J. Phys. Chem. A **1998,** *102*, 2050−2054

Theoretical Calculation of Proton Affinities Using Basis Set Functions Defined by the Generator Coordinate Method

Nelson Henrique Morgon*

Instituto de Química, Universidade Estadual de Campinas (UNICAMP), 13083-970, Campinas-SP, Brasil, CP 6154

Received: October 9, 1997; In Final Form: December 22, 1997

Ab initio calculations have been performed to determine the molecular structure and proton affinity of a set of molecules. The basis sets were developed for pseudopotentials using the GCM procedure. This technique is potentially useful for large molecules for which similar procedures (such as the G2 method and variations) were not feasible. This method achieves performance similar to the G2 method at a lower computational cost. The mean absolute deviation and the mean deviation of the results from experimental are 3.5 and 1.7 kJ mol^{-1}, respectively, compared with 4.6 and 2.2 kJ mol^{-1} for the G2 method.

I. Introduction

Thermochemical data are among the most fundamental and useful information of chemical species which can be used to predict chemical reactivity and relative stability. Thus, it is not surprising that an important goal of computational chemistry is to predict thermochemical parameters with reasonable accuracy.

Reliability is a critical feature of any theoretical model, and for practical purposes the model should be efficient in order to be widely applicable in estimating the structure, energy, and other properties of molecules.[1]

An important thermochemical property from a theoretical and experimental point of view is the proton affinity of molecules since it represents a fundamental gas-phase thermodynamic property. Absolute values of proton affinities are not always easy to obtain and are often derived from relative measurements with respect to reference molecules. On the other hand, theoretical calculations represent one attempt to study absolute values of proton affinities and other thermochemical properties.[2] However, accurate calculations of these properties require sophisticated and high-level methods and a great amount of computational resources. This is particularly true for atoms of the second period and for calculating properties such as the proton affinity of anions,

$$A^- + H^+ \rightarrow HA \qquad (1)$$

A critical feature in these calculations is to establish the best possible basis set. This set should have enough flexibility to give a correct description of the wave function for molecular and atomic environments in ions as well as in neutral systems. This is usually achieved by increasing the basis sets with additional functions, generally diffuse functions.

A theoretical procedure that has found wide acceptability and yields results comparable to experimental values ($\approx$10 kJ mol^{-1}) is the GAUSSIAN-2 (G2) technique[3] and its subsequent derivations, G2(MP2),[4] G2(MP2,SVP),[5] etc. A typical G2 calculation requires the following steps: (1) optimization of the molecular geometry and a vibrational analysis at the HF/6-31G(d) level; (2) optimization of the molecular geometry at the MP2(Full)/6-31G(d) level; and (3) energy correction using the geometry obtained in item (2) at the following levels: QCISD(T,E4T)/6-311G(d,p), MP4/6-311+G(d,p), MP4/6-311G(2df,p), and MP2/6-311+G(3df,2p). The G2 method effectively corresponds to calculations at the QCISD(T)/6-311+G(3df,2p)//MP2/6-31(d) level with ZPE and temperature corrections using harmonic frequency analysis obtained at the HF/6-31G(d) level (scaling of 0.8929) and high-level corrections. Attempts have also been made based on density functional theory using different types of functionals (B-LYP, B-P86, B3-LYP, and B3-P86).[6] The combination of Perdew and Becke's exchange with Proynov's correlation functional has been observed to be the most effective in reproducing proton affinities in close agreement with the corresponding experimental value[7] (this work is on a very small set of systems).

TABLE 1: Optimized Discretization Parameters Obtained by the GCM for the Atoms H, C, N, O, F, Si, P, S, Cl, and Br and Calculated Atomic Energies (in au)[a]

	s			p			d			
	Ω_0	$\Delta\Omega$	N	Ω_0	$\Delta\Omega$	N	Ω_0	$\Delta\Omega$	N	E
H	−2.141	1.276	5	0.000		1				−0.499 580 6
C	−3.668	1.259	7	−2.071	1.191	5	−0.670		1	−5.405 472 8
N	−3.133	1.223	7	−1.723	1.191	5	−0.214		1	−9.733 578 3
O	−1.590	1.199	7	−1.554	1.222	5	0.085		1	−15.822 349 8
F	−1.442	1.237	7	−1.779	1.296	5	0.257		1	−24.061 283 3
Si	−2.097	1.157	7	−3.307	1.168	5	−0.968		1	−6.434 344 4
P	−2.168	1.204	7	−2.581	0.993	5	−0.427		1	−3.746 895 1
S	−2.284	0.758	7	−2.431	1.026	5	−0.842		1	−10.046 298 3
Cl	−1.854	1.200	7	−2.225	1.029	5	−0.158		1	−14.849 155 5
Br	−2.038	1.206	7	−2.318	0.986	5	−0.337		1	−13.287 314 4

[a] The calculated atomic energies were obtained using eq 4.

The objective of the present work is to introduce an alternative methodology for calculating thermochemical properties and in particular for obtaining reliable proton affinities for molecular systems. Our method makes use of the discretized version[8,9] of the generator coordinate method (GCM)[10,11] to model atomic basis sets in conjunction with a pseudopotential (*effective core potential*, *ECP*).[12−14] The basis sets defined by GCM are "similar in spirit" to the *even-tempered* basis sets of Ruedenberg and co-workers.[15]

This paper contains an elaborate description of the procedure, considering (a) construction of small basis sets that adequately describe neutral systems and anions and that allow for correction

---

* E-mail: nelson@iqm.unicamp.br.

S1089-5639(97)03281-7 CCC: $15.00

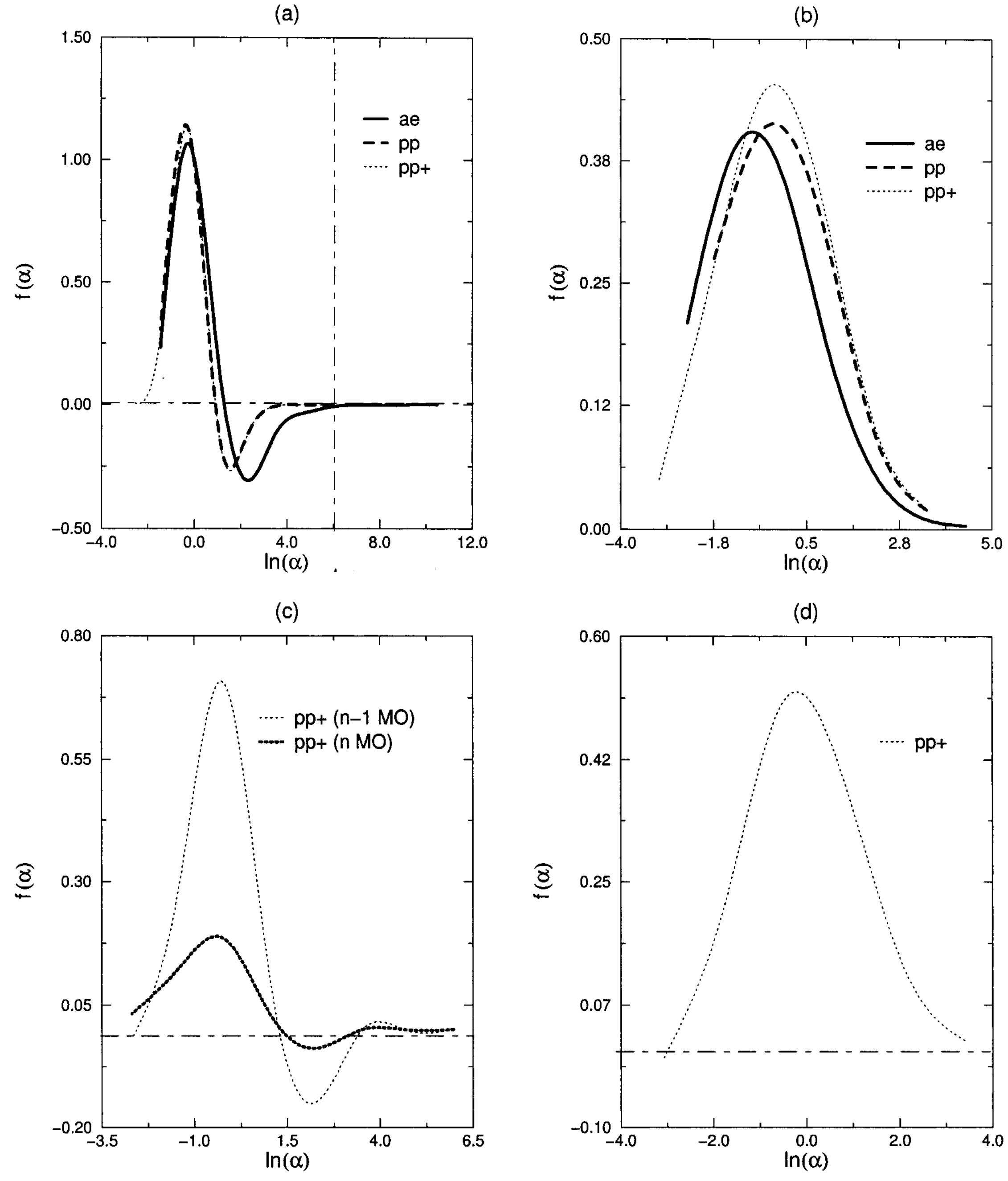

Figure 1. Weight functions for the F atomic orbitals: (a) 2s and (b) 2p for the all-electron (ae) systems, with pseudopotential (pp), and after the addition of extra functions (pp+). Weight functions for the HF outermost molecular orbitals of symmetries (c) σ and (d) π, respectively.

of the diffuse character of anionic species in a simple and systematic way; (b) derivation of adapted basis sets to be used with effective core potentials, which considerably reduced computational demands when compared with calculations employing all electrons;[16] (c) comparison of the calculated results with those obtained by more sophisticated techniques and with available experimental data; and (d) application of the present methodology to systems where use of previous techniques has not been possible because of either computational limitations or the absence of suitable basis sets.[16]

II. Computational Methods

The GCM has been very useful in the study of basis sets.[17−21] It considers the monoelectronic functions $\psi(1)$ as an integral transform,

$$\psi(1) = \int_0^\infty f(\alpha)\, \phi(\alpha, 1)\, d\alpha \qquad (2)$$

where $f(\alpha)$ and $\phi(\alpha, 1)$ are the weight and generator functions, respectively (Gaussian function used in this work) and α is the generator. The existence of the weight functions (graphical display of the linear combination of basis functions) is an essential condition for the use of GCM. Analysis of the behavior of the weight functions by the GCM permits the atomic basis set to be adapted in such a way as to yield a better description of the core electrons (represented by ECP) and the valence orbitals (corrected by addition of the extra diffusion functions), in the molecular environment.

With the exception of some simple systems the analytical expression of the weight functions is unknown. Thus, an analytical solution of the integral transform (eq 2) is not viable in most cases and suggests the need of numerical techniques to

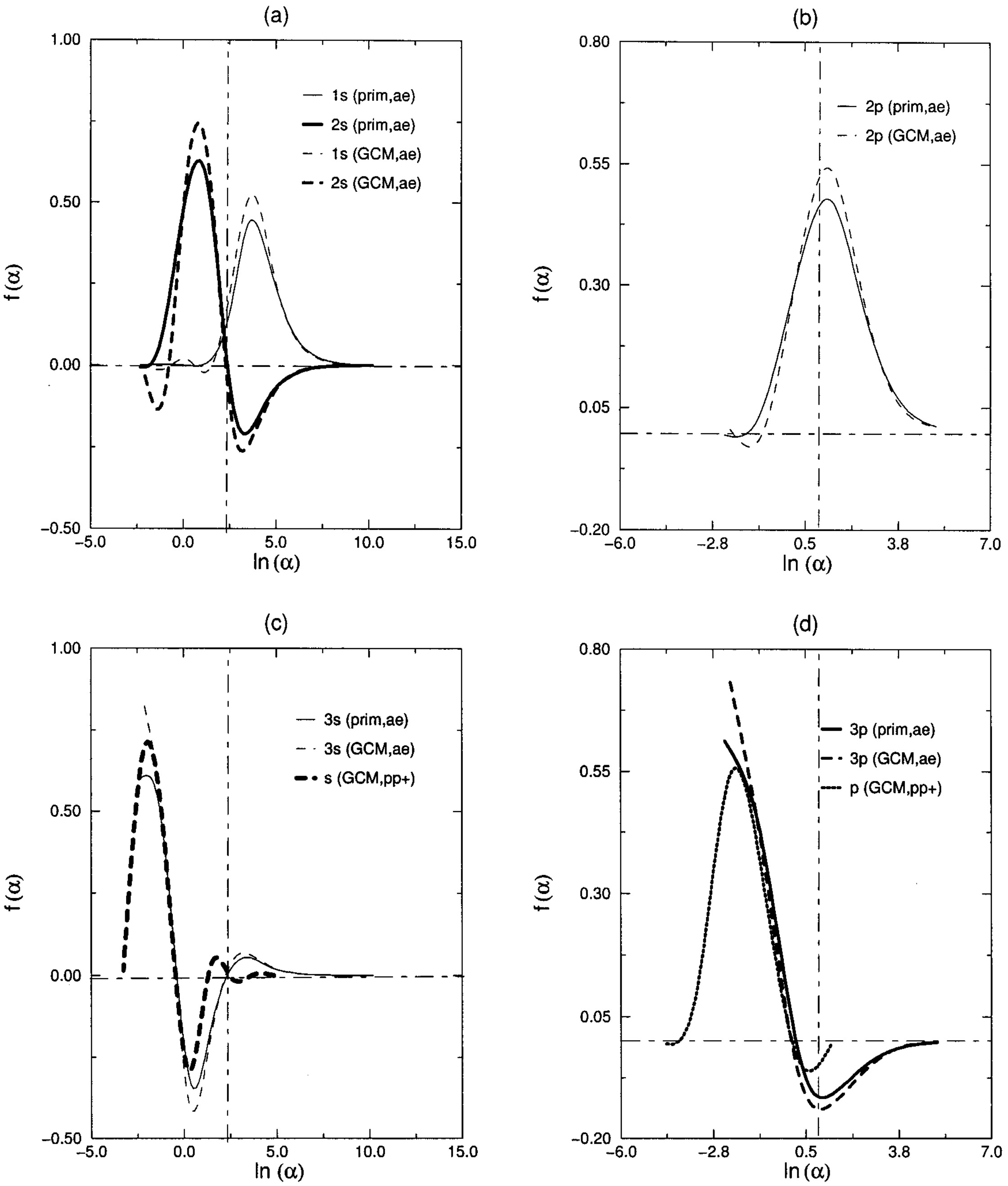

Figure 2. Weight functions of the inner atomic orbitals for Si: (a) 1s, 2s, and (b) 2p in systems for all-electron calculations (ae) and the valence atomic orbitals (c) s and (d) p from calculations using effective core potential (pp). The uncontracted DZV basis set was used to obtain the primitive basis set by GCM. The valence region was corrected by addition of extra functions (+).

solve eq 2.[8] The solution can be carried out by an appropriate choice of discrete points on the generate coordinate, represented by

$$\alpha_{i,(k)} = \exp[\Omega_{0,(k)} + (i-1)\Delta\Omega_{(k)}], \quad i = 1, 2, 3, ..., N_{(k)} \quad (3)$$

The discretization of the set is defined by the following parameters: an initial value (Ω_0), an increment ($\Delta\Omega$), and the number of primitives used (N) for a given orbital k (s, p, d, ...). The search for the best representation is obtained using the total energy of the electronic ground state as the minimization criterion.

The SIMPLEX search method[22] was adapted to the GAUSSIAN/04[23] program to provide the minimum energy of the electronic ground state of the atom corresponding to the optimized discretization parameters. The basic procedure consists of the following steps: (a) The first is a search of the optimum discretized parameter set for the atoms using the GCM for variation on the generator coordinate space. The core electrons are represented by a pseudopotential, where these discretization parameters are defined in conjunction with ECP. The minimum energy criterion is observed, and the characteristics of the atomic orbital weight functions are analyzed. These basis sets are designated as ECP/GCM. (b) The second step is optimization of the molecular geometries of the molecules at the MP2/(ECP/GCM) level. Further vibrational analysis is carried out at the equilibrium geometries. At the MP2 equilibrium geometry corrections to the total energies are performed at a higher level of theory. First, this is done to the QCISD(T)/(ECP/GCM) level and, later, by addition of extra functions (s, p, d and f). These diffuse functions are obtained by observing the convergent behavior of the weight functions of the outer atomic orbitals (s and p). Calculations are then carried out at the MP2/(ECP/GCM+) level. Thus, these results coupled to additive approximations for the energy yield an effective calculation at a high level of theory,

$$E(\mathrm{QCISD(T)/(ECP/GCM+)}) \approx E(\mathrm{QCISD(T)/(ECP/GCM)}) - E(\mathrm{MP2/(ECP/GCM)} + E(\mathrm{MP2/ECP/GCM+})) \quad (4)$$

where in (MP2/ECP/GCM+) the + sign represents the extra functions needed for the correct description of the electronic distribution in an anion (diffuse character of electronic cloud).

III. Results and Discussion

Table 1 shows the optimal atomic discretization parameters obtained according to eq 3.[24] These parameters are adapted to provide the minimum energy of the electronic ground state of the respective atom.[8] An s and one p ($\alpha = 1.0$) function were used for the H atom. The diffuse character was corrected in the molecular environment adding an s function and representing the p orbitals by $\alpha = 0.5$ and $\alpha = 2.0$.[25] The procedure employed to add diffuse functions is given in ref 26.

The atoms of the second and third periods, as shown in Table 2, were represented by a (7s5p1d) basis set. The required additional functions provided sets with (8s6p2d2f). The exponents of the f primitives are the same as those of the d primitives. For example, the oxygen atom, with the parameters $\Omega_{0(s)} = -1.590$, $\Delta\Omega_{(s)} = 1.199$, $N_{(s)} = 7$, $\Omega_{0(p)} = -1.554$, $\Delta\Omega_{(p)} = 1.222$, $N_{(p)} = 5$, $\Omega_{0(d)} = 0.085$, and $N_{(d)} = 1$, using eq 3 results in the following basis sets: $\alpha_{(s)} = \{0.205, 0.679, 2.252, 7.466, 24.755, 82.075, \text{and } 272.123\}$; $\alpha_{(p)} = \{0.211, 0.717, 2.434, 8.259, \text{and } 28.028\}$; $\alpha_{(d)} = \{1.092\}$. The extra functions are $\alpha_{(s)} = \{0.062\}$, $\alpha_{(p)} = \{0.062\}$, $\alpha_{(d)} = \{0.717\}$, and $\alpha_{(f)} = \{1.092 \text{ and } 0.717\}$.

For bromine, additional d and f type extra functions were necessary, resulting in a (8s6p3d3f) basis set.

In all cases, the process involved in obtaining the basis sets from discretization parameters by GCM and the addition of extra functions were carried out by analysis of the behavior of the weight functions of atomic orbitals at the Hartree–Fock level. Figure 1a,b displays the graphical representation of the 2s and 2p atomic orbital weight functions of the F atom. The weight functions calculated with all the electrons are drawn as a continuous line, while those using the pseudopotential are shown by a dashed line. In the valence region (small ln(α)), the weight function presents a deficiency which requires correction by the addition of extra functions (horizontal dot–dashed line). In the core region (Figure 1a) some primitive functions, $f(\alpha) \approx 0$, can be eliminated. Therefore, by analysis of the weight functions it is possible to define the cutoff (vertical dot–dashed line) from where the pseudopotential starts to act, and consequently some primitives can be eliminated.

Figure 1c,d shows the weight functions of fluorine in the molecular environment (HF molecule). In this figure, the weight functions of the outermost σ and π molecular orbitals are represented. The asymptotic behavior of the weight functions can be observed in the inner and valence regions. In the central regions, the weight functions have regular and smooth behavior, which is basic of the GCM.

Using the uncontracted double-zeta valence basis set (DZV)[31] as an example, it can be observed in Figure 2a,b a regular behavior of the weight functions of the atomic orbitals 1s, 2s, and 2p for Si (solid lines for uncontracted DZV). However, the valence orbitals (3s and 3p, Figure 2c,d, respectively) do not display the same asymptotic behavior. In the anionic molecular environment, a correct representation of the diffuse character of the electronic cloud is needed. The cutoff is defined by the GCM in conjunction with the ECP. The core electrons are represented by the pseudopotential, and addition of extra functions yields the correct description of the weight functions.

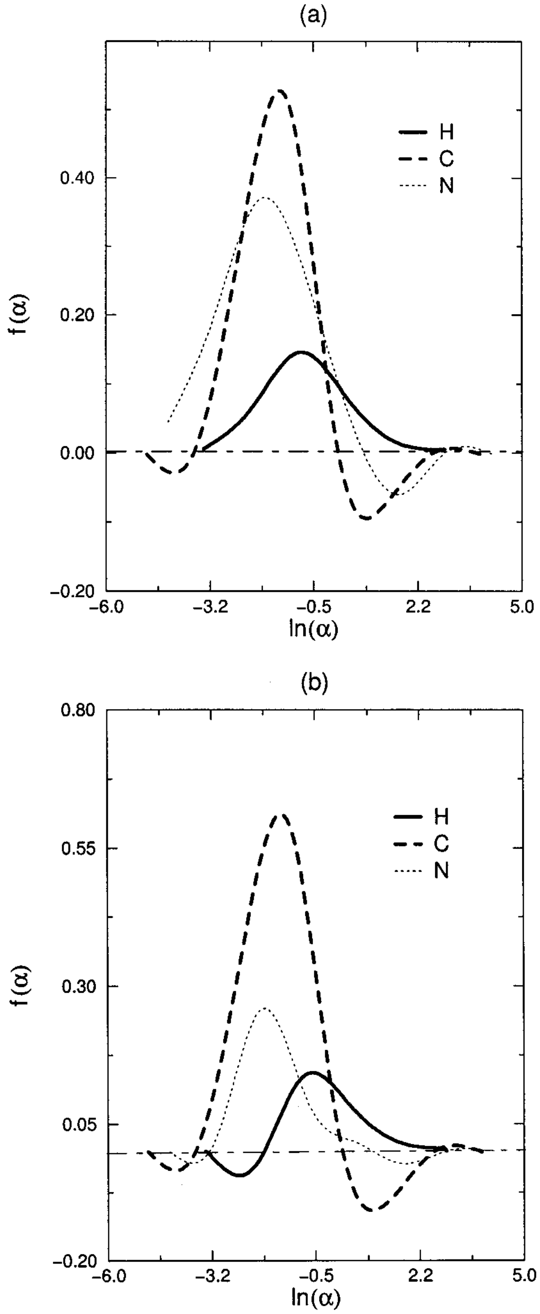

Figure 3. Weight functions of the molecular orbitals of symmetry σ to the (a) anion CH_2N^- and to the (b) neutral molecule CH_2NH, with contributions of the most important atomic orbitals of H (continuous line), C (dashed line), and N (dotted line).

Figure 3a,b is a graphical display of the weight functions for the highest molecular orbitals of symmetry σ for the CH_2N^- anion and the neutral molecule CH_2NH, respectively. The more important atomic contributions on these orbitals are observed. These weight functions are defined from the basis sets given in Table 1. It can also be observed that the weight functions are well-described in a molecular environment. This is an important finding that further lends support to the adopted procedure.

TABLE 2: Acidity (in kJ mol^{-1}) Calculated with the Basis Sets Shown in Table 1 and Comparison with Results Calculated by the G2 Method and with Experimental Values

	PA (298 K)		
system	this work	G2[a]	experimental[b]
H_2	1673 (−2)[c]	1680 (15)	1675
CH_4	1743 (−1)	1751 (7)	1744 ± 7
NH_3	1692 (3)	1690 (1)	1689 ± 3
H_2O	1635 (0)	1631 (−4)	1635
HF	1553 (−1)	1551 (−3)	1554 ± 1
SiH_4	1559 (1)	1564 (6)	1558 ± 8, 1562 ± 10
PH_3	1551 (−7)	1540 (−4)	1544 ± 6, 1552 ± 8
H_2S	1471 (1)	1473 (3)	1470 ± 2, 1469 ± 10
HCl	1397 (2)	1398 (3)	1395 ± 1
HBr	1355 (1)	1355 (1)	1354
CH_3NH_2	1681 (−6)	1688 (1)	1687 ± 5[d]
CH_3OH	1598 (1)	1601 (4)	1597 ± 3[d], 1595 ± 2
CH_3F	1710 (−1)	1717 (6)	1711 ± 17[e]
CH_3SH	1491 (−2)	1498 (5)	1493 ± 12
CH_3Cl	1656 (−1)	1665 (8)	1657 ± 15
$CH_2{=}CH_2$	1703 (−10)	1709 (−4)	1713 ± 3[d], 1699
HCCH	1571 (−10)	1579 (−2)	1581 ± 3[d], 1589 ± 2
$CH_2{=}O$	1646 (0)	1654 (8)	1646 ± 3
HCN	1461 (−8)	1466 (−3)	1469 ± 8, 1461 ± 10
HCO_2H	1439 (−5)	1437 (−7)	1444 ± 12
CH_3CHO	1539 (−6)	1537 (4)	1533 ± 12, 1531 ± 12
CH_2CO	1533 (6)	1533 (6)	1527 ± 11
CH_2NH	1631 (6)	1635 (10)	1625 ± 22

[a] Ref 6. [b] For the hydrogen anion (H^-) was used the exact energy, ref 27. [c] Absolute experimental acidities from ref 28, unless otherwise noted. The numbers in parentheses are differences between theoretical and experimental data. [d] Ref 29. [e] Ref 30.

These functions have regular asymptotic behavior in the molecular environment (mainly anionic environment).

Table 2 shows the results of the acidity (in kJ mol^{-1}) for a set of molecular systems. These results were obtained using the basis sets given in Table 1. A comparison is also presented with selected values obtained by the G2 method and with experimental results (the results of this work are corrected to give the appropriate thermal corrections, ZPE, ΔH_{vib}, ..., such as the G2 method). The mean absolute deviation and the mean deviations of the results from the experimental values are 3.5 and 1.7 kJ mol^{-1}, respectively, compared with 4.6 and 2.2 kJ mol^{-1} of G2.[6] Thus, our calculated values are very near the experimental results and the more sophisticated G2 data in most of the cases (well within 10 kJ mol^{-1}).

The present methodology, which relies on small basis sets (representation of the core electrons by ECP) and an easier and simpler way for correcting the valence regions (mainly of anionic systems), appears as an interesting alternative for the calculation of thermochemical data such as proton affinity for larger systems.

IV. Conclusion

The proton affinities of some simple systems obtained by the procedure outlined in this paper are in very good agreement with experimental values and with those obtained by sophisticated and computationally more expensive calculations.

Acknowledgment. I am grateful to Dr. José M. Riveros and Dr. Rogério Custodio for valuable discussions and critical reading of the manuscript. I would also like to thank the computational facilities of CENAPAD-SP and the financial support from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPESP (Fundação de Amparo à Pesquisa de São Paulo).

References and Notes

(1) Ochterski, J. W.; Petersson, G. A.; Wiberg, K. B. *J. Am. Chem. Soc.* **1995**, *117*, 11299.

(2) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1991**, *95*, 10549,

(3) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

(4) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 1293.

(5) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1995**, *99*, 6468.

(6) Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1995**, *245*, 123.

(7) Chandra, A. K.; Goursot, A. *J. Phys. Chem.* **1996**, *100*, 11596.

(8) Custodio, R.; Goddard, J. D.; Giordan, M.; Morgon, N. H. *Can. J. Chem.* **1992**, *70*, 580.

(9) Custodio, R.; Giordan, M.; Morgon, N. H. Goddard, J. D. *Int. J. Quantum Chem.* **1992**, *42*, 411.

(10) Mohallem, J. R.; Dreizler, R. M.; Trsic, M. *Int. J. Quantum Chem. Symp.* **1986**, *20*, 45.

(11) Mohallem, J. R. *Z. Phys. D* **1989**, *3*, 339.

(12) Stevens, W. J.; Basch, H.; Kraus, M. *J. Chem. Phys.* **1984**, *81*, 6026.

(13) Stevens, W. J.; Kraus, M.; Basch, H.; Jansien, P. *Can. J. Chem.* **1992**, *70*, 612.

(14) Cundari, T. R.; Stevens, W. J. *J. Chem. Phys.* **1993**, *98*, 5555.

(15) Feller, D. F.; Ruedenberg, K. *Theor. Chim. Acta* **1979**, *52*, 231.

(16) The GCM/ECP computational demand is reduced by ≈ 30% when compared with all-electron calculations using the same procedure. The G2 and G2MP2 calculations of systems containing first-row atoms are faster than GCM/ECP. However, for molecules containing at least one second (or third) row atom the use of GCM/ECP is faster than G2 and G2MP2. In an attempt to carry out reliable comparisons of CPU time, the Hessian matrices used in all calculations were obtained numerically.

(17) Morgon, N. H.; Custodio, R.; Riveros, J. M. *Chem. Phys. Lett.* **1995**, *235*, 436.

(18) Morgon, N. H.; Linnert, H. V.; Riveros, J. M. *J. Phys. Chem.* **1995**, *99*, 11667.

(19) Morgon, N. H. *J. Phys. Chem.* **1995**, *99*, 17832.

(20) Tanabe, F. K. J.; Morgon, N. H.; Riveros, J. M. *J. Phys. Chem.* **1996**, *100*, 2862.

(21) Morgon, N. H.; Argenton, A. B.; Silva, M. L. P.; Riveros, J. M. *J. Am. Chem. Soc.* **1997**, *119*, 1708.

(22) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308.

(23) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian/94*, Revision D.2; Gaussian, Inc.: Pittsburgh, PA, 1994.

(24) Table 1 shows all basis sets employed in this work. They are represented as parameters of a geometrical basis set in a format equivalent to that employed by Ruedenberg (ref 15) when describing the *even-tempered* basis functions. Equation 3 describes how to obtain the geometric sequence. References 8 and 9 are suggested for readers interested in more detailed information on the geometric sequence used.

(25) Since geometric sequences are being used as basis functions, the inclusion of diffuse functions is almost trivial. However, since the inclusion of these types of functions is associated with the GCM, ref 25 is suggested for further information.

(26) Morgon, N. H.; Custodio, R.; Tostes, J. G. R.; Taft, C. A. *J. Mol. Struct.* (*THEOCHEM*) **1995**, *335*, 11.

(27) Smith, B. J.; Pople, J. A.; Curtiss, L. A.; Radom, L. *Aust. J. Chem.* **1992**, *45*, 285.

(28) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data Suppl.* **1988**, *1*, 17.

(29) Ervin, K. M.; Gronert, S.; Barlow, S. E.; Gilles, M. K.; Harrison, A. G.; Bierbaum, V. M.; DePuy, C. H.; Lineberger, W. C.; Ellison, G. B. *J. Am. Chem. Soc.* **1990**, *112*, 5750.

(30) Graull, S. T.; Squires, R. R. *J. Am. Chem. Soc.* **1990**, *112*, 2517.

(31) Dunning, T. H., Jr.; Hay, P. J. In *Methods of Electronic Structure Theory*; Shaefer, H. F., III, Ed.; Plenum Press: New York, 1977; Chapter 1, pp 1−27.