Theor Chem Acc (1998) 99:192–196

Theoretical
Chemistry Accounts

Regular article

MIDI! basis set for silicon, bromine, and iodine

Jiabo Li, Christopher J. Cramer, Donald G. Truhlar

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, USA

Received: 7 January 1998 / Accepted: 7 January 1998

Abstract. The MIDI! basis set is extended to three new atoms: silicon, bromine, and iodine. The basis functions for these heteroatoms are developed from the standard 3-21G basis set by adding one Gaussian-type *d* subshell to each Si, Br, or I atom. The exponents of the *d* functions are optimized to minimize errors in the geometries and charge distributions that these basis functions yield when they are used in Hartree-Fock calculations with all atoms represented by the MIDI! basis. The MIDI! basis is defined to use five spherical *d* functions in a *d* subshell. We present a detailed comparison of such calculations to calculations employing six Cartesian *d* functions in each *d* subshell; these studies show that 5D and 6D options give very similar results for molecular geometries and dipole moments, not only for compounds containing Si, Br, and I but also for compounds containing N, O, F, P, S, and Cl. The MIDI! basis set is also tested successfully for hypervalent Si compounds.

Key words: Basis functions – Bond angle – Bond length – Charges, partial atomic – *d* polarization functions – Dipole moment

1 Introduction

The selection of an appropriate basis set is the starting point for essentially all calculations of the electronic structure of molecules. Most basis sets have the same number of *s*, *p*, and *d* subshells on each atom from a given row of the periodic table and are optimized entirely by minimizing atomic and molecular energies. We have recently presented [1] a new basis set called the MIDI! basis set. The goal in creating the MIDI! basis set was to provide a well-balanced and economical basis set that gives reasonably good molecular geometries and partial atomic charges at the Hartree-Fock level. The motivation for creating a basis set that yields accurate geometries and charge distributions at the Hartree-Fock level was presented previously [1], and we simply note here that economical Hartree-Fock calculations can serve as the bottom level of a multi-level scheme (e.g., using the usual slash notation, correlated level/big basis//HF/MIDI!) or as the starting point for solvation energy calculations on large molecules.

The MIDI! basis differs from conventional basis sets in two respects:

1. The final step of the optimization is based on molecular geometry and charge distribution rather than on energy.
2. There are fewer basis functions on C than on N, O, or F.

The MIDI! basis may be designated as a heteroatom-polarized valence-double-zeta basis set (or a heteroatom-polarized split-valence basis set), and it has been defined for H, C, N, O, F, P, S, and Cl. It has been shown [1], by comparing with calculations that include electron correlation by Møller-Plesset second-order perturbation theory (MP2 [2]) with the correlation-consistent polarized valence-double-zeta cc-pVDZ [3] basis set, that the MIDI! basis set yields more accurate geometries and partial atomic charges than does the popular heavy-atom-polarized valence-double-zeta 6-31G* [2] basis set. The present paper uses a procedure similar to that in Ref. [1] to extend the MIDI! basis set to silicon, bromine, and iodine.

Recall that the notation "*" (pronounced "star") denotes including a single set of *d* functions on all atoms with occupied valence *p* orbitals [2]. Similarly, "!" (pronounced "bang") denotes a set of *d* functions on all atoms with occupied valence *p* orbitals except C [1]. The MIDI! basis for N, O, F, P, S, and Cl was created by adding these *d* functions to the MIDI basis [4, 5] for those atoms and using the unpolarized MIDI basis for H and C. The MIDI! basis for Si, Br, and I is created by adding a single set of *d* functions to the 3-21G basis [6–8] for Si, Br, and I. To denote that the basis for Si, Br, and I heteroatoms is designed to be used with the already

---

This article is supplemented by an internet archive which can be obtained electronically from the Springer-Verlag server located at http://link.springer.de/journals/tca

Correspondence to: D.G. Truhlar, C.J. Cramer, or J. Li existing MIDI! basis for H, C, N, O, F, P, S, and Cl, the resulting basis functions for Si, Br, and I are also called MIDI! (rather than, for example, 3-21G!). By definition, the MIDI! basis set uses five linearly independent *d* functions in each set of *d* functions; this may be called the 5D option. Some electronic structure packages are restricted to using six-function Cartesian *d* sets (which are equivalent to a 5D set plus an *s* function); this may be called the 6D option. When this is used, we denote the basis as MIDI!(6D).

Section 2 presents the penalty function used for optimizing the MIDI! basis functions for the three atoms Si, Br, and I and specifies the training sets of molecules and the experimental and high-level theoretical geometries and dipole moments of these molecules. The results are described in Sect. 3. Since both 5D and 6D options for *d*-type Gaussian functions are used in various electronic structure packages, we have also made a comparison of 5D and 6D options in Sect. 4, not only for Si, Br, and I, but also for the other heteroatoms (N, O, F, P, S, and Cl). Section 5 tests the silicon basis for hypervalent compounds.

2 Computational method

The MIDI! basis set for H, C, N, O, F, P, S and Cl is defined in Ref. [1]; for those elements we added polarization functions to the MIDI basis. For the elements Si, Br, and I, we started with the standard 3–21G [6–8] basis functions, and a Gaussian-type *d* set of polarization functions was added to each of them. The exponents of the *d* functions are optimized for the three new elements so that Hartree-Fock calculations with the MIDI! basis set minimize a penalty function based on geometry and charge distribution.

To create the penalty function, we needed accurate values of the properties of a set of test molecules. In the previous MIDI! paper [1], geometries and electrostatic-fitting charges calculated at the MP2/cc-pVDZ level were used for adjusting the basis set parameters. In the present work, we instead used geometries and dipole moments.

We created a training set for each element. For silicon, a set of three neutral molecules, SiH_4, CH_3SiH_3, and H_3SiOH, is used. The training set for bromine consists of the four molecules CH_3Br, HBr, BrF, and BrCl. Three training molecules, CH_3I, HI, and ICl, were used for iodine. For the training molecules containing silicon, all geometries and the dipole moment of H_3SiOH were obtained from MP2/cc-pVDZ calculations, and the other dipole moments were obtained from experiment; the experimental references are given elsewhere [9]. For Br and I, we used only experimental data for both geometries [10, 11] and dipole moments [9]. The geometries and dipole moments used for the training set are given in Figure 1.

The penalty function was defined as:

$$\varepsilon = \left(\frac{\varepsilon_{\text{geom}}^2 + \varepsilon_{\text{dip}}^2}{2}\right)^{1/2} \tag{1}$$

where

$$\varepsilon_{\text{geom}} = \left\{ \frac{1}{N_{\text{bonds}} + N_{\text{angles}}} \left[\sum_{i=1}^{N_{\text{bonds}}} \left(\frac{R_i - R_i^{\text{std}}}{0.02\text{A}} \right)^2 + \sum_{j=1}^{N_{\text{angles}}} \left(\frac{\theta_j - \theta_j^{\text{std}}}{1\ \text{deg}} \right)^2 \right] \right\}^{1/2} \tag{2}$$

and

$$\varepsilon_{\text{dip}} = \left\{ \frac{1}{N_{\text{molecules}}} \left[\sum_{i=1}^{N_{\text{molecules}}} \left(\frac{d_i - d_i^{\text{std}}}{0.1\ \text{D}} \right)^2 \right] \right\}^{1/2} \tag{3}$$

where R_i^{std}, θ_j^{std}, and d_i^{std} denote the objective values of bond distance, bond angle, and dipole moment, respectively, A denotes Ångstrom, and D denotes a Debye unit. The denominators in Eq. (2) are chosen in the same way as in Ref. [1], that is, to reflect the desired accuracies of the individual molecular attributes. The denominator in Eq. (3) is approximately the average error of dipole moments obtained from high-quality theoretical calculations, for instance, B3LYP [12–14] with the cc-pVTZ [3] basis set. In our treatment, only symmetry-unique bond lengths and angles are included in the summations of Eq. (2). If the dipole moment of a molecule is zero by symmetry, we omit that molecule in Eq. (3).

The basis set was optimized to minimize the penalty function.

3 Results

The geometries and dipole moments were calculated at the HF/MIDI! level with Gaussian 94 [15]. The error function was calculated according to Eqs. (1–3). The results are given in Table 1. The values of the penalty function for the 3-21G basis set are also given for comparison. One can see that the penalty function ε is minimal with the *d* exponent equal to 0.4 for each of the three atoms. Interestingly, our exponent for Si, 0.4, is the same as found earlier for this element by Gordon [16].

The errors with the (unpolarized) 3–21G basis set are about 3 times larger than the values obtained with the optimized MIDI! basis set for all three elements. Table 1 also shows that the penalty function is reasonably insensitive to the exponents, while the MIDI! basis set, which has one polarization *d* set on all atoms except H and C, is considerably better than the 3-21G basis set, which has no polarization *d* set.

The MIDI! basis set for Si, Br and I is given in the internet archive of *Theoretical Chemistry Accounts* in Gaussian 94 format. This information is also available in multiple formats on the Internet at the Pacific Northwest National Laboratory's web site [17].

4 5D versus 6D

Some electronic structure programs only support the option of using six Cartesian *d* functions. Although the MIDI! basis set is defined to use the 5D option, it is interesting to compare the performance of the MIDI! basis set using both options. In general, "MIDI!"

Table 1. Error indicator as a function of exponent for *d* functions defining the MIDI! basis set

	Exponent					3-21G
	0.2	0.3	0.4	0.5	0.6	
Si						
$\varepsilon_{\text{geom}}$	0.64	0.73	0.98	1.21	1.34	3.56
$\varepsilon_{\text{dipole}}$	1.59	0.96	0.56	0.44	0.62	2.58
ε	1.21	0.85	0.79	0.91	1.04	3.11
Br						
$\varepsilon_{\text{geom}}$	1.98	1.07	0.94	0.99	1.06	3.96
$\varepsilon_{\text{dipole}}$	2.69	2.39	2.42	2.54	2.66	5.76
ε	2.36	1.85	1.84	1.93	2.03	4.94
I						
$\varepsilon_{\text{geom}}$	1.91	1.22	1.02	1.06	1.21	4.06
$\varepsilon_{\text{dipole}}$	2.35	2.37	2.44	2.54	2.66	8.08
ε	2.14	1.88	1.87	1.94	2.07	6.40

Table 2. Comparison of geometries and dipole moments: 5D versus 6D

Molecule			5D	6D	Expt.[a]
NH_3					
	Bond length (Å)	N-H	1.0185	1.0184	1.012
	Bond angle (deg)	∠HNH	104.63	104.63	106.7
	Dipole moment (D)		1.892	1.905	1.470
H_2O					
	Bond length (Å)	O-H	0.9674	0.9669	0.9575
	Bond angle (deg)	∠HOH	102.56	102.34	104.51
	Dipole moment (D)		2.041	2.067	1.850
CH_3F					
	Bond length (Å)	C-F	1.3735	1.3750	1.382
		C-H	1.0846	1.0846	1.095
	Bond angle (deg)	∠FCH	109.48	109.55	110.45
	Dipole moment (D)		1.976	1.994	1.858
CH_3SiH_3					
	Bond length (Å)	Si-H	1.4806	1.4842	1.485
		C-H	1.0897	1.0899	1.093
		Si-C	1.8836	1.8852	1.867
	Bond angle (deg)	∠HCH	108.31	108.32	107.7
		∠HSiH	108.06	108.01	108.3
	Dipole moment (D)		0.716	0.727	0.735
PH_3					
	Bond length (Å)	P-H	1.4310	1.4350	1.4200
	Bond angle (deg)	∠HPH	94.69	94.46	93.4
	Dipole moment (D)		0.789	0.802	0.574
H_2S					
	Bond length (Å)	S-H	1.3509	1.3528	1.3356
	Bond angle (deg)	∠HSH	93.21	93.06	92.12
	Dipole moment (D)		1.469	1.484	0.970
CH_3Cl					
	Bond length (Å)	C-Cl	1.8067	1.8046	1.785
		C-H	1.0787	1.0787	1.090
	Bond angle (deg)	∠FCH	108.33	108.37	110.8
	Dipole moment (D)		2.3500	2.3510	1.892
CH_3Br					
	Bond length (Å)	C-Br	1.9600	1.9503	1.936
		C-H	1.0792	1.0794	1.090
	Bond angle (deg)	∠FCH	107.81	108.06	107.0
	Dipole moment (D)		2.004	1.990	1.821
CH_3I					
	Bond length (Å)	C-I	2.1745	2.1700	2.139
		C-H	1.0795	1.0795	1.092
	Bond angle (deg)	∠ICH	107.76	107.95	106.7
	Dipole moment (D)		1.798	1.810	1.647
Average unsigned deviations:					
	Bond length:	0.002 (Å)			
	Bond angle:	0.13 (deg)			
	Dipole moment:	0.013 (D)			

[a] Refs. [9–11, 18]

without specification or "MIDI!(5D)" denotes the 5D choice, and "MIDI!(6D)" denotes that the six Cartesian *d* option is used.

Comparing 6D calculations to 5D ones is deemed important because the addition of a diffuse *s* function to such a small basis could conceivably unbalance it. Table 2 shows the optimized geometries and corresponding dipole moments obtained from Hartree-Fock calculations using the MIDI! basis set with both 5D and 6D options for a set of simple molecules containing all of the atoms for which MIDI! is defined, namely H, C, N, O, F, Si, P, S, Cl, Br, and I. The difference in the geometries obtained from the two options is very small. The average deviation of the two options for bond length is about 0.002 Å, and the average difference in bond angles is about 0.1 deg. The difference in dipole moments is also quite small; the average value of the difference is about 0.01 Debye.

Table 2 also compares the results to experiment [9–11]. The differences between 5D and 6D tend to be much smaller than the deviation of either from experiment, so it is not really significant which method is closer to experiment on the average.

As discussed elsewhere, class IV charge models predict much more accurate dipole moments than most other population analyses [9, 19]. Class IV charge models specifically map the atomic partial charges obtained from [9] Mulliken population analysis [20] or from [19] Löwdin population analysis [21] to values that better reproduce dipole moments. Since the MIDI! basis set is designed to predict reasonable atomic partial charges, we also compared Mulliken and Löwdin char-

Table 3. Comparison of Mulliken charges and Löwdin charges obtained from HF/MIDI!(5D) and HF/MIDI!(6D)

Molecule		MIDI!(5D)	MIDI!(6D)
NH_3			
Mulliken	N	−0.839	−0.816
	H	0.280	0.272
Löwdin	N	−0.573	−0.665
	H	0.191	0.222
H_2O			
Mulliken	O	−0.767	−0.746
	H	0.383	0.373
Löwdin	O	−0.558	−0.628
	H	0.279	0.319
CH_3F			
Mulliken	C	−0.008	−0.025
	F	−0.463	−0.450
	H	0.157	0.159
Löwdin	C	0.073	0.116
	F	−0.299	−0.344
	H	0.075	0.076
CH_3SiH_3			
Mulliken	C	−0.870	−0.860
	Si	0.665	0.659
	H(Si)	−0.131	−0.132
	H(C)	0.200	0.199
Löwdin	C	−0.479	−0.435
	Si	0.399	0.264
	H(Si)	−0.088	−0.060
	H(C)	0.114	0.114
PH_3			
Mulliken	P	−0.462	−0.406
	H	0.154	0.132
Löwdin	P	−0.297	−0.406
	H	0.099	0.132
H_2S			
Mulliken	S	−0.566	−0.551
	H	0.283	0.275
Löwdin	S	−0.343	−0.409
	H	0.171	0.204
CH_3Cl			
Mulliken	C	−0.411	−0.411
	Cl	−0.234	−0.234
	H	0.216	0.215
Löwdin	C	−0.208	−0.179
	Cl	−0.128	−0.157
	H	0.112	0.112
CH_3Br			
Mulliken	C	−0.518	−0.544
	Br	−0.174	−0.144
	H	0.231	0.229
Löwdin	C	−0.269	−0.234
	Br	−0.080	−0.115
	H	0.091	0.116
CH_3I			
Mulliken	I	0.045	0.055
	C	−0.761	−0.765
	H	0.238	0.237
Löwdin	I	0.035	0.015
	C	−0.399	−0.378
	H	0.122	0.121
Average unsigned deviation:			
Mulliken charge:	0.01		
Löwdin charge:	0.04		

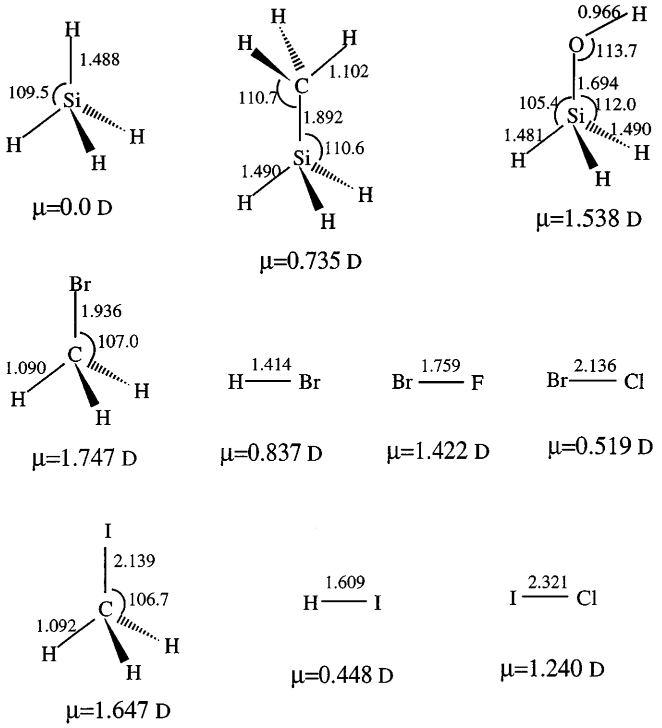

Fig. 1. The geometries (Å and deg) and dipole moments of the training set molecules for silicon, bromine, and iodine

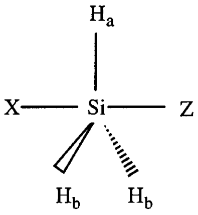

Fig. 2. Schematic structure of hypervalent compounds of silicon

ges obtained from the 5D and 6D options. The results are given in Table 3. Generally, the difference in Mulliken charges is quite small. The average deviation is about 0.01 e. The Löwdin population analysis seems to be more sensitive to the choice of 5D or 6D, and the average deviation is about 0.04 e. However, as reported elsewhere [19], the new class IV charge model, Charge Model 2 (CM2), works very well for both MIDI!(5D) and MIDI!(6D).

5 Hypervalent Si compounds

Although the new basis was not designed for hypervalent compounds, it is interesting to test its robustness for this class of structures. For this test we choose four compounds of the type H_3NSiH_3X and $HOSiH_3X^-$, where X = H or F. The former, neutral compounds were studied previously by Gordon et al. [22], and the latter, anionic species were studied previously by Damrauer et al. [23]. The geometries are defined by Fig. 2, and the results are in Table 4.

For comparison with the HF/MIDI! results in Table 4, we also show results calculated by HF/6-31G* and MP2/cc-pVDZ, where the latter is taken as a standard of higher accuracy. Table 5 shows average devia-

Table 4. Structures and dipole moments for $X{-}SiH_3{-}Z$ compounds [a,b]

X	Z	Si-X	Si-H_a	Si-H_b	Si-Z	H_a-Si-Z	H_b-Si-Z	Dipole moment
H	NH_3	1.488	1.475	1.475	2.921	73.4	73.4	3.18
		1.483	1.473	1.473	3.201	72.2	72.2	2.66
		1.495	1.485	1.485	3.007	72.7	72.7	2.82
F	NH_3	1.593	1.473	1.473	2.729	74.5	74.5	4.42
		1.608	1.468	1.468	2.819	74.6	74.6	4.66
		1.675	1.482	1.482	2.490	76.4	76.4	5.77
H	OH^-	1.631	1.538	1.547	1.777	87.4	93.6	2.93
		1.605	1.511	1.524	1.786	88.8	92.6	2.51
		1.607	1.526	1.536	1.832	86.7	92.9	2.34
F	OH^-	1.658	1.522	1.536	1.748	86.5	91.7	1.16
		1.710	1.502	1.516	1.787	87.4	91.7	1.47
		1.767	1.513	1.526	1.824	86.9	92.6	1.50

[a] The first line for each compound gives the HF/MIDI! structure and dipole moment; the second line gives the HF/6-31G* structure and dipole moment; the third line gives the MP2/cc-pVDZ results.

[b] Bond lengths are given in Å, angles in deg, and dipole moments in Debye. For the ions the dipole moment is calculated with the origin at the center of nuclear mass

Table 5. Average deviations of Hartree-Fock results from MP2/cc-pVDZ calculations for compounds with hypervalent Si

Quantity (units)	HF/MIDI	HF/6-31G*
Bond length (Å)	0.045	0.066
Bond angle (deg)	0.88	1.00
Dipole moments (D)	0.66	0.36

tions of the two sets of Hartree-Fock results from the correlated results. We see that the MIDI! basis gives more accurate geometries than the larger 6-31G* basis, but it gives less accurate dipole moments.

6 Conclusions

The MIDI! basis set is extended to three new elements: silicon, bromine, and iodine. The new basis functions are derived from the standard 3-21G basis functions by adding one Gaussian-type *d* set with exponent 0.4 for each of the three elements. The new basis set gives geometries and dipole moments that are about three times more accurate than are obtained using the 3-21G basis set without polarization *d* functions. The MIDI! basis set is reasonably stable to expanding the five *d* functions to a set of six.

Acknowledgement. This work was supported in part by the National Science Foundation.

References

1. Easton RE, Giesen DJ, Welch A, Cramer CJ, Truhlar DG (1996) Theor Chim Acta 93:281
2. Hehre WJ, Radom L, Schleyer PvR, Pople JA, Ab initio molecular orbital theory, Wiley: New York, 1986
3. a) Dunning TH Jr (1989) J Chem Phys 90: 1007; b) Woon DE, Dunning TH Jr (1993) J Chem Phys 98:1358
4. Tatewaki H, Huzinaga S (1980) J Comput Chem 1:205
5. Huzinaga S, Andzelm J, Klobukowski M, Radzio-Andzelm E, Sakai Y, Tatewaki H (1984) In: Huzinaga S (ed) Gaussian basis sets for molecular calculations. Elsevier, Amsterdam
6. Pietro WJ, Francl MM, Hehre WJ, Defrees DJ, Pople JA, Binkley JS (1982) J Am Chem Soc 104:5039
7. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ (1983) J Am Chem Soc 104:2797
8. Dobbs KD, Hehre WJ (1987) J Comput Chem 8:861, 880
9. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) J Comput Aided Mol Des 9:87
10. King WT, Mills IM, Crawford B Jr (1957) J Chem Phys 27:455
11. Huber KP, Herzberg G (ed) (1979) Molecular spectra and molecular structure, vol IV. Constants of diatomic molecules. Van Nostrand-Reinhold, New York
12. Lee C, Yang W, Parr RG (1988) Phys Rev B37:785
13. Becke AD (1993) J Chem Phys 98:5648
14. Stevens PJ, Devlin FJ, Chabalowski CF, Frisch MD (1994) J Phys Chem 98:11623
15. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman JR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JJP, Head-Gordon M, Gonzalez C, Pople JA (1995) Gaussian 94, Revision D.4. Gaussian, Inc., Pittsburgh, Pa.
16. Gordon MS (1980) Chem Phys Lett 76:163
17. Pacific Northwest National Laboratory's web site at *http://www.emsl.pnl.gov:2080/forms/basisform.html*
18. Lide DR (ed) (1995) CRC handbook of chemistry and physics, 76th edn. CRC Press, Boca Raton
19. Li J, Zhu T, Cramer CJ, Truhlar DG, J Phys Chem A, (in press)
20. Mulliken RS (1955) J Chem Phys 23:1833
21. Löwdin P-O (1950) J Chem Phys 18:365
22. Gordon MS, Davis LP, Burggraf LW (1989) Chem Phys Lett 163:371
23. Damrauer R, Burggraf LW, Davis LP, Gordon MS (1988) J Am Chem Soc 110:6601