Vol. 43 No. 4 SCIENCE IN CHINA (Series B) August 2000

An effective scheme for selecting basis sets for *ab initio* calculations

ZHANG Ruiqin (张瑞勤)[1], HUANG Jianhua (黄建华)[2], BU Yuxiang (步宇翔)[1, 3], HAN Keli (韩克利)[2], LEE Shuit-Tong (李述汤)[1] & HE Guozhong (何国钟)[2]

1. Centre of Super-Diamond and Advanced Films (COSDAF) & Department of Physics and Materials Science, City University of Hong Kong, Hong Kong, China;
2. State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China;
3. Present correspondence address: Institute of Theoretical Chemistry, Shandong University, Jinan 250100, China

Correspondence should be addressed to Zhang Ruiqin (email: aprqz@cityu.edu.hk)

Received August 25, 1999

Abstract An effective scheme for selecting economical basis sets for *ab initio* calculations has been proposed for wide-range systems based on the analysis of different functions in the currently used basis sets. Accordingly, the selection of the basis sets should be made according to the different properties and real chemical surrounding of the atoms in the systems. For normal systems, the size and level of the basis sets used for the descriptions of the constituent atoms should be increased from left to right according to the position of the atom in the periodic table. Moreover, the more the atom is negatively charged, the more the basis functions and suitable polarization functions and diffuse functions should be utilized. Whereas, for the positively charged atoms, the size of basis set may be reduced. It is not necessary to use the polarization and diffuse functions for the covalently saturated atoms with normal valence states. However, for the system involving hydrogen-bonding, weak interactions, functional groups, metallic bonding with zero valence or low positive valence, and other sensitive interactions, the polarization and diffuse functions must be used. With this scheme, reliable calculation results may be obtained with suitable basis sets and smaller computational capability. By detailed analysis of a series of systems, it has been demonstrated that this scheme is very practical and effective. This scheme may be used in Hartree-Fock, Møller-Plesset and density functional theoretical calculations. It is expected that the scheme would find important applications in the extensive calculations of large systems in chemistry, materials science, and life and biological sciences.

Keywords: ***ab initio*** **calculation, basis sets, population analysis, polarization function, diffuse function.**

Quantum chemical *ab initio* calculations have been shown to be very successful in investigating the structures and properties of molecular systems. Such a theoretical approach solves the quantum-mechanical many-electron Schrödinger equation by considering all the electrons involved[1—4]. By using complete basis sets including electron correlation, satisfactory agreement with experiments has been found for many medium- or small-size molecular systems in terms of their energetic, spectroscopic and other properties. The relevant properties and regularity of some unknown and metastable molecules have been accurately predicted on the basis of the calculations. It indicates the feasibility to study various systems from a pure theoretical way. However, with the development of natural science in various fields and due to their hybridization, many issues relating to the complicated systems in materials science, clusters, solvents and biology need understanding or interpretation by means of theoretical study. This leads to the required CPU time and the disk space in computations of these systems using *ab initio* or density functional theories (DFT) increased rapidly. In the conventional approaches for using these first-principle theories, there is no clear guiding principle for selecting a suitable basis set for practical calculations so that the number of basis sets used for calculations is generally determined empirically. Almost every quantum chemist needs to conduct a large number of tests so as to gain sufficient experience. In order to achieve satisfactory accuracy and reliability, very large basis sets are normally used, leading to the overuse of the computational facilities and making the study very costly. It would be of great significance to search for a new scheme for selecting basis sets which can maintain the reliability of the result while with the minimal computational cost. In the conventional scheme of selecting the basis sets, the same number and the same kind of basis functions are used to describe all atoms involved except for the hydrogen atoms, regardless of what kind of atoms they are[5]. Obviously, such kind of selection is not optimal. An accurate description for each atom in the molecular system should consider the nature and the environment of the atom. In this work, by analyzing such an issue and by examining a large number of examples, an economical and effective scheme for selecting basis set has been demonstrated.

1 Basis sets and an economical and effective scheme for selecting basis set

1.1 Basis sets

The *ab intio* theory of quantum chemistry uses the atomic orbitals of all the atoms involved to construct the basis space[1 –3]. The basis functions are composed of angular and radial parts. The angular part is spherical harmonic functions while the radial part generally uses STO and GTO sets. By taking the advantages of both functions, STO-GTO basis sets have been proposed to be the linear combination of GTO to form STO. According to the difference of the contribution from the inner- and outer-shell electrons in construction of a molecule, the mixed z set is normally used, which is the most popular basis set in the current *ab initio* calculations. By considering the difference of the electrons in the molecule and in the atoms as well as the deformation of their orbitals, polarization functions have also been introduced to make correction[2]. Although the use of polarization functions has increased the size and level of the basis sets used, the described electron motion in the molecule is much more accurate. Further correction has been made by adopting the diffuse functions[6, 7]. Such functions are useful in describing the motion of the electrons in electron-rich area. Obviously, it is particularly important in describing the electron which is far from the nucleus in an electron-rich molecule, such as the one with lone-pair electrons or an anionic system which possesses a large radius.

1.2 Theoretical scheme for selecting economical and effective basis set

To deal with a practical system using quantum-chemistry *ab initio* calculations, it would be necessary to use more complete basis set in addition to the consideration of electron correlation. The more complete the basis set utilized is, the more reliable the result obtained is and the better agreement with experiments would be achieved. Obviously, such a treatment would encounter much difficulty when calculating a large system. Following our previous work using an economical basis set to treat protonated systems[8], a general scheme for selecting economical and effective basis set for wide-range heteroatomic systems is proposed here. In this new scheme, the differences in the role of different basis functions adopted in *ab initio* basis set are considered. Moreover, for a same kind of basis function, a different role may be found when it is used for different systems. Thus, the amount and type of the basis functions for describing each atom in the molecule should be determined according to the real occupation and behavior of its valence electron. The less electron the atom possesses, the less basis functions it may require. Hence, the amount and type of the basis functions to describe an atom would be increased in the order from left to right of its appearance in the periodic table. Furthermore, if the atom is negatively charged, more basis functions should be used. In contrast, if positively charged, the use of the basis function could be cut down. On the other hand, the polarization and diffuse functions are the correction functions which can describe the deformation and diffuse of the electronic cloud. When forming bond, electron transfer occurs, leading to poor uniformity of the electron cloud. When losing an electron, the electron cloud of the atom would be contracted. Such kind of deformation of the electron cloud can be easily described possibly without using the polarization function. However, for the atom which has accepted some extra electrons, the electron cloud becomes deformed and expanded as well as diffused sometimes. For such a case, the use of polarization and diffuse functions is necessary. Therefore, an appropriate amount of polarization and diffuse functions could be selected for the atoms in a molecule according to the bonding feature. Although the environment of an atom in different molecules could be very different, the situation should be similar for the same kind of molecular systems.

In the following, the effects of various basis functions as well as polarization and diffusion functions on the geometrical structures are examined in the calculations using HF (Hartree-Fock), MP2 (MØller-Plesset) and B3LYP/DFT for some selected hydrogen-bonded systems, hydroxylates, carboxylic acids, amines, alkyl-alkene hydrocarbons, oxides and molecules or compounds involving metals. All the computations were carried out using G94W package[9].

2 Results and discussion

In this paper, geometrical optimizations have been done at various levels of theory and by using different basis sets for the following systems: $(BH_3)_2$, $(NH_3)_2$, $(H_2O)_2$, $(HF)_2$, CH_3OH, CH_3CH_2OH, CH_3COCH_3, CH_3COOH, CH_3CONH_2, $CH_2=CH-CH=O$, CH_4, CH_3CH_3, $CH_3CH_2CH_3$, $CH_2=CH_2$, $CH_2=CH-CH=CH_2$, $(HCOOH)_2H^+$, $(SiO)_2$, $(SiO)_3$, Fe^+OH_2, $Al(OH)_3$, $Al(CN)_3$, and NaCl. The geometrical structures of some of them are illustrated in fig. 1, while their corresponding geometrical parameters are listed in tables 1—4. Table 5 represents their

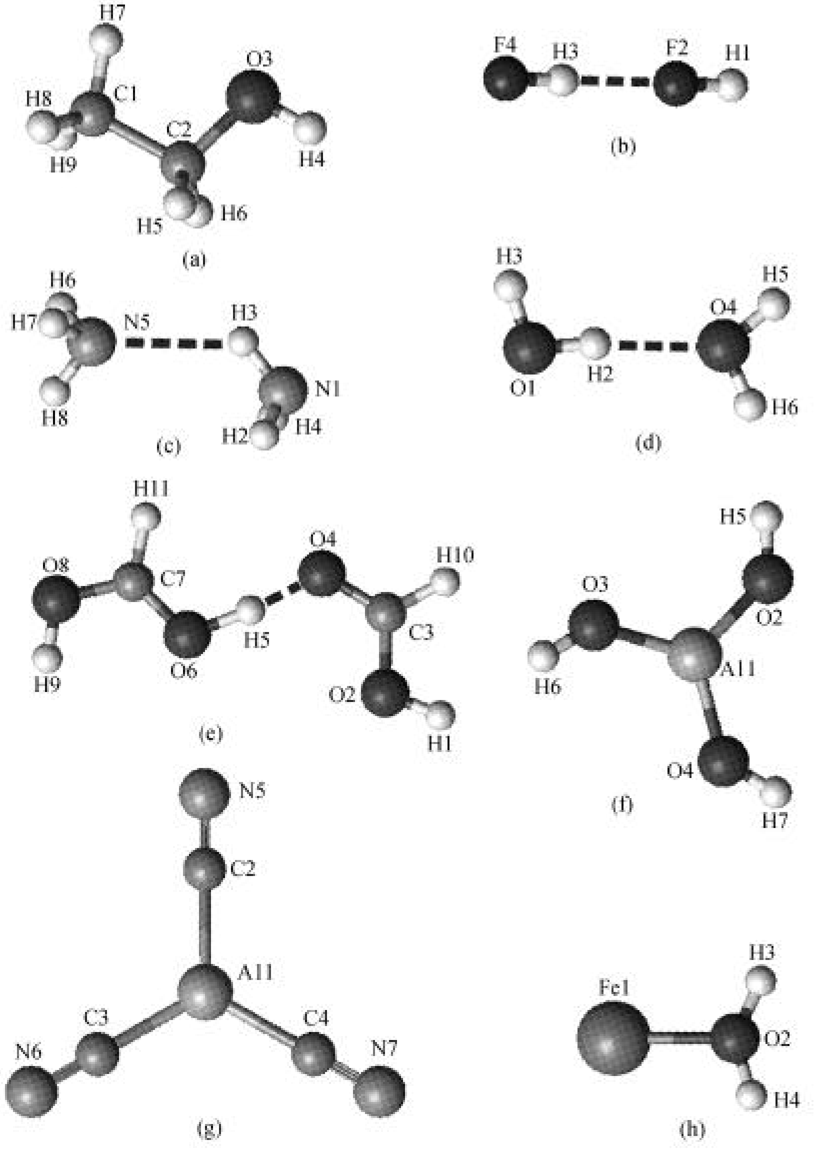

Fig. 1. The molecular structures of CH_3CH_2OH (a), $(HF)_2$ (b), $(NH_3)_2$ (c), $(H_2O)_2$ (d), $(HCOOH)_2H^+$ (e), $Al(OH)_3$ (f), $Al(CN)_3$ (g) and Fe^+OH_2 (h).

net charge distribution on atoms. It is shown that quite large charge transfers have taken place for many of the systems selected. For the atoms carrying significant negative charges, the descriptions of them obviously cannot use the basis set determined according to their performances in the neutral or slightly charged systems. The following discusses the examinations of these systems on their geometrical changes under the use of different basis sets.

For the hydride systems, H atom normally carries positive charge. Due to the less electrons which are strongly bound to the nucleus, the electron cloud distributes closely near the nucleus in

Table 1 Deviations of the bond lengths (nm) and bond angles (°) for CH_3CH_2OH (C_s) calculated using selected basis sets from those using 6-31++G** at B3LYP-DFT level

Basis set	C—C	C—O	O—H	∠COH	∠CCO
6-31++G**	0.153 45	0.143 48	0.096 57	109.04	108.04
3-21G	–0.001 18	–0.003 15	–0.003 01	1.45	1.77
6-31G	0.000 01	–0.002 86	–0.001 26	–1.15	0.8
6-31G*	–0.000 16	0.000 72	–0.000 37	–0.7	–0.26
6-31G**	–0.000 13	0.000 73	–0.000 03	1.22	0
6-31+G	0.000 21	–0.003 97	–0.001 24	–2.51	1.01
6-31++G	0.000 21	–0.003 97	–0.001 24	–2.51	1.01
O/6-31G*/C, H/6-31G	–0.000 37	–0.001 11	–0.000 47	1.62	0.37
C/6-31G*/O, H/6-31G	–0.000 05	0.000 57	–0.001 22	–1.57	0.25
C2, O3/6-31G*/C1, H/6-31G	–0.000 33	0.000 83	–0.000 36	1.14	0.06
C2, O3/6-31G*/C1, H/3-21G	–0.000 03	0.000 83	–0.000 36	1.15	–0.11

Table 2 Deviations of a part of H…A bond lengths (nm) and related bond angles (°) of $(NH_3)_2$ (C_s), $(H_2O)_2$ (C_s) and $(HF)_2$ ($D_{\infty h}$) calculated using various basis sets from those using 6-31G**

Level	Basis set	$(NH_3)_2$					$(H_2O)_2$			$(HF)_2$
		N1—H3	N5…H3	∠312	∠531	∠653	O…H	O4—H5	∠542	F…H
HF	6-31G**	0.100 33	0.255 55	107.43	128.19	126.47	0.202 55	0.094 84	107.51	0.192 21
	3-21G	0.000 1	–0.017 8	5.17	–2.10	–2.80	–0.019 9	0.001 64	7.55	–0.000 28
	6-31G	–0.000 8	0.001 33	7.81	–0.66	–5.17	–0.013 9	0.000 12	11.83	0.011 96
	6-31G*	0.000 05	0.000 08	0.11	–0.09	0.08	–0.000 8	0.000 04	–0.05	0.023 27
MP2	6-31G**	0.101 86	0.245 17	106.78	123.79	126.58	0.195 85	0.097 07	99.91	0.185 53
	6-31G	–0.000 42	–0.001 3	7.01	3.45	–3.72	–0.007 1	0.000 35	15.03	–0.001 19
	6-31G*	0.000 03	0.000 05	0.21	0.08	0.05	–0.000 3	0.000 06	–0.01	0.003 84
B3LYP (DFT)	6-31G**	0.101 92	0.242 27	106.49	123.33	126.80	0.193 1	0.096 68	105.88	0.182 19
	6-31G	–0.000 62	–0.005 6	8.34	3.01	–4.96	–0.013 8	0.000 78	6.02	0.009 27
	6-31G*	0.000 14	–0.000 6	0.06	0.02	–0.02	0.008 67	–0.000 3	1.38	–0.000 96

Table 3 Deviations of the bond lengths (nm) and bond angles (°) for $(HCOOH)_2H^+$ (C_s) calculated using selected basis sets from the results of using 6-31G**

Level	Basis set	H1—O2	O2—C3	C3—O4	O4—H5	O4—O6	O6—C7	C7—O8	C3—H10
HF	6-31G*	0.095 26	0.129 63	0.120 02	0.154 58	0.253 09	0.123 94	0.125 89	0.108 22
	6-31G	0.000 18	0.001 41	0.002 19	–0.014 68	–0.009 46	0.001 02	0.002 09	–0.000 66
	6-31G#	0.000 08	0.001 24	0.000 65	0.001 75	0.001 53	0.000 94	0.000 96	–0.000 29
MP2	6-31G*	0.097 95	0.131 73	0.123 38	0.140 16	0.248 06	0.125 93	0.128 80	0.109 37
	6-31G	0.000 38	0.002 97	0.002 53	–0.003 79	–0.001 11	0.002 16	0.002 94	–0.000 22
	6-31G#	0.000 09	0.001 57	0.001 00	0.000 52	0.000 81	0.001 12	0.001 32	0.000 21
B3LYP (DFT)	6-31G*	0.097 62	0.131 00	0.123 04	0.133 52	0.244 52	0.125 15	0.128 75	0.109 55
	6-31G	0.000 51	0.001 53	0.001 98	–0.008 55	–0.003 42	0.001 22	0.002 22	–0.000 55
	6-31G#	0.000 08	0.001 09	0.000 73	0.000 71	0.000 74	0.000 83	0.000 94	–0.000 22

Level	Basis set	∠123	∠234	∠345	∠346	∠467	∠678	∠789	∠23H10
HF	6-31G*	113.83	121.47	128.59	134.53	104.61	121.72	114.28	117.27
	6-31G	6.16	–0.84	8.25	5.82	14.79	–0.24	5.82	1.26
	6-31G#	–0.82	–0.49	–2.47	–2.38	–1.34	–0.39	–0.79	0.34
MP2	6-31G*	111.92	120.86	120.99	125.06	107.55	122.11	111.65	118.52
	6-31G	4.34	–0.69	7.25	5.74	10.56	–0.26	4.74	0.69
	6-31G#	–0.59	–0.84	–1.87	–2.39	0.18	–0.76	–0.48	0.34
B3LYP (DFT)	6-31G*	112.45	121.32	122.72	125.91	111.12	122.51	112.07	118.54
	6-31G	4.72	–0.77	6.31	5.01	11.39	–0.12	4.41	1.12
	6-31G#	–0.71	–0.35	–1.57	–1.68	–0.91	–0.29	–0.47	0.16

6-31G#: O/6-31G*/C, H/6-31G.

Table 4 Deviations of the bond lengths (nm) and bond angles (°) for $M(AB)_3$ molecules calculated using selected basis sets from those using 6-31++G**at B3LYP level

Basis sets	$Al(OH)_3$ (C_{3h})			$Al(CN)_3$ (D_{3h})		Fe^+-OH_2 (C_{2v})		
	M = Al	A = O	B = H	M = Al A = C B = N		M = Fe^+	A = O	B = H
	Al—O	O—H	∠AlOH	Al—C	C—N	Fe—O	O—H	∠HOH
6-31++G**	0.170 82	0.096 14	123.45	0.191 13	0.116 53	0.210 74	0.097 10	107.60
3-21G	0.007 90	–0.001 61	–6.92	–0.000 81	–0.000 34	0.011 40	–0.002 42	–0.91
6-31G	0.000 14	0.000 22	–16.03	–0.000 38	–0.001 07	0.004 18	–0.000 84	–2.72
6-31G*	0.000 12	–0.000 39	4.26	0.000 26	0.000 04	0.001 92	–0.000 59	0.55
6-31G**	0.000 33	0.0	3.49	0.000 26	0.000 04			
6-31+G	–0.000 08	0.000 19	–20.50	–0.000 69	–0.001 02	0.003 07	–0.001 02	–2.96
6-31++G	–0.000 08	0.000 19	–20.50	–0.000 69	–0.001 02			
6-31+G*	–0.000 19	–0.000 40	0.75	0.0	0.0	0.000 73	–0.000 73	0.0
A/6-31G*/M,B/6-31G	–0.000 70	–0.000 40	4.12	–0.001 36	–0.000 25	0.001 60	–0.000 61	0.38
A/6-31G*/M,B/3-21G	–0.000 58	–0.000 57	2.76	–0.000 67	0.000 36			
M/6-31G*/A,B/6-31G	0.001 31	–0.000 17	–10.81	–0.001 07	0.000 09	0.005 28	–0.000 88	–2.78
M/6-31+G/A,B/6-31G						0.004 95	–0.000 98	–3.56
A/6-31+G/M,B/6-31G						0.002 28	–0.000 96	–2.69
M,A/6-31+G/B/6-31G						0.003 37	–0.000 99	–3.15

Table 5 The net charge distribution on atoms of the selected systems from Mulliken populations calculated at HF/6-31G* level

Molecule	Net charge (a.u.)						
$(BH_3)_2$	B1: 0.13	H3: –0.05	H7: –0.03				
$(H_2O)_2$	O1: –0.72	H2: 0.36	H3: 0.33	O4: –0.67	H4: 0.35	H5: 0.35	
$(NH_3)_2$	N1: –0.54	H2: 0.05	H3: 0.28	H4: 0.04	N5: –0.54	H6: 0.26	H8: 0.28
$(HCOOH)_2H^+$	H1: 0.51	O2: –0.62	C3: 0.59	O4: –0.62	H5: 0.60	O6: –0.59	C7: 0.61
	O8: –0.55	H9: 0.53	H10: 0.24	H11: 0.31			
$(HF)_2$	H1: 0.54	F2: –0.52	H3: 0.53	F4: –0.55			
Fe^+OH_2	Fe1: 0.91	O2: –0.95	H3: 0.52				
$Al(OH)_3$	Al1: 1.38	O2: –0.92	H5: 0.46				
$Al(CN)_3$	Al1: 0.68	C2: 0.06	N5: –0.28				
CH_3COCH_3	C1: -0.57	C2: 0.49	C3: –0.57	O4: –0.53	H5: 0.17	H6: 0.21	
CH_3CH_2OH	C1: -0.52	C2: 0.01	O3: –0.74	H4: 0.44	H5: 0.14	H7: 0.20	H8: 0.17

The atomic numbers correspond to those in fig. 1. Some equivalent atoms are omitted here.

the s-orbital. Therefore, a certain kind of s-type function can give a satisfactory description. The effect of the various types of basis functions on the description of hydrogen atom will not be emphasized in this work. However, it can be reflected by many of the examples treated here. Generally speaking, for a simple system involving hydrogen, regardless of whether using HF, MP2 or B3LYP methods, the p polarization function of the H can only cause little influence on the structure. The changes of the bond length are at most about 10^{-4} nm, whereas the variation of the bond angles is normally within 1°. It is still the case for the organic systems. Such effect has been shown in many previous studies[10, 11], which coincides with the point of view of this paper. This is possibly the reason that the H is distinguished from the other elements in the assignment of basis sets in the literature.

2.1 Basis sets for alkane-alkene series

The molecules in the alkane series possess high orders of bonding and tetrahedral frameworks. Since there is no lone-pair electron, the covalent electron cloud distributes uniformly in the line of the bond. The corresponding molecular orbital can easily be described. Therefore, quite accurate results are expected by using only a small basis set. In this work, various levels of SCF methods have been used to optimize the geometric parameters with the different levels of basis sets.

For the alkane series, no matter what kind of basis set is selected, the result is almost the same. It is particularly the case for the C—H bond on which the influence of polarization and diffuse functions is very small. Because the description of H uses a spherical function, the deformation of the orbital is relatively small when H bonds with the carbon atom. It can be deduced that for all the molecules in the alkane series, the C—H bond length and ∠HCH bond angle can be accurately determined without using polarization functions and diffuse functions. For C—C bond, the addition of the polarization functions can slightly shorten the bond length to assume the accurate value. There is almost no effect for the use of diffuse functions. In fact, the carbon atoms are slightly negatively charged while the H atom shows positive charging. For example, the charge distribution of the ethane is $C^{-0.356}C^{-0.087}H^{0.10}$ at the MP2(full)/6-311++G** level of calculation. If dividing the charge distribution into groups, the molecule shows $(CH_3)^{-0.046}(CH_2)^{0.092}(CH_3)^{-0.046}$ with a very small dipole moment of 0.082 Debye. There is neither significant charge transfer nor orbital polarization in the whole molecule. For such a case, the use of a medium basis set (such as 6-31G) can produce quite accurate results.

The same trend has been found for alkenes with all three levels of calculations. For a general alkene, the polarization and diffuse functions can only have slight influence on the descriptions of the C—H bond and ∠HCH, ∠CCC and ∠HCC angles. However, for the C—C bond length, there is a certain effect of the polarization function, while no obvious effect of the diffuse function has been found, compared to that for the alkane series. Although there are π bond interactions weaker than σ bond, the corresponding p-orbitals are not diffuse and there is no polarization induced by strong charge centers. Furthermore, the angle changes between the frameworks formed by the σ bonds in the molecular plane do not influence π bond interaction between the atoms forming σ bonds. Therefore, it is not necessary to consider the diffuse functions for the description of the general C═C bonding systems. But the polarization can only influence C—C bonding slightly. In addition, the similar trend has been found in the calculations for various kinds of alkylbenzene systems.

2.2 Alcohols, ketone-aldehydes, carboxylic acids, amines and other hetero-atomic systems involving O and N atoms

In the areas of biology and materials science, the systems containing active groups such as —OH, —COOH, or —NH2, are of particular importance. Their functional groups can determine many of their properties. As mentioned above, for the $C=C$ bonding systems, the accurate geometric parameters can be predicted without using any polarization functions. However, when a $-C=$ in a molecule is substituted by one O or $=NH$ group, the double bond would become polarized. In addition, there are lone-pair electrons appearing at the end of the O or N, leading to the state of being electron-rich near the O or N and electron deficient on the C atoms. Such a situation provides good evidence for us to decide whether or not the polarization or diffuse functions are necessary. In this work, the geometric optimizations for a large number of acohols, ketone-aldehydes, carboxylic acids, amines and other organic hetero-atomic systems involving O, N, S and P atoms have been performed at the B3LYP level for calculation. Only the results for CH_3CH_2OH are listed in table 1. For CH_3CH_2OH, diffuse function can only cause slight influence on the geometric parameters. However, the addition of polarization functions to C and O can lead to results which are very close to those at 6-311++G** level. If applying polarization functions only to O atom, the results are still close to the accurate ones. On the other hand, the addition of polarization functions only on the carbon atoms can produce accurate $C-C$ and $C-O$ bond lengths while the improvement to the $O-H$ and $C-H$ bond lengths as well as the $\angle COH$ and $\angle OCH$ angles are quite small. Therefore, the analysis for the results shows that the addition of polarization function only to O atom for CH_3CH_2OH molecule can achieve quite similar results to those using polarization functions simultaneously to C, O and H atoms. Obviously, the former scheme can reduce largely the number of the basis functions. The fact does closely relate to the charge distribution in CH_3CH_2OH molecule. At B3LYP/6-31++G** level, the charge distribution is $H_3^{0.48}C^{-0.52}C^{-0.06}H_2^{0.26}O^{-0.52}H^{0.36}$. Obviously, polarization functions must be included for the description of the significantly negatively charged C and O atoms where the charge density is very high. However, the total number of electrons relating to C is less than that of O. Moreover, H atoms connected to C atoms carry a certain number of positive charges. By grouping them into CH_3 and CH_2, the charge distribution can be described as $(CH_3)^{-0.021}(CH_2)^{-0.2}O^{-0.52}H^{0.36}$. It indicates that CH_3 and CH_2 groups are quite the same as the alkanes in nature and thus can be described by only using a fair basis set. Therefore, the addition of polarization functions only to the oxygen atoms can achieve the result quite close to the accurate ones. Similar cases have been found for CH_3COCH_3 and CH_3COOH where the charge distributions were found to be $(H_3^{0.54}C^{-0.48})_2C^{0.28}O^{-0.38}$ and $H_3^{0.53}C^{-0.46}C^{0.40}O^{-0.36}O^{-0.5}H^{0.36}$. By regrouping, the charge distributions are $(CH_3)^{0.06}C^{0.28}O^{-0.38}(CH_3)^{0.06}$ and $(CH_3)^{0.07}C^{0.40}O^{-0.36}(O^{-0.50}H^{0.36})$. Therefore, the use of polarization functions only for the O atoms would produce accurate results.

The principle can still hold for some other oxide systems. In this work, geometric optimizations have also been performed for $(SiO)_3$ and $(SiO)_2$ systems and the results show that no matter which level of the first-principle theory is used, the addition of polarization functions only to the oxygen atoms can result in the data close to those using polarization functions for all atoms involved.

2.3 Protonated formic acid dimer and other systems with weak bonding

It is well known that the description of hydrogen bonding and other weak interactions is a particularly difficult topic in computational chemistry. Since the weak bonding is quite different from covalent bonding, the commonly used chemical bonding theory cannot be applied to such systems. In general, the bond length between the two weakly interacting atoms is relatively long and the corresponding orbital overlap is small. Therefore, an accurate description is normally required for such a system. Table 2 presents the bond lengths and bond angles relating to the H-bond of such kind of hydride dimers, in order to show the effects of the basis set. For $(HF)_2$, when two HF molecules bond each other, a little charge (0.01 e) transfers from F of one fragment to H of another fragment, leading to the polarization of the electron cloud in order to enhance the formed H-bond. Therefore, the H‥F bond length is relatively long (MP2/6-31G**: 0.182 19 nm) while the interaction between H and F is weak so as to be sensitive to the environment. On the other hand, the electron cloud among the H‥F has been deformed comparing to that before the formation of the H-bond so that the influence of polarization functions is found considerably large. If adding polarization functions only to all F atoms, large improvement has been found while the effect of diffuse functions is found very small. Furthermore, if the addition of polarization functions is only to the F atom forming the H-bond, quite good agreement can be achieved in comparison with the accurate geometric values. For the same reason, the treatments of the N‥H bond in $(NH_3)_2$ and the O‥H bond in $(H_2O)_2$ should follow the same way. The reasonable bond lengths of N‥H and O‥H can be obtained by only applying the polarization function to N and O atoms, respectively. Meanwhile, the addition of the polarization function can also sensitively improve the bond angles near the heavy atoms relating to the H-bond. In $(NH_3)_2$, the bond angles among the H‥N bond and other N—H bonds have changed about 7° while the bond angles among the H of the H‥N and other N—H bonds have varied around 4°. The improvement of the bond angles in $(H_2O)_2$ due to the introduction of the polarization functions is found much more significant, the value of which can reach up to 15°.

For the H-bonding system with the joining of the proton H^+, the effect of the polarization functions is much more significant. Table 3 lists some selected bond lengths and bond angles of the protonated formic acid dimer calculated at the various levels of theory with different basis sets. The geometric parameters are compared by reference to the values obtained with standard 6-31G* basis set. Big differences have been found between the results using 6-31G and those using 6-31G* while the addition of the d polarization function on O atoms has resulted in significant improvement. The most sensitive bond to the basis set is the hydrogen bond connecting the two formic acid units, i.e. O4—H5. The change of bond lengths by polarization functions from those without such a treatment is quite great (HF: 0.014 68 nm; MP2: 0.003 79 nm). In fact, O atoms carry significant negative charges while C and H carry positive charges. Therefore, the d polarization functions are much more important for O atoms. The effect of polarization functions can also be seen by examining the variations of the bond angles at various levels of basis set. When using

6-31G for calculation, the largest deviations of the bond angle (∠O4O6C7) from the value of 6-31G* can be up to 14.792°, 10.557° and 11.39° in HF, MP2 and B3LYP calculations, respectively. However, when adding the polarization d functions to O atoms, the deviations are reduced to –1.338°, 0.180° and –0.810°, respectively. The same trends have been found for the other bond angles. Therefore, the addition of the polarization d functions only to the oxygen atoms can produce very similar results to those with the polarization function added to all the atoms of O, C and H involved. In such a way, the use of the basis set has saved one third in number. With such a treatment for the protonated formic and acetic acid clusters, the geometric optimizations have been successfully performed for the systems as large as hexamer, $(HCOOH)_6H^+$. Consequently, the evaporation energy of a monomer from the series of protonated clusters has been obtained to agree well with experimental values and the dimer evaporation phenomenon has been explained reasonably[8]. Further applications of the basis set treatment have been performed on the protonated formic acid clusters with the addition of a water molecule, and satisfactory results have been achieved[12].

2.4 Metallic compounds

The compounds with the joining of metallic elements are important in general. The unique characteristics of the metallic elements and their interactions with each functional group may control some important functions of the biological and materials molecules. They also play key roles in catalytic processes. The compounds may show different properties in relation to the different electronic structures and location of the involved metallic element in the periodic table. Therefore, different treatments should be considered for use in the various theoretical calculations according to the different characteristics of the metallic atoms and the different properties of the compounds they join. In general, for the metal elements in groups ⅠA and ⅡA, the calculated geometric parameters may be quite accurate even using only medium-size basis sets excluding the polarization or diffuse functions, because these elements can easily lose their high energy level s-electron and show very small electronegativity so as to be in cation form in the compounds with high stability. For instance, for the NaCl system good agreement has been found between the calculated bond length (0.238 74 nm) at B3LYP/Cl/6-31G*/Na/6-31G level and that (0.239 07 nm) at B3LYP/Na, Cl/6-31G* level. However, for Na in an electron-rich system, such as $Na\text{-}NH_3$, the polarization functions on Na are still quite important because of the diffuse electron cloud. It is necessary to apply polarization functions to both Na and N atoms.

For metallic elements in the p-block of the periodic table, the situation can be different. For these elements, their d orbitals may be full or empty and their properties are determined by p orbitals. However, due to the differences in their main quantum numbers and the numbers of their valence electrons, different bonding properties can be found. Table 4 lists the geometric parameters optimized by using B3LYP with different levels of basis set for both $Al(OH)_3$ and $Al(CN)_3$ compounds. For $Al(OH)_3$, since Al shows property of positively charging, the attraction of the nucleus to its outer-shell electrons is very strong, leading to the compact electronic structure. On the other hand, the electron in O atoms is very rich, indicating that accurate results may be obtained using corrections with polarization functions and diffuse functions only for the oxygen atoms. The analysis of table 4 also indicates that the Al —O bond is not sensitive to the diffuse functions. However, if adding the polarization functions and diffuse functions simultaneously on the oxygen atoms, the produced geometric parameters would be very close to the one from 6-31+G*. The bond angle ∠AlOH is much sensitive to the polarization functions, particularly to the polarization functions on the oxygen atoms. For the case of adding polarization function to the oxygen atoms, quite satisfactory geometries can be obtained even using very small basis sets (e.g. 3-21G) for Al and H atoms (the improvement can be up to 20^o for ∠AlOH). It is quite obvious, since ∠AlOH is mainly determined by the electronic structures of the electron-rich O atoms. The same trends have been found for $Al(CN)_3$ compound as well. In such a case, the addition of polarization functions to Al atom can only influence the C —N bond slightly and lead to the decrease of the Al —C bond length. If applying polarization functions to C and N simultaneously, the C —N bond length is very close to the value obtained using a large basis set. Meanwhile, the Al —C bond has been stretched slightly. However, for AlR_3 (R represents alkyl), it is necessary to make correction with polarization functions.

For transition metallic elements, since their d orbitals are their valence shell and join bonding with variable spin states, their ionic radiuses are relatively large with d orbital being diffused and deformed. These kinds of metal atoms generally behave as the functional active center in many relevant compounds such as biological and materials molecules. Therefore, the basis set effect should be carefully taken into account in the calculations involving the transition metal elements. Table 4 lists the calculated results of Fe^+-OH_2 ion at 11 basis set levels. With the great charge on Fe^+, the bonding of Fe^+ with H_2O results in the charge transfer of 0.244 a.u. (at MP2(full)/6-311++G** level) from H_2O to Fe^+. This part of transferred charges is mainly distri- buted over Fe and O atoms, which strengthens the Fe —O bonding. Further this charge transfer also results in the polarization and deforms of the molecular orbitals in H_2O molecule. Thus polarization and diffuse functions must be used to describe these polarized and deformed molecular orbitals. From table 4 it can be known that the addition of diffuse functions to 6-31G basis set only yields a small effect on the bond O —H and the bond angle ∠HOH, while the addition of polarization functions to 6-31G decreases the bond angle ∠HOH by more than 3^o. In fact, this decrease of bond angle should be mainly attributed to the addition of the polarization functions of O atom. Therefore, the consideration of only O polarization functions in the corresponding calculations can give the correct geometry of ligand H_2O. However, for Fe —O bond, the situation may be different, because the Fe —O bonding is closely related to the atomic orbitals and the charge populations of Fe and O atoms. Table 4 shows that there is a large deviation between the value (0.206 56 nm) at the medium level of basis set (6-31G) and that (0.210 74 nm) at the higher level of basis set (6-311++G**).

But a great improvement (0.209 14 nm) has been obtained by taking into account the modification with polarization functions only for O. However, if only the polarization functions of Fe atom are added to 6-31G basis set, the Fe—O bond length is of great deviation (~0.005 3 nm) from the accurate value. The same regularity may be obtained by considering the effect of the diffuse functions. This analysis has indicated that in the calculations on these kinds of systems, the accurate results can be obtained by increasing polarization functions only for O. Similarly, the calculations on $FeOH_2$, $Fe^{2+}OH_2$ and $Fe^{3+}OH_2$ species have also been performed with the same methods and at levels of basis set as mentioned above. The results have demonstrated that the polarization functions of Fe and O atoms are important for $FeOH_2$, and those of only O atoms for $Fe^{2+}OH_2$ and $Fe^{3+}OH_2$ systems.

2.5 CPU time comparison

The basis set effect on the molecular geometries has been discussed above. This analysis has indicated that the additions of the polarization and diffuse functions to only a part of atoms in particular chemical surroundings can yield more accurate results for geometric optimizations. This provides scientific basis for reducing the total number of basis functions and for improving computation efficiency so as to enable the calculations of large biological and materials molecules to have a satisfactory accuracy. The decrease of the basis-set size would help to overcome the limitation in computations of the computer disk space and memory, and may significantly save CPU time. Table 6 gives the relationship between basis-set size and the CPU time (Pentium-II/2×350 personal computer) in single-point calculations with different levels of basis set for the proton-bound dimer of the formic acid. It is shown from this table that the CPU time in the calculations is very closely related to the basis-set size. As mentioned above, the calculations with consideration of polarization functions only for O can give results very close to those with polarization function for all atoms involved in molecule. However the number of basis set is consequently decreased by 41, while the computer CPU time is reduced by 1—6 times and becomes very close to that at 6-31G basis set. It is obvious that, for the calculations of larger molecular systems, the saving of computer CPU time and disk-space or memory will be very observable.

Table 6 CPU time comparison at various levels of basis set in the single-point calculations of $(HCOOH)_2H^+$

Basis set	No. of basis functions	CPU time/s[a)]		
		HF	B3LYP	MP2
6-31G	64	30	280	49
6-31G*	100	57	729	164
6-31G**	115	95	748	421
6-31+G	88	44	404	127
6-31++G	93	49	550	147
6-31++G**	144	320	1 756	1 249
O/6-31G*/C,H/6-31G	84	44	355	110
O4,O6/6-31G*/others/6-31G	74	37	352	65
O,H5/6-31G*/C,H/6-31G	84	44	383	112

a) Pentium-Ⅱ/2×350 personal computer.

3 Conclusions

The above theoretical analysis and the demonstrations with many examples can lead to the following conclusions.

(i) In general, the level of basis sets including polarization and diffuse functions of the atoms involved in the calculations should be increased from left to right according to their locations in the element periodic table and their actual increases of negative charge carried. For atoms bonded by non-polar or weak-polar covalent bonds, medium basis sets without polarization functions may be suitable. For atoms bonded with strong-polar or ionic bonds, medium basis sets without polarization functions may be used for the positively charged atoms, while large basis sets with polarization or diffuse functions may be required for the negatively charged atoms.

(ii) It should be particularly noted that for the various kinds of compounds such as large clusters formed by alcohol, carboxylic acid, water and other polar solvent molecules and the organic polymers with hetero-atoms involved in the biological and materials sciences, the large basis sets with polarization functions are generally required to use for the hetero-atoms with great electronegativity, such as N, O, P, S, etc. In contrast, for usual C and H atoms, medium basis sets without polarization functions may be qualified for accurate calculations. For the alkyl and alkene groups, the minimum or medium basis sets without polarization functions may be used. For the key groups involved in the studied systems, large basis sets with polarization and diffuse functions are necessarily adopted.

(iii) For metal atoms with metallic property, or with negative or low positive charges, polarization and diffuse functions should be included in the calculations. But for the metal atoms with a high positive charge (> +1 a.u.), the medium basis sets without polarization and diffuse functions are applicable.

(iv) For the systems with intermolecular hydrogen-bond or other weak-interaction, the rules (i) —(iii) mentioned above are still applicable to the small molecular fragment units. But for the active atoms used to form the hydrogen-bond or the weak interaction, it is necessary to use polarization and diffuse functions in the calculations.

References

1. Tang Aoqing, Yang Zhongzhi, Li Qianshu, Quantum Chemistry (in Chinese), Beijing: Science Press, 1982.
2. Xu Guangxian, Li Lemin, Wang Demin, Quantum Chemistry: Basic Principle and *Ab Initio* Calculational Method (Part B) (in Chinese), Beijing: Science Press, 1985.
3. Liao Muzhen, Wu Guoshi, Liu Honglin, *Ab Initio* Calculation Method of Quantum Chemistry (in Chinese), Beijing: Qinghua University Press, 1984.
4. Labanowski, J. K., Andzelm, J. W., Density Functional Methods in Chemistry, New York: Springer-Verlag, 1991.
5. Hehre, W. J., Ditchfield, R., Pople, J. A., Self-consistent molecular orbital methods, XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules, J. Chem. Phys., 1972, 56: 2257.
6. Duke, A. J., A Hartree-Fock study of the methyl anion and its inversion potential surface: Use of an augmented basis set for this species, Chem. Phys. Lett., 1973, 21: 275.
7. Chandrasekhar, J., Andrade, J. G., Schleyer, P. v. R., Efficient and accurate calculation of anion proton affinities, J. Am.

Chem. Soc., 1981, 103: 5609.

8. Zhang, R. Q., Lifshitz, C., *Ab initio* calculations of hydrogen-bonded carboxylic acid cluster systems: Dimer evaporations, J. Phys. Chem., 1996, 100: 960.
9. Frisch, M. J., Trucks, G. W., Head-Gordon, M. et al., GAUSSIAN 94W (Revision A), Pittsburgh, PA: Gaussian, Inc., 1994.
10. Boggs, J. E., Cordell, F. R., Accurate *ab initio* gradient calculation of the structures and conformations of some boric and fluoroboric acids, Basis-set effects on angles around oxygen, J. Mol. Struct. Theochem., 1981, 76: 329.
11. Pang, F., Pulay, P., Boggs, J. E., The structure of some nitrogen heteroaromatics, J. Mol. Struct. Theochem., 1982, 88: 79.
12. Aviyente, V., Zhang, R. Q., Varnali, T. et al., Structures and energetics of proton-bound formic acid-water clusters, Inter. J. Mass Spectrom. Ion Processes, 1997,161: 123.