

The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges

R. Evan Easton, David J. Giesen, Andrew Welch, Christopher J. Cramer,
Donald G. Truhlar

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

Received 30 October 1995/Final revision received 8 January 1996/Accepted 8 January 1996

Abstract. We present a series of calculations designed to identify an economical basis set for geometry optimizations and partial charge calculations on medium-size molecules, including neutrals, cations, and anions, with special emphasis on functional groups that are important for biomolecules and drug design. A new combination of valence basis functions and polarization functions, called the MIDI! basis set, is identified as a good compromise of speed and accuracy, yielding excellent geometries and charge balances at a cost that is as affordable as possible for large molecules. The basis set is optimized for molecules containing H, C, N, O, F, P, S, and Cl. Although much smaller than the popular 6-31G* basis set, in direct comparisons it yields more accurate geometries and charges as judged by comparison to MP2/cc-pVDZ calculations.

Key words: Bond length – Bond angle – Atomic partial charges – Ab initio – *d* polarization functions

1. Introduction

There are many, many basis sets available in the literature for electronic structure calculations on molecules [1], but there are few systematic comparisons of their abilities, especially if one looks for comparisons of basis sets developed by different authors, e.g., comparison of Pople group basis sets [2] to Dunning sets [3] (for a recent exception see [4]). Questions remain even for first row atoms, including the widely studied C, N, O, and F, and the relative merits of various basis sets are especially unclear for second-row atoms (Si, P, S, Cl).

It is now well recognized that good convergence of energy calculations requires both (i) very large basis sets, B , with at least triple zeta quality in the valence space plus polarization functions and (ii) higher grades, G , of electronic structure theory, for example, CCSD(T) [5]. Such calculations can be very expensive, and so an approach has arisen of optimizing the geometry with smaller basis sets B' and lower grades, G' , then performing single-point calculations for such geometries at the G/B level. This is denoted $G/B//G'/B'$. This // approach to geometry optimization is an example of a dual-level method. Dual-level approaches are also used in

dynamics, where $G/B///G'/B'$ denotes [6] calculating a reaction path at the G'/B' level and correcting the energetics along it by fixed-point calculations at level G/B . All these approaches evince a need for moderate-size basis sets that yield accurate geometries.

Another application area for moderate-size basis sets is the calculation of electrostatic potentials [7], partial charges [8–12], and electron densities of medium-size molecules, e.g., drug candidates, halogenated polyaromatic environmental toxins, and small biomolecules. In order to be useful for such calculations, a basis set must give a well-balanced description of polar bonds so that the extent of intramolecular charge redistribution is predicted reliably. Furthermore, a similar strategy to the dual-level structural ($//$) and dual-level dynamics ($///$) approaches can be used to include the effects of solvation by calculating the solute-solvent electrostatic interactions at low levels (such solvation calculations would typically be sensitive to the low-level charge distribution) and correcting the gas-phase energetics at a higher level [13].

Questions one would like answered are typified by the following: If I am studying compounds containing H, C, O, and P, how accurate are typical bond distances, bond angles, or partial charges computed at the 3-21G^(*) level [2], and is it worth the cost to go to higher levels, i.e., how much does the accuracy improve for various increases in computational cost?

The present paper is an attempt to provide guidance along these lines, and it also defines a new basis set optimized on the criteria of geometries and partial charges. Section 2 defines a large number of basis sets that we will test and establishes a notation for discussing new basis sets defined specifically for the present work. Section 3 presents computational details. Section 4 establishes a composite error index (inverse quality index), ϵ , that we will use to judge these basis sets. Section 5 presents the test sets, which are groups of molecules used to test the basis sets. Section 6 presents the tests, those conclusions from the early round of tests that were used to plan later stages, and the final choice of a new basis set. Section 7 contains further discussion. We close with a summary of major conclusions in Section 8.

An unusual aspect of the present paper is that the new basis set is optimized to reproduce results (geometries and charges) from a larger basis set and a higher level of theory rather than experimental results.

2. Basis sets

All basis sets discussed herein are of the contracted Gaussian type. When d functions are involved, all results presented in this paper are for sets of five true d functions, not six “Cartesian d functions,” one of which would transform like an s function. This is because using six d functions is more expensive and – in preliminary tests on several molecules – we found it never led to significantly better results and sometimes actually led to worse results for small, well-balanced basis sets.

We tested both existing basis sets and new ones. The existing basis sets [3, 14–26] we tested are listed in Table 1, along with one new basis set (MIDI!) discussed below. At this point we explicitly remind the reader of one confusing aspect of existing notation created by previous workers, namely the meaning of the symbol * for polarization functions, where “polarization functions” are p functions for H and d functions for B through F and Al through Cl. For calculations on first-row

Table I. Basis sets taken from the literature compared to MIDI! basis set^a

Name of set	Ref.	Number of primitive functions			Number of contracted functions			H	C	H	Highest l value	Smallest exponent of any d function in basis set						
		H	C	N-F	P-CI	H	C					N	O	F	P	S	Cl	
STO-2G(*)	14, 15, 18	2	10	10	23	0	1	1	1	1	1	1	1	1	2	2	0.39	2
		1	5	5	14	0	1	1	1	1	1	1	1	1	2	2	0.39	2
STO-3G(*)	14, 15, 18	3	15	15	29	0	1	1	1	1	1	1	1	1	2	2	0.39	2
		1	5	5	14	0	1	1	1	1	1	1	1	1	2	2	0.39	2
CEP-31G	25	4	16V ^b	16V	16V	0	1	1	1	1	1	1	1	1	1	1	1	1
		2	8V	8V	8V	0	1	1	1	1	1	1	1	1	1	1	1	1
3-21G	19, 24	3	15	15	27	0	1	1	1	1	1	1	1	1	1	1	1	1
		2	9	9	13	0	1	1	1	1	1	1	1	1	1	1	1	1
3-21G(*)	19, 22, 24	3	15	15	31	0	1	1	1	1	1	1	1	1	2	2	0.55	2
		2	9	9	18	0	1	1	1	1	1	1	1	1	2	2	0.65	2
MINI**	20, 21	6	20	20	32	1	2	0.61	2	2	2	2	2	2	2	2	0.75	2
		5	10	10	14	0	2	0.61	2	2	2	2	2	2	2	2	0.46	2
CEP-31G*	25, 26 ^c	4	21V	21V	21V	0	2	0.75	2	2	2	2	2	2	2	2	0.56	2
		2	13V	13V	13V	0	1	0.75	2	2	2	2	2	2	2	2	0.60	2
MIDI	20, 21	3	15	15	27	0	1	1	1	1	1	1	1	1	1	1	1	1
		2	9	9	13	0	1	1	1	1	1	1	1	1	1	1	1	1
MIDI!	new ^d	3	15	20	32	0	1	1	2	2	2	2	2	2	2	2	0.40	2
		2	9	14	18	0	2	0.85	0.80	0.80	1.00	1.00	1.00	1.00	2	2	0.25	2
6-31G*	16, 17, 23	4	27	27	51	0	2	2	2	2	2	2	2	2	2	2	0.65	2
		2	14	14	14	0	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	2	2	0.55	2
MIDI**	20, 21	6	20	20	32	1	2	0.80	0.80	0.80	0.80	0.80	0.80	0.80	2	2	0.65	2
		5	14	14	18	0	2	0.61	1.16	1.16	1.50	1.50	1.50	1.50	2	2	0.46	2
cc-pVDZ	3	6	39	39	93	1	2	0.61	1.16	1.16	1.50	1.50	1.50	1.50	2	2	0.46	2
		5	14	14	18	0	0.55	0.817	1.185	1.185	1.64	1.64	1.64	1.64	2	2	0.47	2

^a As explained in the text, d sets in the present paper are composed of five functions as the default, and the basis function counts in this table are all based on sets of five d functions. In Table 7, we use the symbol 6D to denote using six d functions. We note that in the GAUSSIAN code, the 3-21G(**), MINI*, MIDI*, and 6-31G* basis sets involve using six d functions

^b In tables, V after a number of basis functions means that these functions need only span the valence space, i.e., core basis functions are missing because an effective core potential is used

^c d exponents for O, P, S, and Cl are from Ref. [25]; d exponents for C, N, and F are from Ref. [26]

^d MIDI! is based on the MIDI-1 basis of Refs. [20, 21] augmented for N through Cl by polarization functions optimized in the present paper

atoms containing H, C, N, O, and F, 6-31G* was defined to denote the inclusion of six Cartesian d functions on C, N, O, and F with no polarization functions on H [17], where Cartesian d functions are equivalent to five true d functions and one s function. Later, 6-31G* was extended to Na through Ar [23], again employing Cartesian d functions. 6-31G** denotes 6-31G* plus p functions on H [17, 23]. However, in the STO-3G* basis sets, * was redefined to denote the inclusion of d functions on second-row atoms (e.g., P, S, Cl) but not C-F [18]. Furthermore, whereas the 6-31G* basis set includes six Cartesian d functions, the STO-3G* basis set includes only the five true d functions [17, 18, 23]. Further notational complexity was introduced with the 3-21G(*) basis set, in which (*) denotes the inclusion of six Cartesian d functions for second-row atoms (Al-Cl) only [22]. However, MINI-1* was defined to mean p functions on H as well as d functions on B-Ne [20]. In the present paper we are concerned only with H, C, N, O, F, P, S, and Cl, and we adopt the following notational conventions:

- * denotes d functions on C-F, P-Cl with no polarization functions on H
- ** denotes p functions on H and d functions on C-F, P-Cl.
- (*) denotes d functions on P-Cl with no polarization functions on H, C-F
- ! denotes d functions on N-F, P-Cl with no polarization functions on H or C

We pronounce * and ! as “star” and “bang”, respectively. We do not use these symbols to specify whether the d set consists of five or six functions (we recommend this always be specified separately), and in this paper the convention is that we use five true d functions in all cases. Just as * denotes the addition of d functions in a general sense but 6-31G* denotes a specific set of d exponents, we will use MIDI! to denote the specific set of d exponents optimized in this paper.

Some new basis sets are listed in Table 2. For new basis sets we use the conventions:

- $B + O$ start with existing basis set B but add an sp set on oxygen with exponential parameter 0.085
- $B[d_A 1x]$ start with existing basis B but change the exponent of the d function on A to x

Table 2. New basis sets created for this study

Notation	Existing basis in Table 1	Comments
CEP-31G $\{d_N\}$	CEP-31G	Add d on N
CEP-31G $\{d_O\}$	CEP-31G	Add d on O
CEP-31G $\{d_N d_O\}$	CEP-31G	Add d on N, O
3-21G $\{d_N\}$	3-21G	Add d on N
3-21G $\{d_O\}$	3-21G	Add d on O
3-21G $\{d_N d_O\}$	3-21G	Add d on N, O
MIDI $\{d_N\}$	MIDI-1	Add d on N
MIDI $\{d_O\}$	MIDI-1	Add d on O
MIDI $\{d_C d_N\}$	MIDI-1	Add d on C, N
MIDI!	MIDI-1	Add d on N-F, P-Cl

$B[d_A d_B x x']$	start with existing basis B but <i>change</i> the exponents of the d functions on A and B to x and x' , respectively
$B\{d_A x\}$	start with existing basis B and <i>add</i> d functions on atom A with exponent x

The meaning of $B[d_A d_B d_C |x x' x'']$ and $B\{d_A d_B |x x'\}$ and so forth should be clear by straightforward extension of these conventions. In another notational extension, $B\{d_A\}$ denotes a generic $B[d_A |x]$ basis set in the text when the precise value of x is not essential to the discussion or in a table when it is most convenient to specify x in a footnote or where several values of x are considered.

The MINI and MIDI basis sets are each defined for variable numbers of primitive functions. In this paper we always use the minimum number of primitive functions, which is technically denoted MINI-1 and MIDI-1, but we will drop the “-1” and just say MINI and MIDI, or with the optimized d functions on N-F, P-Cl, we will say MIDI! The size of the primitive basis set becomes important when one optimizes geometries because it has a significant effect on the time for integral and gradient optimization [27]. (See the appendix for a more precise specification of the MIDI! basis).

3. Computational methods

We use two methods of electronic structure theory: Hartree-Fock (HF) and Møller-Plesset second-order perturbation theory (MP2). For systems with an even number of electrons, we restricted all orbitals to be doubly occupied. For odd-electron systems we used unrestricted Hartree-Fock and unrestricted MP2, although we do not add U to the method abbreviations in this paper. Also we find it convenient in this paper, when the context makes the meaning clear, to use the popular convention that denoting a basis without a method implies HF; thus 6-31G* is synonymous with HF/6-31G*.

All electrostatic fitting [9] was carried out using the ChelpG [11] method.

All *ab initio* electronic structure calculations were carried out using the GAUSSIAN92/DFT [28] computer program.

4. Error indicator

In order to judge the qualities of the basis sets we will consider various sets of test molecules, for each of which we will optimize the geometry at the MP2/cc-pVDZ [3] level. We also calculate partial charges from the first-order perturbed wave function by electrostatic fitting. For each unique bond i , unique valence bond angle j , and unique atom k , we calculate the bond length, bond angle, and electrostatic-fitted partial charge at the the MP2/cc-pVDZ level, and we label this as the “standard” value. These are called R_i^{std} , θ_j^{std} , and q_k^{std} , respectively. (One exception: For acetate in Table 5, the accurate geometries and ChelpG partial charges are calculated at the MP2/aug-cc-pVTZ level rather than MP2/cc-pVDZ.) Note that if a bond length, bond angle, or atomic charge is identical by symmetry to one already included, we omit it from the test set.

Then we calculate bond lengths R_i , bond angles θ_i , and partial charges for the level (G/B) being tested. We calculate two kinds of partial charges, Mulliken

charges [29] q_i^M and electrostatic-fitted charges q_i^E . In our “error” analysis we will compare both of these to the electrostatic-fitted standard partial charge q_i^{std} of the standard MP2/cc-pVDZ calculation.

The motivation for comparing q_i^E to q_i^{std} is to test the quality of the lower-level wave function since the errors in electrostatic-fitted charges should reflect errors in the full continuous charge distribution. Although perhaps less obvious, there are three motivations for comparing q_i^M to q_i^{std} . First of all, Mulliken analysis gives “reasonable” partial charges only for well-balanced basis sets. Thus, comparing q_i^M to q_i^{std} is a test of basis set balance. Second, Mulliken charges are very useful for molecular dynamics and for treating very large molecules because they are less expensive than electrostatic-fitted charges, thus the comparison q_i^M to q_i^{std} is a test of the suitability of the basis set for such applications. The third motivation is that Mulliken partial charges are linearly related to the matrix elements in the basis set representation of the electron density. Thus, if one is to create a nonlinear Schrödinger equation to describe environmental effects (e.g., solvation) by adding terms to the Hamiltonian involving the interaction of an external medium with the system at hand through its partial charges [30, 31], the required derivatives of the energy functional with respect to the electron density matrix elements (which are required to derive the Fock operator) will be practical to obtain. This is essential either for using the Mulliken partial charges directly in calculations of environmental effects or for using them as the basis of class IV charge models [32] in calculations of environmental effects. (Class IV charges are obtained by a parameterized linear mapping of Mulliken charges, and one would expect the mapping to be more robust when the underlying Mulliken charges are more accurate).

Our quantitative assessment of basis-set quality will be based on a single combined error indicator based on geometries, Mulliken charges, and electrostatic fitting charges. We define a geometry error indicator $\varepsilon_{\text{geom}}$, a charge error indicator $\varepsilon_{\text{chrg}}$, and an overall error indicator, as follows:

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

$$\varepsilon_{\text{chrg}} = \left\{ \frac{1}{N_{\text{atoms}}} \sum_{k=1}^{N_{\text{atoms}}} \left[\left(\frac{q_i^E - q_i^{\text{std}}}{0.1} \right)^2 + \left(\frac{q_i^M - q_i^{\text{std}}}{0.2} \right)^2 \right] \right\}^{1/2}$$

and

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

where partial charges are measured in atomic units, and N_{bonds} , N_{angles} , and N_{atoms} are the numbers of unique (i.e., non-symmetry-identical) bonds, angles, and atoms. The denominators are chosen to reflect the desired accuracies of the individual molecular attributes.

Note that our error indicator contains no information on dihedral angles. We did examine the accuracy of dihedrals, as a result of which we concluded that including them in $\varepsilon_{\text{geom}}$ did not increase its usefulness as a quantitative indicator of geometric quality. We did, however, monitor the accuracy of dihedrals to be sure that errors were reasonably small.

5. Test molecules

We created seven test sets, each defined by the molecules constituting the set.

1. *HCNO test set no. 1*. This set includes the three neutral molecules, MeONH_2 , MeC(O)NH_2 , and CH_2NOH , shown in Fig. 1.

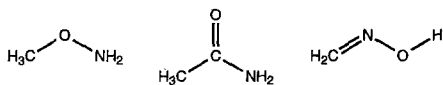


Fig. 1

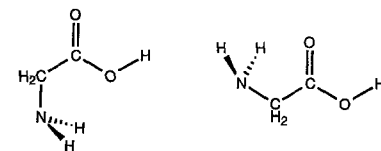
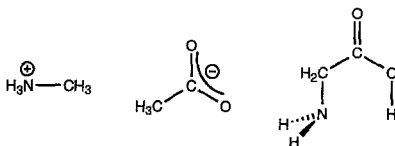


Fig. 2

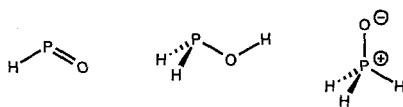
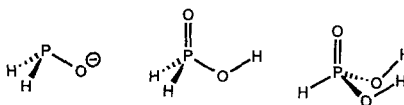
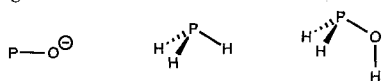


Fig. 3

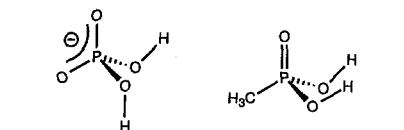


Fig. 4

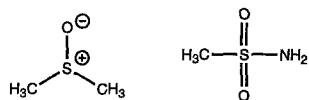
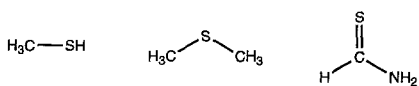
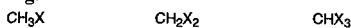


Fig. 5

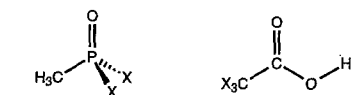
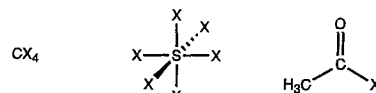


Fig. 6

Fig. 1. HCNO test set no. 1

Fig. 2. Amino acid functional group test set

Fig. 3. HOP test set

Fig. 4. Carbon-containing members of the HCOP test set

Fig. 5. HCNOS test set

Fig. 6. Halogen test set ($\text{X} = \text{F}$ or Cl)

2. *The amino acid functional group test set.* This test set consists of the molecules in Fig. 2. The set contains a carboxylate anion, a protonated amine, and three amino acid structures in non-zwitterionic form; in particular acetate anion, methylammonium cation, and three neutral glycine structures.

3. *HOP test set.* This set consists of the three compounds in Fig. 3.

4. *HCOP set.* This includes the HOP set plus the eight compounds in Fig. 4. Altogether there are three anions and eight neutrals.

5. *HCNOS test set no. 1.* This includes the five neutral molecules in Fig. 5; in particular, thioformamide, dimethyl sulfide, methanethiol, methanesulfonamide, and dimethyl sulfoxide.

6. *Fluorine and chlorine test sets.* These sets are shown in Fig. 6, where X represents F or Cl. Each test set contains eight neutral molecules.

7. *Combined HCNO test set.* Combining HCNO test set no.1 and the amino acid functional group test set yields eight molecules in all, consisting of one cation, one anion, one oxime, one amide, one *O*-alkyl hydroxylamine, and three conformations of the simplest amino acid (glycine) in its non-zwitterionic form. This will be called the combined HCNO test set.

8. *HCNOS test set no. 2.* This test set consists of methyl vinyl sulfide ($\text{CH}_3\text{SCH}=\text{CH}_2$), thiacyclobutane (trimethylene sulfide), thiacyclopropane (ethylene sulfide), methyl thiocyanate (H_3CSCN), methyl isothiocyanate (H_3CNCS), 2-mercaptoethanol ($\text{HSCH}_2\text{CH}_2\text{OH}$), and thioacetic acid ($\text{CH}_3\text{C}(\text{O})\text{SH}$).

6. Tests

A primary goal of the tests is to find an economical basis set that, at the Hartree-Fock level, gives the smallest ε (of order of magnitude unity) for the least cost. A second goal of the tests is to prepare informative tables that may be used by researchers interested in general performance characteristics of various basis sets. To meet these goals our tests proceeded in several stages.

It is well recognized that *d* functions must be included on P, S, and Cl for qualitatively correct results, and we did not test any basis set in which such functions are omitted. But we did check the importance of *d* functions on C, N, and O, using HCNO test set no. 1. The *d* exponents were taken from Pople's 6-31G* basis set, which uses 0.8 for C, N, and O. The results are given in Table 3. We here draw the conclusion that *d* functions on C and O are much less important for structural predictions than *d* functions on N. Table 3 shows that *d* functions are very important for nitrogen; however, the cases for oxygen and carbon are less clear. However, we found that bond angles of the form H-O-X are more sensitive to *d* functions on O than are bond angles of the form H-C-X to *d* functions on C. An example of this is shown in Table 4. We finally decided to include *d* functions on O but not on C.

In the next stage we again considered HCNO test set no. 1, and we used the MIDI $\{d_N d_O | 0.8 \ 0.8\}$ basis set that had been selected as the best combination of accuracy and economy on the basis of the test above. In this basis set the hydrogen

basis consists of two contracted functions made from three primitives. We considered basis sets of the form MIDI $\{d_N d_O | 0.8 \ 0.8\} [s_H s_H s_H | x x' x'']$ and we varied x , x' , and x'' simultaneously, taking them as much as 20% larger or 20% smaller than their original values. The results are shown in Table 5. Variations of 17% are seen in the error indicator, but no choice of scale factor was consistently better than 1.0 for both geometry and charges, so we decided not to vary the H exponents. We also tested simultaneous variations in the exponents and contraction coefficients (by using Gaussian expansions of hydrogenic 1s orbitals with different scaling factors) and drew the same conclusion.

At this point we had settled on the MIDI basis set plus d functions on all heteroatoms, i.e., atoms other than H or C. On the basis of past experience, we

Table 3. Results for d function tests with HCNO test set no. 1^a

Level	ϵ_{geom}	ϵ_{chrg}	ϵ
MIDI	1.22	1.34	1.31
MIDI $\{d_N\}$	0.90	1.34	1.14
MIDI $\{d_C\}$	1.29	1.22	1.25
MIDI $\{d_O\}$	1.29	1.20	1.25
MIDI $\{d_C d_N\}$	0.77	1.21	1.02
MIDI $\{d_C d_O\}$	1.27	1.13	1.21
MIDI $\{d_N d_O\}$	0.77	1.21	1.01
MIDI $\{d_C d_N d_O\}$	0.51	1.11	0.87

^a d exponents are 0.8 for all these runs

Table 4. Sensitivity of H–Y–X bond angles to d functions on Y in CH₂NOH

	HF/MIDI	HF/MIDI $\{d_Y 0.8\}$	MP2/cc-pVDZ
H–Y–X angle (deg)			
H–C–N	122.7	122.5	122.4
H–O–N	103.6	102.0	101.5

Table 5. Results for MIDI $\{d_N d_O\} [s_H s_H s_H | x x' x'']$ basis sets with HCNO test set no. 1

Scale factor on H exponents	ϵ_{geom}	ϵ_{chrg}	ϵ
0.80	0.76	0.89	0.83
0.85	0.72	0.95	0.84
0.90	0.70	1.01	0.87
0.95	0.71	1.06	0.90
1.00	0.68	1.11	0.92
1.05	0.68	1.15	0.95
1.10	0.68	1.18	0.97
1.15	0.68	1.21	0.98
1.20	0.68	1.23	1.00

dismissed the possibility of p functions on H providing an improvement commensurate with their cost, so the final recommended basis will have a double zeta s set on H, a valence double zeta sp set on C, and a polarized valence double zeta set on heavier atoms. We next set out to find optimum d exponents for N through Cl. The resulting MIDI $\{d_N d_O d_F \dots\}$ basis with our newly determined exponents will be called the MIDI! basis set. Note that it differs from the usual * type basis sets [2] in not having a d set on C. Thus, we introduce the symbol ! (“bang”) instead of * (“star”).

In order to optimize the d exponents (i.e., in order to minimize ε with respect to the d exponents) we carried out two-dimensional grid searches to find balanced values for pairs of atoms. We were particularly concerned with finding good d exponents for phosphorus and sulfur since one finds a wide variation in the literature. [E.g., the value used for phosphorus in the 6-31G* basis set is 0.55, whereas that used for phosphorus in the cc-pVDZ basis set is 0.373 (32% smaller), and the percentage deviations in some other basis sets for P and in standard basis sets for S are even larger – as discussed further in Sect. 7.]

We first performed a grid search for optimum P and O exponents on the full HCOP test set (Table 6), then we searched a grid for optimum N and O d exponents on the combined HCNO test set (not shown), and finally we searched for optimum N, O, and S d exponents on the HCNOS test set (not shown). All these tests showed low sensitivity to O d exponents with optimum values in the range 0.6–0.9. Examination of these results convinced us that it is unnecessary to optimize the various d exponents simultaneously, and on the basis of our survey we settled on 0.8 as an excellent compromise for O.

With this value fixed we successively optimized the other d exponents in this order: (1) optimize exponent for N to minimize ε in combined HCNO set (Table 7); (2) optimize P exponent to minimize ε for HCOP test set (Table 6); (3) optimize S exponent to minimize ε for HCNOS test set (Table 8); (4) optimize F exponent on fluorine test set (Table 9); (5) optimize Cl exponent on chlorine test set (Table 10). The final d exponents of the MIDI! basis set are given in Table 11. The exponent for N is larger than that for O; otherwise, the trend is monotonic within a given period.

Table 6. Results for HCOP test set

Basis set	$\varepsilon_{\text{geom}}$	$\varepsilon_{\text{chrg}}$	ε
3-21G(*)	4.18	1.79	3.22
MIDI!	0.80	0.70	0.75
MIDI $\{d_O d_P 0.64\ 0.272\}$	0.86	0.92	0.89
MIDI $\{d_O d_P 0.64\ 0.34\}$	0.98	0.90	0.94
MIDI $\{d_O d_P 0.64\ 0.408\}$	1.20	0.95	1.08
MIDI $\{d_O d_P 0.8\ 0.200\}$	0.87	0.81	0.84
MIDI $\{d_O d_P 0.8\ 0.272\}$	0.82	0.67	0.74
MIDI $\{d_O d_P 0.8\ 0.34\}$	1.00	0.64	0.84
MIDI $\{d_O d_P 0.8\ 0.408\}$	1.23	0.72	1.03
MIDI $\{d_O d_P 0.96\ 0.272\}$	0.81	0.85	0.83
MIDI $\{d_O d_P 0.96\ 0.34\}$	1.04	0.86	0.95
MIDI $\{d_O d_P 0.96\ 0.408\}$	1.34	0.92	1.15
MIDI $\{d_O d_P 0.87\ 0.25\}$	0.79	0.86	0.83

Table 7. Results for HF/MIDI $\{d_N d_O\}$ calculations on the combined HCNO test set^a

Exponents of d_N	ϵ_{geom}	ϵ_{chrg}	ϵ
0.64	0.99	1.09	1.04
0.80	0.87	1.12	1.00
0.96	0.84	1.15	1.01

^a Exponent of d_O is 0.80**Table 8.** Results for HF/MIDI $\{d_N d_O d_S\}$ calculations on HCNOS test set^a

Exponents of d_S	ϵ_{geom}	ϵ_{chrg}	ϵ
0.40	0.72	0.87	0.80
0.50	0.80	0.84	0.82
0.60	0.93	0.90	0.92

^a Exponent of d_N is 0.85, and exponent of d_O is 0.80**Table 9.** Results for HF/MIDI $\{d_F\}$ calculations on fluorine test set^a

Exponent of d_F	ϵ_{geom}	ϵ_{chrg}	ϵ
no d on F	0.89	2.92	2.16
0.20	0.82	2.67	1.98
0.40	0.79	2.22	1.67
0.60	0.87	2.28	1.73
0.80	0.89	2.23	1.69
1.00	0.85	2.23	1.69
1.20	0.82	2.28	1.71
1.40	0.79	2.34	1.74
1.60	0.76	2.40	1.78

^a Exponent of d_N is 0.85, d_O is 0.80, d_P is 0.25, and d_S is 0.40**Table 10.** Results for HF/MIDI $\{d_{Cl}\}$ calculations on chlorine test set^a

Exponent of d_{Cl}	ϵ_{geom}	ϵ_{chrg}	ϵ
0.40	0.86	1.58	1.27
0.60	0.75	1.28	1.05
0.80	0.73	1.35	1.08

^a Exponent of d_N is 0.85, d_O is 0.80, d_P is 0.25, and d_S is 0.40**Table 11.** Exponents for d functions defining the MIDI! basis set

Atom	Exponent
N	0.85
O	0.80
F	1.00
P	0.25
S	0.40
Cl	0.65

7. Further tests and discussion

It is well known from previous work [33–36] that including d functions on N is necessary for a correct description of nitrogen pyramidalty. Some work has included limited experimentation with basis sets containing polarization functions only on atoms with lone pairs [33, 34, 36]. It has also been pointed out that valence angles around oxygen are, like those around nitrogen, more sensitive to inclusion of polarization functions than are those around carbon (or boron) [37]. Our results are consistent with these studies.

Table 12 shows results for the HOP test set. In this table, N_{prim} and N_{bas} denote the number of primitive and contracted functions, respectively, for H_3PO . We conclude from this table that there is a considerable improvement in proceeding from single zeta basis sets like STO-2G^(*) and STO-3G^(*) to double zeta sets like the MIDI! basis. A very important finding is that the MIDI! basis performs *much* better than the more expensive 6-31G* basis set. The MINI-1** basis set in Table 12, which is also bigger than MIDI! for H_3PO , also performed considerably less well. Adding correlation effects at this stage does not help either, at least with the basis sets for which this was tested.

To ascertain further the relative merits of various basis sets, we made tests on the amino acid functional group test set. A particularly critical aspect of this test set is that it includes both an anion and a cation. Our tests re-examined the MIDI, 3-21G, and CEP-31G basis sets, with and without polarization functions, and they also included examining the effect of a diffuse sp set on oxygen. The results are in Table 13. For the individual molecules, the basis sets are arranged in a logical fashion for comparing the effects of adding functions. In the summary section they are arranged in order of increasing ϵ . The MIDI $\{d_{\text{N}}\}$ set performs much better than 3-21G $\{d_{\text{N}}\}$ or CEP-31G $\{d_{\text{N}}\}$. Adding the diffuse sp set on O actually made ϵ_{chrg} worse. A particularly striking aspect of these results, is that the MIDI $\{d_{\text{N}}d_{\text{O}}\}$ basis (i.e., the MIDI! basis) performs significantly better than the MIDI $\{d_{\text{C}}d_{\text{N}}\}$ basis, confirming the greater importance of d functions on O than on C.

The acetate section of Table 13 contains some MP2 calculations with additional basis sets. We note that the computer times for the MP2/cc-pVDZ and MP2/aug-cc-pVDZ calculations are about 10 times larger than those for HF/MIDI!, and the computer time for MP2/cc-pVTZ is about 170 times larger.

It is interesting to compare the d exponents in Table 11 to values in the literature. Some such comparisons are presented in Table 1. Further comparisons for P, S, and Cl are also of interest. Collins et al. [18] reported new optimized d exponents for these atoms and also summarized previous work. In their work and the previous work, the optimum d orbitals ranged over 0.14–0.50 for P, 0.35–0.58 for S, and 0.44–0.60 for Cl. Collins et al. elected to use an average value of 0.39 for all three atoms; our value for Cl is considerably higher. Later Francl et al. [23] performed further optimizations, obtaining 0.46–0.62 for P, 0.53–0.72 for S, and 0.60–0.71 for Cl; they adopted 0.55, 0.65, and 0.75, respectively. These optimizations were all based on molecular energies. Our values for P and S are significantly lower.

An issue of particular concern with valence double zeta basis sets (also called split valence basis sets), like MIDI and 6-31G, and even some valence triple zeta basis sets is whether the most diffuse functions are diffuse enough [4, 38–40]. Thus, Table 14 presents the exponential parameters of the most diffuse s and p functions for the basis considered here. The table shows that the MIDI basis functions tend to be comparable in diffuseness to the other basis sets.

Table 13. Results for various levels with the amino acid functional group test set^a

Level ^b	ϵ_{geom}	ϵ_{chrg}	ϵ	N_{prim}^c	N_{bas}^d
Acetate ^e ($\text{H}_3\text{C}_2\text{O}_2^-$)					
3-21G	0.76	0.90	0.84	69	42
3-21G + O	0.50	1.50	1.12	77	50
CEP-31G	0.56	1.66	1.24	76V	38V ^f
CEP-31G + O	0.57	1.51	1.14	84V	44V
MIDI	0.81	0.83	0.82	69	42
MIDI + O	0.44	1.37	1.02	77	50
MIDI $\{d_{\text{C}} 0.66\}$	1.22	1.23	1.23	79	52
MIDI $\{d_{\text{O}} 0.4\}$	0.82	0.89	0.86	79	52
MIDI $\{d_{\text{O}} 0.6\}$	1.25	0.88	1.08	79	52
MIDI!	1.42	0.73	1.13	79	52
MIDI $\{d_{\text{O}} 1.0\}$	1.44	0.64	1.11	79	52
MIDI $\{d_{\text{O}} 1.2\}$	1.38	0.62	1.07	79	52
MP2/cc-pVDZ	0.87	1.69	1.35	174	71
MP2/aug-cc-pVDZ	0.45	0.84	0.67		
MP2/cc-pVTZ	0.39	1.37	1.01		162
Methylammonium					
3-21G	1.07	1.76	1.46	48	30
CEP-31G	0.77	1.53	1.21	54V	28V
MIDI	1.07	1.64	1.39	48	30
3-21G $\{d_{\text{N}}\}$	0.45	1.69	1.24	53	35
CEP-31G $\{d_{\text{N}}\}$	0.55	1.57	1.18	59V	33V
MIDI!	0.54	1.68	1.25	53	35
MIDI $\{d_{\text{N}}\}$	0.54	1.70	1.26	53	35
MIDI $\{d_{\text{C}}d_{\text{N}} 0.66, 0.80\}$	0.37	1.53	1.12	58	40
Glycine (1)					
3-21G	2.65	1.48	2.15	90	55
CEP-31G	3.49	2.04	2.86	100V	50V
MIDI	2.63	1.40	2.11	90	55
3-21G $\{d_{\text{N}}\}$	2.12	1.37	1.79	95	60
CEP-31G $\{d_{\text{N}}\}$	2.67	1.82	2.29	105V	55V
MIDI $\{d_{\text{N}}\}$	2.01	1.29	1.69	95	60
3-21G + O	3.24	1.94	2.67	50	63
CEP-31G + O	3.42	1.91	2.77	108V	58V
MIDI + O	3.31	1.83	2.68	98	63
3-21G $\{d_{\text{N}}\}$ + O	2.64	1.89	2.29	103	68
CEP-31G $\{d_{\text{N}}\}$ + O	2.65	1.71	2.23	113V	63V
MIDI $\{d_{\text{N}}\}$ + O	2.63	1.81	2.26	103	68
MIDI $\{d_{\text{C}}d_{\text{N}} 0.66, 0.80\}$	1.56	1.25	1.41	105	70
MIDI $\{d_{\text{N}}d_{\text{O}} 0.8, 0.4\}$	0.61	0.90	0.77	105	70
MIDI $\{d_{\text{N}}d_{\text{O}} 0.8, 0.6\}$	0.51	0.92	0.75	105	70
MIDI $\{d_{\text{N}}d_{\text{O}} 0.8, 0.8\}$	0.58	0.88	0.74	105	70
MIDI $\{d_{\text{N}}d_{\text{O}} 0.8, 1.0\}$	0.74	0.87	0.81	105	70
MIDI $\{d_{\text{N}}d_{\text{O}} 0.8, 1.2\}$	0.91	0.90	0.90	105	70
MIDI!	0.58	0.89	0.75	105	70
Glycine (2)					
3-21G	3.87	1.46	2.92	90	55
CEP-31G	3.51	1.93	2.83	100V	50V
MIDI	2.58	1.38	2.07	90	55

Table 13. Continued

Level ^b	ϵ_{geom}	ϵ_{chrg}	ϵ	N_{prim}^c	N_{bas}^d
3-21G{ d_N }	2.45	1.40	1.99	95	60
CEP-31G{ d_N }	2.79	1.75	2.33	105V	55V
MIDI{ d_N }	2.07	1.33	1.74	95	60
3-21G + O	3.18	1.90	2.62	50	63
CEP-31G + O	3.44	1.81	2.75	108V	58V
MIDI + O	3.23	1.80	2.62	98	63
3-21G{ d_N } + O	2.70	1.89	2.33	103	68
CEP-31G{ d_N } + O	2.78	1.65	2.28	113V	63V
MIDI{ d_N } + O	2.67	1.80	2.28	103	68
MIDI{ $d_C d_N$ 0.66, 0.80}	1.68	1.26	1.48	105	70
MIDI{ $d_N d_O$ 0.8, 0.4}	0.85	0.98	0.92	105	70
MIDI{ $d_N d_O$ 0.8, 0.6}	0.73	1.02	0.89	105	70
MIDI{ $d_N d_O$ 0.8, 0.8}	0.79	0.98	0.89	105	70
MIDI{ $d_N d_O$ 0.8, 1.0}	0.95	0.97	0.96	105	70
MIDI{ $d_N d_O$ 0.8, 1.2}	1.09	0.99	1.04	105	70
MIDI!	0.79	0.99	0.90	105	70
Glycine (3)					
3-21G	2.03	1.34	1.72	90	55
CEP-31G	3.13	1.54	2.47	100V	50V
MIDI	2.03	1.12	1.64	90	55
3-21G{ d_N }	2.04	1.28	1.71	95	60
CEP-31G{ d_N }	3.04	1.60	2.43	105V	55V
MIDI{ d_N }	2.05	1.18	1.67	95	60
3-21G + O	2.28	1.55	1.95	98	63
CEP-31G + O	3.10	1.39	2.40	108V	58V
MIDI + O	2.35	1.44	1.95	98	63
3-21G{ d_N } + O	2.38	1.63	2.04	103	68
CEP-31G{ d_N } + O	2.98	1.35	2.31	113V	63V
MIDI{ d_N } + O	2.42	1.54	2.03	103	68
MIDI{ $d_C d_N$ 0.66, 0.80}	1.90	1.12	1.56	105	70
MIDI{ $d_N d_O$ 0.8, 0.4}	0.91	1.00	0.96	105	70
MIDI{ $d_N d_O$ 0.8, 0.6}	0.71	1.05	0.89	105	70
MIDI{ $d_N d_O$ 0.8, 0.8}	0.86	1.00	0.93	105	70
MIDI{ $d_N d_O$ 0.8, 1.0}	1.05	0.98	1.01	105	70
MIDI{ $d_N d_O$ 0.8, 1.2}	1.21	0.97	1.10	105	70
MIDI!	0.84	1.01	0.93	105	70
Combined results for amino acid functional group test set (order of increasing ϵ)					
MIDI!	0.68	0.87	0.78	447	297
MIDI{ $d_N d_O$ 0.8, 0.4}	0.79	1.11	0.96	447	297
MIDI{ $d_N d_O$ 0.8, 0.6}	0.78	1.13	0.97	447	297
MIDI{ $d_N d_O$ 0.8, 0.8}	0.89	1.09	1.00	447	297
MIDI{ $d_N d_O$ 0.8, 1.0}	1.00	1.08	1.04	447	297
MIDI{ $d_N d_O$ 0.8, 1.2}	1.10	1.08	1.09	447	297
MIDI{ $d_C d_N$ 0.66, 0.80}	1.56	1.28	1.43	452	302
MIDI{ d_N }	1.79	1.30	1.57	407	257
3-21G{ d_N }	1.93	1.36	1.67	407	257
MIDI	2.14	1.31	1.77	387	237
MIDI{ d_N } + O	2.23	1.67	1.97	439	289
3-21G{ d_N } + O	2.22	1.74	2.00	439	289
CEP-31G{ d_N } + O	2.42	1.57	2.04	482V	242V

Table 13. Continued

Level ^b	ϵ_{geom}	ϵ_{chrg}	ϵ	N_{prim}^c	N_{bas}^d
3-21G	2.57	1.43	2.08	387	237
CEP-31G $\{d_N\}$	2.45	1.69	2.11	450V	236V
MIDI + O	2.60	1.64	2.18	419	269
3-21G + O	2.55	1.75	2.19	419	269
CEP-31G + O	2.87	1.66	2.35	462V	246V
CEP-31G	2.92	1.77	2.41	430V	216V

^a When not specified, all additional d exponents are 0.8

^d Results for identical basis sets are not listed (e.g. CEP-31G $\{d_N, d_O\}$ and CEP-31G $\{d_N\}$ are identical for methylammonium)

^c N_{prim} is number of primitive basis functions

^d N_{bas} is number of contracted basis functions

^e Reference level: MP2/aug-c-pVTZ

^f See footnote b in Table 1

Table 14. Exponential parameters of most diffuse s and p functions

Basis	H	C	N	O	F	P	S	Cl
s and p functions have the same exponents in the following bases:								
STO-2G	0.23	0.29	0.37	0.49	0.63	0.20	0.25	0.30
STO-3G	0.17	0.22	0.29	0.38	0.49	0.19	0.22	0.23
CEP-31G	0.18	0.11	0.16	0.20	0.25	0.081	0.10	0.13
3-21G	0.18	0.20	0.28	0.37	0.48	0.12	0.14	0.17
6-31G	0.16	0.17	0.21	0.27	0.36	0.10	0.12	0.14
s exponents are listed above p exponents in the following bases:								
MINI, MIDI ^a	0.15	0.15	0.21	0.28	0.36	0.11	0.14	0.17
		0.20	0.29	0.37	0.48	0.086	0.11	0.13
cc-pVDZ	0.12	0.16	0.22	0.30	0.39	0.12	0.16	0.19
	0.73	0.15	0.22	0.28	0.35	0.12	0.14	0.16
cc-pVTZ	0.10	0.13	0.18	0.24	0.30	0.11	0.13	0.16
	0.39	0.12	0.17	0.21	0.27	0.10	0.11	0.13

^a The MINI and MIDI basis sets employ the same primitive functions, but the MINI set is contracted as a minimal basis whereas the MIDI set is contracted as a split valence set

The good accuracy of the MIDI! basis set for charge distributions has some support in the literature. For example, Alkorta [41] found that calculations with a polarized MIDI basis give more accurate dipole moments than HF/6-31G* for sulfur compounds. Six of the compounds in his test set for which experimental dipole moments are known have carbon as well as sulfur, and we calculated the dipole moments with the MIDI! basis set. Our results are shown in Table 15. We obtain a mean unsigned deviation of only 18%, further confirming that it is reasonable to omit polarization functions on C.

To provide another test of the usefulness of the basis sets we created a new sulfur test set that includes a greater variety of carbon types, to test further the reliability of omitting d functions on carbon and to provide a test for sulfur-containing compounds similar to the critical test for phosphorus-containing compounds in Table 12. The new sulfur test set is called HCNOS test set no. 2 (see Section 5), and the results are in Table 16. For these seven compounds, none of

Table 15. Dipole moments

Molecule	MIDI!		Experiment
	N_{bas}	$\mu(D)$	$\mu(D)$
CS	27	1.34	1.98
H ₂ CS	31	2.19	1.65
CH ₃ SH	35	1.87	1.52
(CH ₃) ₂ SO	62	4.20	3.96
(CH ₃) ₂ SO ₂	76	4.78	4.43
H ₂ C-CH ₂ \ /\ SO ₂	72	4.76	4.41

Table 16. Results for HCNOS test set no. 2

		HF/3-21G(*)	HF/MIDI!	HF/6-31G(*)
CH ₃ SCH=CH ₂	ϵ_{geom}	0.81	0.84	0.82
	ϵ_{chrg}	1.16	0.60	0.91
	ϵ	1.00	0.73	0.86
H ₂ C-S H ₂ C-CH ₂	ϵ_{geom}	1.25	1.65	1.47
	ϵ_{chrg}	1.65	1.15	1.28
	ϵ	1.47	1.32	1.22
H ₂ C-CH ₂ \ /\ S	ϵ_{geom}	0.87	0.80	0.63
	ϵ_{chrg}	0.92	0.50	0.63
	ϵ	0.89	0.67	0.63
HSCH ₂ -CH ₂ OH	ϵ_{geom}	1.48	0.60	1.16
	ϵ_{chrg}	1.15	0.62	0.95
	ϵ	1.32	0.61	1.06
CH ₃ C(O)SH	ϵ_{geom}	0.96	0.77	1.14
	ϵ_{chrg}	1.30	1.23	1.02
	ϵ	1.14	1.02	1.09
H ₃ CSCN	ϵ_{geom}	1.14	1.16	1.16
	ϵ_{chrg}	1.44	0.78	1.17
	ϵ	1.30	0.99	1.17
H ₃ CNCS	ϵ_{geom}	1.08	1.26	1.12
	ϵ_{chrg}	1.95	2.16	1.25
	ϵ	1.58	1.77	1.18
7 compounds combined	ϵ_{geom}	1.08	0.95	1.02
	ϵ_{chrg}	1.39	1.08	1.07
	ϵ	1.25	1.02	1.04
First 6 compounds	ϵ_{geom}	1.07	0.90	0.99
	ϵ_{chrg}	1.31	0.87	1.05
	ϵ	1.16	0.88	1.02
7 compounds combined	N_{prim}	613	640	1025
	N_{bas}	370	390	470

which were used in the parameterization of the MIDI! basis, the MIDI! basis set outperforms both the 3-21G(*) basis set and the 6-31G* basis set.

Note that the largest error for MIDI! in Table 16 occurs for the least likely to be encountered functionality in that table, name H₃CNCS. Examining the calculations for this molecule in more detail shows that ϵ_{chrg} is much bigger than ϵ_{geom} and that ϵ_{chrg} is dominated by the deviations in the Mulliken charges, which are qualitatively different from the ChelpG charges not only for MIDI! but also for 3-21G(*) and HF/6-31G* in the case. We conclude that the deviations for this molecule may be telling us more about Mulliken analysis than about the basis sets; therefore, we have also tabulated the results without this unusual functionality. The combined error indicator for the other molecules is 0.88 for MIDI!, 1.02 for 6-31G*, and 1.16 for 3-21G(*).

Another interesting test of the new basis set concerns its ability to predict the remarkable series of bond angles in the compounds NH₃, NF₃, PH₃, and PF₃ [42]. The results are shown in Table 17. The MIDI! basis predicts these bond angles much better than the 3-21G(*) basis.

Based on test data generated here and the good success we obtained with the MIDI! basis set, we can now use it to answer an interesting question that arose in our earlier work [32] on semiempirical methods for calculating partial charges. In that work we found that the AM1 and PM3 methods give *very* different Mulliken charge distributions in phosphorus-containing compounds, but there were no reliable results available for comparison. A test using the present MP2/cc-pVDZ electrostatic-fitted partial charges is shown in Table 18. This table illustrates that

Table 17. Results for X–N–X and X–P–X bond angles

Molecule	3-21G	3-21G(*)	MIDI!	Expt. ^a
H ₃ N	112.4	112.4	104.7	106.7
F ₃ N	101.6	101.6	102.5	102.4
H ₃ P	96.1	95.0	94.7	93.4
F ₃ P	96.2	97.2	97.1	97.8

^a Ref. [42]

Table 18. Results for phosphorous partial charges

Molecule	MP2/cc-pVDZ ChelpG	MIDI! Mulliken	AM1 Mulliken	PM3 Mulliken
H ₂ PO ⁻	-0.3	-0.1	0.3	0.5
H ₂ PO ₄ ⁻	1.2	1.1	1.2	2.1
H ₂ POH(a)	-0.1	-0.1	0.2	0.8
H ₂ POH(b)	0.0	0.0	0.4	0.9
H ₂ POOH	0.9	0.6	0.6	2.5
H ₃ PO	0.8	0.1	0.3	2.7
H ₃ PO ₃	1.0	0.9	0.9	2.4
MePO(OH) ₂	1.1	1.0	0.9	2.3
PH ₃	-0.2	-0.4	-0.1	0.6
PO ⁻	-0.5	-0.4	-0.5	-0.5
Average unsigned deviation		0.2	0.3	1.0

partial charges calculated by Mulliken analysis of MIDI! wave functions are somewhat more accurate than those calculated from the AM1 semiempirical parameterized NDDO wave functions and that those calculated from the PM3 semiempirical parameterized NDDO wave functions are much less accurate than either the AM1 or MIDI! Mulliken charges.

We note that d functions on C are known to have a significant effect for calculations of non-classical structures, notably carbonium ions, and for cyclopropane rings. This should be considered in deciding whether to use MIDI! or MIDI*.

One area for future work is the extension of the MIDI! basis set to transition metals. We note that the starting point for such an extension is the basis called split STO-SET (or SSTO-SET) by Tatewaki and Huzinaga for Sc–Zn [43, 44]; the SSTO-SET was later renamed [20] MIDI-1. MIDI-1 basis sets for Ga–Cd were presented by Sakai et al. [45].

8. Concluding remarks

Traditional methods for optimizing basis functions are based entirely on energies. Modern strategies for electronic structure calculations, however, often use dual-level techniques in which geometries or reaction paths are optimized at a low level, and energies are calculated at the resulting geometries by employing a larger basis or a higher level of electron correlation. In such a case the ability of the low-level calculation to predict accurate geometries is more important than its ability to predict accurate energies. To provide useful basis sets for geometry optimization, Sargent and Hall [46] have optimized basis sets for transition metals on the basis of the difference between calculated and experimental geometries.

An even more recent development is the method of class IV charge models [32]. Such models can yield accurate partial charges by a linear mapping of the partial charges of a low-level calculation. One would expect intuitively that such mappings would be more accurate when the charge balances in the low-level calculation are more accurate to begin with. Thus, the utility of a low-level calculation for such applications may depend more on the quality of the charge balance than on the energy. For applications to large molecules it is preferable to base the mapping on Mulliken charges rather than electrostatic-fitting charges so we are motivated to seek basis sets that not only predict accurate charge balances but do so in such a way that even the partial charges predicted by Mulliken analysis are accurate.

With these motivations, the present paper is an attempt to open a new chapter in the history of basis set development by optimizing a basis set based on the criteria of geometries and partial charges rather than energies. The resulting basis set is called the MIDI! basis, and the present paper defines and tests this basis for H, C, N, O, F, P, S, and Cl.

Acknowledgments. We are grateful to Susan Gustafson and Greg Hawkins for assistance with the calculations. D.J.G. gratefully acknowledges a Kodak Fellowship. This work was supported in part by the National Science Foundation and the Army Research office.

Appendix

There is more than one MIDI basis set in the literature, and therefore we close this paper with a precise specification of the MIDI! basis to avoid potential confusion.

Table 19. Hartree–Fock energies

CH ₄	– 39.956 427
NH ₃	– 55.862 841
H ₂ O	– 75.577 181
HF	– 99.440 678
PH ₃	– 340.769 055
H ₂ S	– 396.754 496
HCl	– 457.907 038

We note first that there are four MIDI basis sets (called MIDI-1 through MIDI-4) for each atom in Refs. [20, 47]. In addition, there are several basis sets of this type in the book of Huzinaga [21]. For the present work, we used the basis in footnote d of Table 11 of Ref. [20] for H, and we took the other MIDI basis sets from the book. In particular, we used the (33/33) basis for C–F, and we used the (333/33) basis for P–Cl, contracted as in the MIDI basis sets. The resulting MIDI basis set is identical to MIDI-1 for H–F and similar to (but differing in both exponents and contraction coefficients from) one of those in Ref. [21] for P–Cl. The exponents and contraction coefficients used here for all atoms precisely match those for the basis set called simply MIDI in both the MELDF program [48] and in the Gaussian Basis Set Library of Pacific Northwest Laboratory [49]. Starting with this MIDI basis set, the MIDI! basis set is formed by adding the five-function *d* sets specified in Table 11 of the present paper. The MIDI! basis set is available on the World Wide Web [49].

As a further aid to users wishing to verify their exponents and contraction coefficients, Table 19 presents restricted Hartree–Fock energies for the hydrides of C, N, O, F, P, S, and Cl obtained using the MIDI! basis set and the GAUSSIAN94 electronic structure package. These calculations are performed at the optimized geometry for the MIDI! basis set.

References

1. Feller D, Davidson ER (1990) *Rev Comp Chem* 1:1
2. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) *Ab initio* molecular orbital theory. John Wiley, New York
3. Dunning TH Jr (1989) *J Chem Phys* 90:1007. Woon DE, Dunning TH Jr (1993) *J Chem Phys* 98:1358
4. Del Bene JE, Aue DH, Shavitt I (1992) *J Am Chem Soc* 114:1631
5. Raghavachari K, Trucks GW, Pople JA, Head-Gordan M (1989) *Chem Phys Lett* 157:479
6. Corchado JC, Espinosa-Garcia J, Hu W-P, Rossi I, Truhlar DG (1995) *J Phys Chem* 99:687
7. Politzer P, Truhlar DG (eds) (1981) *Chemical applications of atomic and molecular electrostatic potentials*. Plenum, New York
8. Reed AE, Weinstock RB, Weinhold F (1985) *J Chem Phys* 83:735
9. Chirlian LE, Francl MM (1987) *J Comput Chem* 8:894
10. Burley SK, Petsko GA (1988) *Adv Protein Chem* 39:125
11. Breneman CM, Wiberg KB (1990) *J Comput Chem* 11:361
12. Merz KM (1992) *J Comput Chem* (1995) 13:749
13. Cramer CJ, Truhlar DG (1993) *J Am Chem Soc* 115:8810
14. Hehre WJ, Stewart RF, Pople JA (1969) *J Chem Phys* 51:2657
15. Hehre WJ, Ditchfield R, Stewart RF, Pople JA (1970) *J Chem Phys* 52:2769
16. Hehre WJ, Ditchfield R, Pople JA (1972) *J Chem Phys* 56:2257

17. Hariharan PC, Pople JA (1973) *Theor Chim Acta* 28:213
18. Collins JB, Schleyer PR, Binkley JS, Pople JA (1976) *J Chem Phys* 64:5142
19. Binkley JS, Pople JA, Hehre WJ (1980) *J Am Chem Soc* 102:939
20. Tatewaki H, Huzinaga S (1980) *J Comput Chem* 1:205
21. 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
22. Pietro WJ, Francl MM, Hehre WJ, DeFrees DJ, Pople JA, Binkley JS (1982) *J Amer Chem Soc* 104:5039
23. Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ, Pople JA (1982) *J Chem Phys* 77:3654
24. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ (1983) *J Am Chem Soc* 104:2797
25. Stevens WJ, Basch H, Krauss M (1984) *J Chem Phys* 81:6026
26. Dunning TH Jr, Hay PJ (1977) In: Schaefer HF III (ed) *Methods of electronic structure theory*. Plenum, New York, p. 1
27. Pulay P, Fogarasi G, Pang F, Boggs JE (1979) *J Am Chem Soc* 101:2250
28. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Wong MW, Foresman JB, Robb MA, Head-Gordon M, Replogle ES, Gomperts R, Andres JL, Raghavachari K, Binkley JS, Gonzalez C, Fox DJ, DeFrees DJ, Baker J, Stewart JJP, Pople JA (1993) *GAUSSIAN92/DFT*. Gaussian Inc., Pittsburgh, PA
29. Mulliken RS (1955) *J Chem Phys* 23:1833
30. Cramer CJ, Truhlar DG (1995) *Rev Comp Chem* 6:1
31. Cramer CJ, Truhlar DG (in press) In: Tapia O, Bertrán J (eds) *Solvent effects and chemical reactivity*. Kluwer, Dordrecht
32. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) *J Computer-Aided Molec Des* 9:87
33. Boggs JE, Niu Z (1985) *J Comput Chem* 6:46
34. Niu Z, Boggs JE (1984) *J Mol Struct (Theochem)* 109:381
35. Wong MW, Wiberg KB (1992) *J Phys Chem* 96:668
36. Šponer J, Hobza P (1994) *J Mol Struct (Theochem)* 304:35
37. Boggs JE, Cordell FR (1981) *J Mol Struct (Theochem)* 76:329
38. Frisch MJ, Pople JA, Binkley JS (1984) *J Chem Phys* 80:3265
39. Scuseria G, Schaefer HF III (1989) *J Chem Phys* 90:3269
40. Pople JA, Head-Gordon M, Fox DJ, Raghavachari K, Curtiss LA (1989) *J Chem Phys* 90:5622
41. Alkorta I (1994) *Theor Chim Acta* 89:1
42. Gilheany DG (1994) *Chem Rev* 94:1339
43. Tatewaki H, Huzinaga S (1979) *J Chem Phys* 71:4339
44. Tatewaki H, Huzinaga S (1980) *J Chem Phys* 72:399
45. Sakai Y, Tatewaki H, Huzinaga S (1982) *J Comput Chem* 3:6
46. Sargent AL, Hall MB (1991) *J Comput Chem* 12:923
47. Sakai Y, Tatewaki H, Huzinaga S (1981) *J Comput Chem* 2:100
48. McMurchie L, Elbert S, Langhoff S, Davidson ER, Feller D, Cave R, Rawlings D, Frey R, Daasch R, Nitchie L, Phillips P, Iberle K, Jackels C (1984) *MELDF-X program suite*. Indiana University, Bloomington, IN
49. Feller D, Schuchardt K, Jones D (1995) *Extensible computational chemistry environment basis set database, Version 1.0*, Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory (<http://www.emsl.pnl.gov:2080/forms/basisform.html> and http://www.dl.ac.uk/emsl_pnl/basisform.html)