a broad range of fields, including biology, the life sciences, and materials science.

In the study reported herein, various schemes for selecting basis sets were examined in SIESTA calculations of a proton-bound dimer of formic acid and the Na-8-quinolinol compound. The efficiency of the economical basis set was then demonstrated by using it to study the meridional tris(8-hydroxyquinoline)aluminum (Alq<sub>3</sub>) molecule.

## Theoretical approach and computational details

The SIESTA that we used adopted DFT with the local density approximation for exchange–correlation [13] and norm-conserving pseudopotentials [14, 15]. Various

Correspondence to: R. Q. Zhang  
e-mail: aprqz@cityu.edu.hk

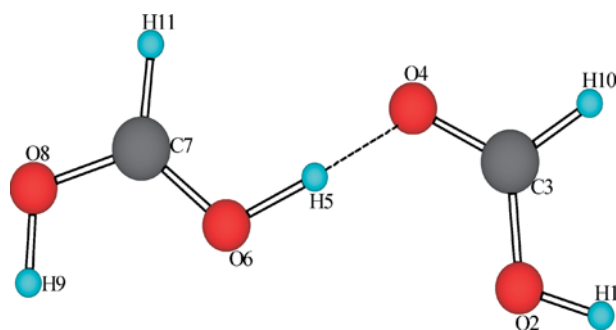
basis sets, ranging from single-zeta (SZ) to double-zeta plus polarization orbitals (DZP) were applied. The energy cutoff (for real-space mesh points) was chosen to be 100 Ry for all of the calculations. All of the geometries were optimized by conjugate gradients, with the remaining forces being less than 0.01 eV/Å. The cutoff radii of the basis functions, which are defined by a parameter, “energy shift”, were chosen to be 0.02 Ry. Although no periodic boundary conditions were considered for the molecular systems selected in this work, the economical basis sets demonstrated are expected to be applicable to systems described with periodic boundary conditions (e.g., in a supercell approach). Because the validity of SIESTA in treating various systems has been established [16], this work is limited to a comparison of the results that were obtained by using a high-level basis set (DZP) and lower basis sets.

## Results and discussion

### *Proton-bound dimer of formic acid*

Proton-bound formic acid clusters can form hydrogen-bond-containing networks [17]. Much effort has been exerted to study the proton-bound formic acid clusters using ab initio Hartree–Fock and density functional methods [17, 18, 19]. Proton-bound formic acid clusters are systems that are very sensitive to the basis set because hydrogen bonds are involved. To determine an economical basis set for general usage on systems that involve hydrogen bonds, a proton-bound formic acid dimer was chosen for calculations using SIESTA with different basis set combinations.

The structural parameters that were calculated for the proton-bound formic acid dimer are defined in Fig. 1. The deviations of the geometric parameters that were optimized with different combinations of basis sets from those calculated using the high-level basis set (DZP) are listed in Table 1. The geometric structures that were obtained strongly depended on the basis set employed in the calculations. The deviations of the bond lengths and bond angles that were calculated using lower-level basis sets, such as SZ, double zeta (DZ), and single zeta plus



**Fig. 1.** Chain structure of a proton-bound dimer of formic acid  $(\text{HCOOH})_2\text{H}^+$ . The dashed line represents the hydrogen bond involved

polarization (SZP), from the results of DZP used by all atoms were remarkable. For example, the largest relative deviation of the bond lengths/bond angles that was predicted using the SZ basis set for all atoms was up to 2.63%/3.76%. When the basis functions of oxygen atoms were improved to DZP, the results greatly improved. The structural parameters described using a DZP basis set on an O centers and DZ basis sets on H and C atoms (this composite basis set is denoted as DZP#) were quite similar to those deduced from the high-level basis set (DZP for all atoms), while the CPU time per structural update was reduced by about 33%. It should be mentioned that if DZP basis functions are applied to carbon atoms only, then the results do not improve as remarkably as when DZP is applied to oxygen atoms only. This is related to the fact that oxygen has higher electronegativity than carbon. When oxygen atoms form bonds with other atoms, electron transfers occur, and higher basis functions are needed to describe the deformation and diffusion of the electron cloud. However, atoms of lower electronegativity lose electrons when forming bonds with other atoms. Their electron clouds contract and are easy to describe even with lower-level basis functions. Compared with calculations that use the high-level basis set, the composite basis set is more economical in the calculation of large organic molecules, containing a large number of carbon atoms, where more computational resources are normally needed.

### *Na-8-quinolinol*

Interactions between an alkali metal and organic functional groups exist widely in molecules of living bodies and materials, and play key roles in many important physical and chemical processes. Recently, alkali-metal interaction with organic molecules was found to be an important issue in organic light-emitting devices, in which a metal–organic interface forms [20, 21]. The electronic structure of the organic molecules can be modified by the interaction with metal atoms. To better understand and improve the charge injection properties of the metal–organic interface and thus the efficiency of the devices, theoretical studies on metal–organic interactions and related issues, such as the origin of new electronic states, charge transfers, and dipoles at the interface, are very necessary. Theoretical studies of the interaction between metal atoms and large biomolecules such as proteins are also necessary in biophysics and biochemistry, and may be helpful for understanding many biochemical processes in living bodies.

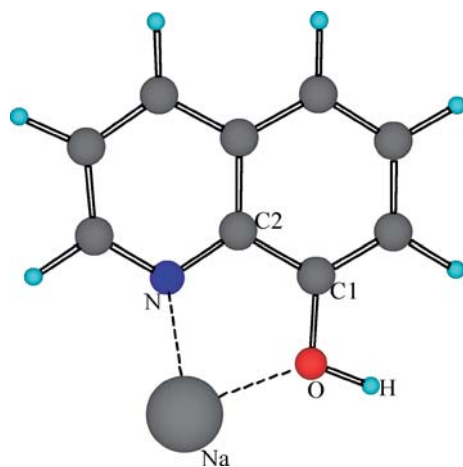
We studied Na-8-quinolinol using SIESTA calculations with different basis set schemes. By comparing the molecular structure of this compound as predicted by different basis sets, a scheme for selecting an economical basis set was demonstrated for systems that involve weak interaction between an organic molecule and metal atoms. The most economical basis set was then applied

**Table 1.** Deviations of the bond lengths and bond angles for a proton-bound dimer of formic acid  $(\text{HCOOH})_2\text{H}^+$  calculated using various basis sets from the results of a standard basis set (double-zeta basis plus polarization, *DZP*) in SIESTA calculations. The bond lengths are in angstroms and the bond angles are in degrees. The basis set combinations adopted are defined as *DZP#*: H:DZ/C:DZ/O:DZP, *Gen1*: H:SZ/C:SZ/O:DZP, and *Gen2*: H:SZP/C:SZP/O:DZP

	DZP	SZ	SZP	DZP#	Gen1	Gen2
H1-O2	0.989	0.026	0.113	0.003	0.059	0.032
O2-C3	1.290	0.010	0.120	0.020	0.010	0.000
C3-O4	1.220	0.020	0.110	0.020	0.010	0.000
O4-H5	1.290	-0.040	-0.020	-0.014	-0.040	-0.010
O6-C7	1.240	0.020	0.110	0.020	0.000	0.000
C7-O8	1.270	0.020	0.120	0.020	0.030	0.010
C3-H10	1.110	-0.010	0.110	-0.010	0.110	0.080
$\angle 1-2-3$	112.520	2.390	-0.300	2.340	-1.620	1.630
$\angle 2-3-4$	126.440	-0.439	2.080	-0.440	3.480	2.170
$\angle 3-4-5$	149.410	5.620	-8.030	-0.360	-17.650	-2.090
$\angle 6-7-8$	122.820	-0.430	-1.770	-0.260	2.350	1.350
$\angle 7-8-9$	112.070	2.390	-0.960	1.990	-3.520	-0.520
$\angle 2-3-10$	112.270	-0.160	-2.300	0.100	-1.550	-1.550

to study  $\text{Alq}_3$  molecules. The configuration of Na-8-quinolinol is shown in Fig. 2. Selected bond lengths and bond angles that were obtained by using different basis sets are listed in Table 2.

The deviations from the DZP results of the bond lengths and bond angles that were calculated using lower-level basis sets, such as SZ, SZP, and DZ, for all atoms were remarkable. The results that were obtained by applying the high-level basis set (DZP) only to Na, N, and O atoms, while applying the lower-level basis set (DZ) to carbon atoms and hydrogen atoms (as denoted by DZP#), are very close to those obtained by applying the high-level basis set (DZP) to all atoms, and the CPU time per cycle was reduced by about 40%. However, when DZP basis functions were only applied to Na, N, or O atoms, the deviations of the results from those of the high-level basis set (DZP) were still very large. This indicates that applying the highest basis functions only to metal atoms and atoms of high electronegativity in the calculation of metal-containing organic molecules is an effective method to maintain the reliability of the results with lower computational costs. We additionally examined the effect of the range of the basis function on the structural parameters. The energy shift, which defines the cutoff radii of the basis functions, corresponds



**Fig. 2.** The geometric structure of Na-8-quinolinol

to the excitation energy of pseudo-orbital wavefunctions owing to the confinement to a finite range. The results shown in Table 2 indicate that the remarkable differences between the SZ and DZP results are not reduced by changing the range of the basis functions.

#### *Application of the most economical basis set to calculation of $\text{Alq}_3$*

$\text{Alq}_3$  is one of the widely used organic materials for organic light-emitting devices because of its excellent electron-transporting and light-emitting properties. We calculated an  $\text{Alq}_3$  molecule by using SIESTA with the highest basis set (DZP) and the economical basis set (DZP#). In DZP#, DZP was applied only to nitrogen, oxygen, and aluminum atoms, while the lower basis set, DZ, was applied to carbon and hydrogen atoms. Selected bond lengths and bond angles of  $\text{Alq}_3$  that were obtained by using different basis sets are listed in Table 3. The corresponding geometric structure is shown in Fig. 3. It is clear that the use of the economical basis set can produce results that are very close to those obtained with the high-level basis set (DZP); however, about half of the CPU time is saved. We also studied the total energy of  $\text{Alq}_3$  by using single-point energy calculations with the high-level basis set (DZP) based on the structure that was optimized with the high-level basis set (DZP) and the economical basis set (DZP#), respectively, and found that the difference was only 0.082 eV per atom. This indicates that the use of an economical basis set in which the high-level basis set (DZP) is applied only to metal atoms and atoms with high electronegativity is an efficient approximation in SIESTA calculations of systems that involve the interaction between metal atoms and organic molecules. We should point out here that the efficiency of the economical basis set in the SIESTA calculation of  $\text{Alq}_3$  is not as high as that in calculations with the accurate DFT method, where the CPU time is reduced by nearly 16 times with the use of the economical basis set as compared with the use of the 6-31+G\* basis set [22]. This may be due to the fact that many approximations have been adopted in SIESTA to save computational resources, which reduces

**Table 2.** Selected bond lengths and bond angles of Na-8-quinolinol obtained by using SIESTA calculations with different basis sets. The data of the lower-level basis set are the deviations from the results of the standard basis set (DZP). The bond lengths are in angstroms and the bond angles are in degrees. The basis set

	DZP	SZ	SZ <sup>a</sup>	SZ <sup>b</sup>	DZ	SZP	DZP#	Gen1	Gen2	Gen3
Na–O	2.206	−0.087	−0.082	−0.076	−0.069	−0.029	0.000	−0.074	−0.048	−0.057
Na–N	2.230	−0.060	−0.080	0.070	−0.030	0.010	0.050	−0.020	−0.030	−0.010
O–C1	1.360	0.070	0.060	0.060	0.027	0.030	0.000	0.020	0.010	0.030
C1–C2	1.420	0.050	0.050	0.060	−0.010	0.020	0.040	0.000	0.000	−0.010
N–C2	1.360	0.060	0.060	0.060	0.010	0.030	−0.011	0.000	0.010	0.010
H–O	0.970	0.100	0.100	0.100	0.030	0.060	0.060	0.030	0.020	0.030
C–H	1.100	0.080	0.070	0.070	0.000	0.050	0.080	0.000	0.000	0.000
∠N–Na–O	75.870	5.880	5.320	5.620	1.810	2.160	0.160	1.460	1.150	1.570
∠Na–O–C1	114.370	−2.210	−1.650	−1.810	−0.190	−0.920	0.540	0.280	0.170	−0.250
∠Na–N–C2	112.560	−2.340	−1.230	−0.620	−1.390	−1.450	−0.510	−0.870	−0.490	−1.390
∠N–C2–C	119.720	−2.410	−0.950	−0.710	−1.730	−2.200	−0.270	−4.500	−2.340	−1.600

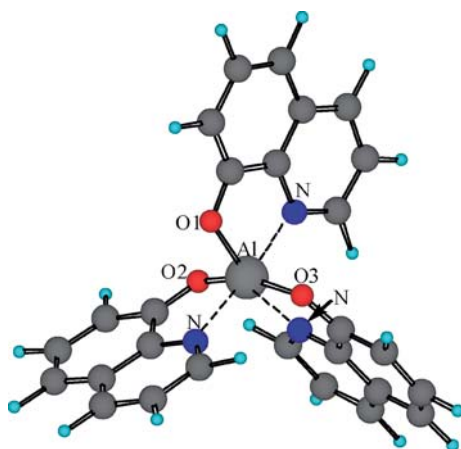
<sup>a</sup>The energy shift was chosen to be 0.025 Ry

<sup>b</sup>The energy shift was chosen to be 0.03 Ry

**Table 3.** The geometric structure of tris(8-hydroxyquinoline) aluminum obtained by using SIESTA calculations with a standard basis set and an economic basis set. The bond lengths are in angstroms, and the bond angles are in degrees. The CPU time is in seconds per cycle for a personal computer with a 1.2 GHz CPU

Basis sets	Al–O	Al–N	∠N–Al–O	∠O–Al–O	CPU time
DZP	1.906	2.027	83.88	101.19	22.04
DZP#(H:DZ/C:DZ/N:DZP/O:DZP/Al:DZP)	1.870	2.040	85.37	100.94	10.85

the difference of time expenditure between calculations with a lower-level basis set and those with the high-level basis set. However, the economical basis set scheme for SIESTA has potential applications, especially for the calculation of systems that involve large organic molecules.



**Fig. 3.** Structure of the meridional tris(8-hydroxyquinoline)aluminum molecule optimized by using a SIESTA calculation with economic basis set DZP# (H:DZ/C:DZ/N:DZP/O:DZP/Al:DZP)

combinations adopted are defined as DZP# H:DZ/C:DZ/N:DZP/O:DZP/Na:DZP, Gen1 H:DP/C:DP/N:DZP/O:DZ/Na:DZ, Gen2 H:DZ/C:DZ/N:DZ/O:DZP/Na:DZ, and Gen3 H:DZ/C:DZ/N:DZ/O:DZ/Na:DZP

## Conclusion

The present economical basis set, which uses the high-level basis set (DZP) only for metal atoms and atoms of large electronegativity in the computation of large organic systems that involve weak interaction, such as hydrogen bonds and interactions between metal atoms and organic molecules, can predict geometric structures very close to those obtained using the high-level basis set (DZP) for all atoms in SIESTA calculations. Energetic results of good accuracy may be achieved by further performing a single-point calculation on the geometric structure determined using the economical basis set. The principle for basis set selection may be regarded as a general guideline in SIESTA computations of large systems.

*Acknowledgement.* The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. CityU 1033/00P).

## References

1. Kohn W, Sham LJ (1965) *Phys Rev* 140:A1133
2. Sankey OF, Niklewski DJ (1989) *Phys Rev B* 40:3979
3. Lewis JP, Glaesemann KR, Voth GA, Fritsch J, Demkov AA, Ortega J, Sankey OF (2001) *Phys Rev B* 64:195103
4. Ordejón P, Artacho E, Soler JM (1996) *Phys Rev B* 53:R10441
5. Sánchez-Portal D, Ordejón P, Artacho E, Soler JM (1997) *Int J Quantum Chem* 65:453
6. Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D (2002) *J Phys Condens Matter* 14:2745
7. de Pablo PJ, Moreno-Herrero F, Colchero J, Gómez-Herrero J, Herrero P, Baró AM, Ordejón P, Soler JM, Artacho E (2000) *Phys Rev Lett* 85:4992
8. Zhang YK, Lee TS, Yang WT (1999) *J Chem Phys* 110:46
9. Zhang RQ, Wong NB, Lee ST, Zhu RS, Han KL (2000) *Chem Phys Lett* 319:213
10. Zhang RQ, Xie XG, Liu SX, Lee CS, Lee ST (2000) *Chem Phys Lett* 330:484
11. Zhang RQ, Chu TS, Lee ST (2001) *J Chem Phys* 114:5531
12. Zhang RQ, Lu WC, Cheung HF, Lee ST (2002) *J Phys Chem B* 106:625

13. Perdew JP, Zunger A (1981) Phys Rev B 23:5048
14. Troullier N, Martins JL (1991) Phys Rev B 43:1993
15. Kleinman L, Bylander DM (1982) Phys Rev Lett 48:1425
16. <http://www.uam.es/siesta>
17. Zhang RQ, Lifshitz C (1996) J Phys Chem 100:960
18. Boyd SL, Boyd RJ (1997) J Am Chem Soc 119:4214
19. Colominas C, Teixido J, Cemeli J, Luque FJ, Orozco M (1998) J Phys Chem B 102:2269
20. Lee ST, Hou XY, Mason MG, Tang CW (1998) Appl Phys Lett 72:1539
21. Rajagopal A, Kahn A (1998) J Appl Phys 84:355
22. Zhang RQ, Lu WC, Lee CS, Hung LS, Lee ST (2002) J Chem Phys 116:8827