Basis Set Dependency of Molecular Electronic Structures

Hiroshi ICHIKAWA,* Keiichiro SAMESHIMA, and Yukiko EBISAWA
Hoshi College of Pharmacy, Shinagawa-ku, Tokyo 142
(Received April 14, 1986)

The total atomic charges as well as the kinetic and potential energies calculated using 22 different basis sets were compared. It was shown that polarization functions must be added to atoms of the first row as well as the hydrogen atom in order to obtain well-balanced wave functions. The STO-*N*G (*N*=3, 4, 5, or 6) method, which is considered to give poor wave functions, is shown to give well-balanced wave functions.

It is unnecessary to say that ab initio molecular orbital (MO) theories have been frequently used in explaining various molecular properties in terms of electronic structures. Energy partitioning based on ab initio MO theories has been effectively applied in analyses of chemical bonds[1] and aromaticity.[2,3]

So far, quantum chemists have proposed various types of basis sets (i.e. MO theories). However, much effort in creating a new basis set has been devoted to the lowering of the total energy and to reproducing the observed geometries; the electron distribution has been treated less effectively. As a result, electron distributions can be greatly changed according to the type of basis set used. There have been method-to-method comparisons reported between different basis sets, mainly by the originaters of new basis sets. However, the number of such comparisons is too small for determining the general aspects of various MO theories. This paper is to show how the electronic structures and partitioned energies are dependent on the adopted basis set. For this purpose we have examined those quantities using 22 different basis sets in the GAUSSIAN-80 program.[4,5]

Partitioning of Molecular Energies

If wave functions are to provide an understanding of the behavior of electrons in a molecule, one must try to extract, in some way, information on the physically essential features of electronic energies. One of the trends along this line is energy partitioning.[6] Since in LCAO MO theory each MO is expanded by a linear combination of atomic orbitals (AO's), and since via the Fock matrix three- and four-center integrals out of the two-electron integrals can be reduced to two-center terms, the energy (E) based on either the restricted[7] or unrestricted Hartree-Fock equation[8] is *formally* expressed by the sum of monocentric (E_A) and bicentric (E_{AB}) terms,

$$E = \sum_{A} E_A + \sum_{A>B} E_{AB}. \quad (1)$$

These terms were further partitioned as

$$E_A = E_A^V + E_A^J + E_A^T \quad (2)$$

and

$$E_{AB} = E_{AB}^V + E_{AB}^J + E_{AB}^N + E_{AB}^T, \quad (3)$$

where E^V, E^J, and E^T are the one-electron potential, two-electron potential and kinetic energies of the electrons. E_{AB}^N is the nuclear-repulsion energy between atoms A and B and is, of course, one of the potential energies in the given system. A detailed calculation method has been given elsewhere.[3]

According to the method of Mulliken population analysis,[9] the kinetic and potential energies concerning electrons belonging to a specified atom, X, KE(X) and PE(X) are, respectively, defined as,

$$\mathrm{KE(X)} = E_X^T + 1/2 \sum_{A} E_{XA}^T \quad (4)$$

and

$$\mathrm{PE(X)} = E_X^V + E_X^J + 1/2 \sum_{A>B} (E_{XA}^V + E_{XA}^J + E_{XA}^N). \quad (5)$$

Since the kinetic and potential energies are considered to be basic quantities with apparent physical meanings, we examined them in addition to the total atomic charges of the Mulliken population analysis.

Results and Discussion

1. Total Atomic Charge. The Mulliken population analysis[9] is probably the most-widely applied method in the analyses of the electronic structures in the various fields of chemistry. It is important to note that the total atomic charge obtained in a Mulliken population analysis must be dependent on the basis set since any assignment of charges to atoms in a molecule is necessarily arbitrary. Therefore, the absolute charges calculated by a Mulliken population analysis also tend to be dependent on the basis set. First, we show how much they are dependent.

Figure 1 shows the total atomic charge of the hydrogen atoms of vinylamine by 22 different ab initio MO methods,[5] where a geometry optimization was carried out for each MO calculation. We also obtained results for the hydrogen of methane, which appears to be almost the same as those of H_a. It is clearly shown that the calculated total atomic charge varies significantly according to the calculation method. However, there seem to be some tendencies in the results. It can

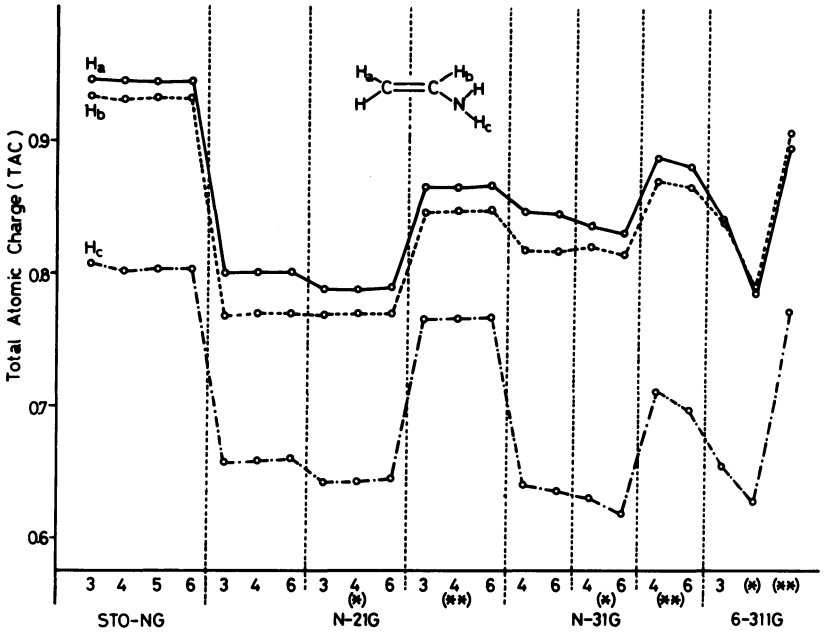

Fig. 1. Dependence of total atomic charge on basis set in hydrogen atoms of vinylamine.

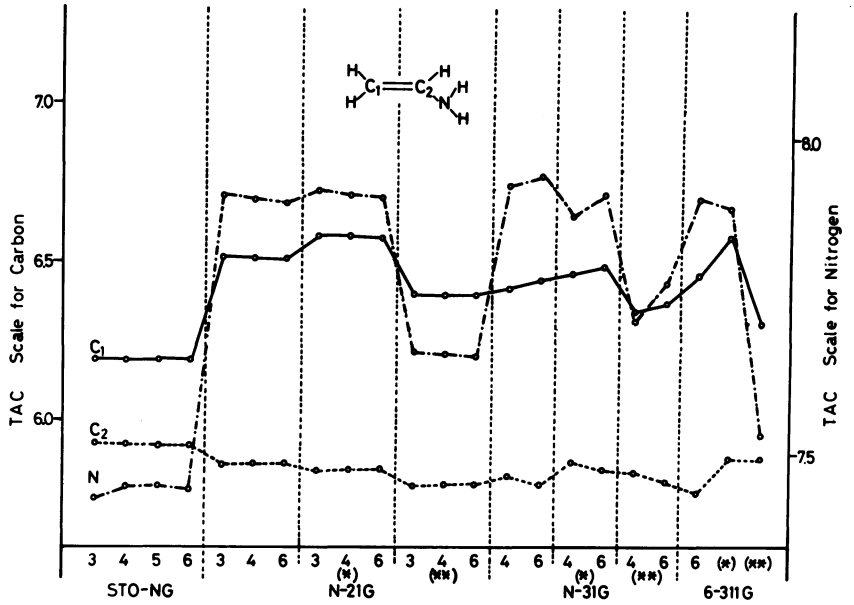

Fig. 2. Dependence of total atomic charge on basis set in carbon and nitrogen atoms of vinylamine.

be generally said that similar methods give similar electronic structures. The STO-*NG* series give almost the same total atomic charges, independently of the contraction number *N*.

In the split-valence methods, the number of the primitive gaussian functions for an inner-shell 1*s* AO has almost nothing to do with the total atomic charge of the hydrogen. "Single-star" methods generally give a total atomic charge for hydrogen that is too low, while "double-star" methods compensate such a loss of electrons on the hydrogen atoms. A good example is the difference of the results between the 6-311G* and 6-311G** methods, where the difference is most prominent.

Figure 2 shows the total atomic charges of C_1, C_2, and N for optimized structures of vinylamine. If we consider that the 6-311G** method gives the best-balanced wave function at present, any other split-

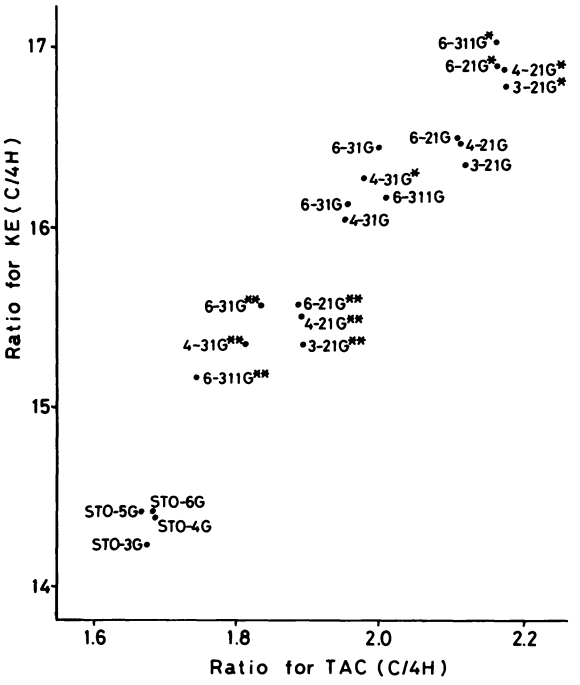

Fig. 3. Ratio of kinetic energies (C/4H) vs. ratio of total atomic charge (C/4H).

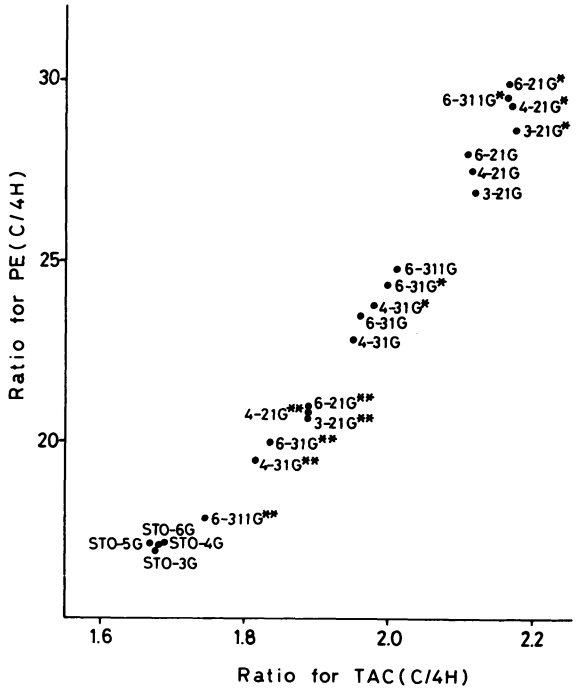

Fig. 4. Ratio of potential energies (C/4H) vs. ratio of total atomic charges (C/4H).

valence method without a "double-star" seems to produce off-balanced wave functions. Interestingly, the STO-*N*G method gives a very similar result to that of 6-311G**.

The above results suggest that the addition of polarization functions allows much freedom for electrons, resulting in a large electronic population on atoms with polarization functions. This indicates that polarization functions must be added to both the atoms of the first row and hydrogen atoms in order to obtain a well-balanced wave function. It should be stressed here that the total atomic charge obtained by STO-*N*G is very close to that obtained by 6-311G**, which is considered to be very close to the Hartree-Fock limit.

2. The Relationship between Total Atomic Charge and Partitioned Energies. Since the total energy is dependent on the basis set, a direct comparison of the partitioned energy with that of a different basis set is meaningless. As a matter of fact, we usually do not need the absolute energy, but only the relative energy or relative distribution of energy in a molecule with respect to the change of the molecular geometry. Accordingly, we examined the ratios of the total atomic charge and partitioned energies.

The horizontal axis of Figs. 3 or 4 represents the ratios of the total atomic charge of the carbon atom to the sum of those for four hydrogen atoms in methane. The ratios range from 1.685 to 2.174. It can again be said that a similar method gives similar results: a "single-star" series give results that are very different from that of 6-311G** because of an *unfair* addition of electronic freedom (only for carbon). The addition of a polarization function on hydrogen improves such a deformed wave function.

The vertical axes of Figs. 3 and 4 are for the ratios of the kinetic and potential energies, respectively. There is a close linear relationship between the ratios of the total atomic charge and that of the potential energy or the kinetic energy.

3. Basis Set Dependency of Partitioned Potential and Kinetic Energies. It was very fortunate to be able to obtain the result that STO-*N*G can reproduce a well-balanced wave function since STO-*N*G is a feasible method that can be applied to systems that are as large as those in which organic chemists show interest. We extended a similar examination in order to determine whether or not STO-*N*G gives a well-balanced wave function, regardless of the position of the bisection. Since it became clear that a similar method gives similar results, we only examined STO-3G, 3-21G, 4-31G, 6-31G*, and 6-311G**, and assumed that 6-311G** gives the best-balanced wave function out of all the methods.

Table 1 shows the results for carbon monoxide, formaldehyde, and ketene, where the ratios of the kinetic energies (KE(X)/KE(Y)) and the potential energies (PE(X)/PE(Y)) are also listed. We chose the oxygen-containing compounds as models, since such compounds are most liable to show a large depen-

Table 1. Ratios of the Potential and Kinetic Energies in Divided Systems

X†Y		STO-3G	3-21G	4-31G	6-31G*	6-311G**
C†O	KE[a]	0.5116	0.5003	0.5021	0.5060	0.5057
	PE[b]	0.4993	0.4835	0.4874	0.4958	0.4982
H_2†C=O	KE	0.0117	0.0108	0.0108	0.0107	0.0116
	PE	0.0100	0.0058	0.0066	0.0069	0.0102
H_2C†O	KE	0.5224	0.5185	0.5205	0.5214	0.5219
	PE	0.5166	0.5027	0.5049	0.5087	0.5124
H_2†C=C=O	KE	0.0082	0.0069	0.0071	0.0070	0.0077
	PE	0.0064	0.0012	0.0026	0.0023	0.0047
H_2C†C=O	KE	0.3483	0.3460	0.3462	0.3470	0.3461
	PE	0.3489	0.3450	0.3460	0.3501	0.3470
H_2C=C†O	KE	1.0338	1.0236	1.0265	1.0277	1.0284
	PE	1.0207	0.9918	0.9979	1.0019	1.0099

a) Ratio KE(X)/KE(Y). b) Ratio PE(X)/PE(Y).

Table 2. Ratios of the Potential and Kinetic Energies of Two Functional Groups

X†Y		STO-3G	3-21G	4-31G	6-31G*	6-311G**
CH_3†$CH=CH_2$	KE[a]	0.5105	0.5114	0.5115	0.5117	0.5109
	PE[b]	0.5111	0.5111	0.5116	0.5121	0.5074
CH_3†C≡CH	KE	0.5151	0.5178	0.5178	0.5192	0.5175
	PE	0.5162	0.5130	0.5114	0.5179	0.5122
CH_3†CN	KE	0.4267	0.4278	0.4282	0.4288	0.4276
	PE	0.4267	0.4246	0.4243	0.4271	0.4220
H†CN	KE	0.0057	0.0052	0.0053	0.0053	0.0063
	PE	0.0045	0.0006	0.0014	0.0016	0.0042
CH_3†NC	KE	0.4242	0.4254	0.4261	0.4276	0.4273
	PE	0.4212	0.4142	0.4135	0.4192	0.4196
H†NH_2	KE	0.0117	0.0101	0.0099	0.0097	0.0114
	PE	0.0082	0.0041	0.0034	0.0028	0.0079
CH_3†NH_2	KE	0.7121	0.7092	0.7096	0.7113	0.7106
	PE	0.7087	0.6985	0.6971	0.7014	0.7031
H†OH	KE	0.0089	0.0072	0.0072	0.0067	0.0089
	PE	0.0057	0.0013	0.0006	0.0004	0.0052
CH_3†OH	KE	0.5252	0.5223	0.5228	0.5237	0.5238
	PE	0.5225	0.5119	0.5119	0.5127	0.5159

a) Ratio KE(X)/KE(Y). b) Ratio PE(X)/PE(Y).

dence on the basis set. Comparing those ratios, we could again conclude that the STO-3G method reproduces the results of 6-311G** very well.

Next, we examined the partitioned energies bisected by a functional group. Table 2 shows the ratios of the kinetic and potential energies of the two functional groups. The ratios are similar, independent of the method unless one of the functional groups is hydrogen. As shown in the former section, the total atomic charge of hydrogen is most susceptible to a change in the basis set: the split-valence methods and the split-valence methods with polarization functions only for first row atoms ("single-star" series) are liable to cause a low electron distribution on hydrogen resulting in a large difference in the energy ratios from the STO-*N*G or 6-311G** methods. It was again ascertained that the STO-3G method gives the closest ratios to those by the 6-311G** method in all kinds of functional partitionings.

Thus, we have obtained very instructive results regarding the choice of the basis set in a MO calculation. It is well known that the larger is the adopted basis set, the better the molecular wave function. It can be obtained simply because of the lower total energy. However, this does not apply to the electron distribution in the system. The addition of polarization functions to only first row atoms gives the most deformed wave functions. Instead, the STO-*N*G method, which was considered to give a poor wave function, was found to produce a well-balanced wave function. Recently, contour maps have been widely used in an explanation of the molecular properties and the reaction processes in cases where the shape of molecular wave function is the decisive factor.

Finally, Fig. 5 shows a comparison of the computation time (cpu time) of 22 different basis sets for some small molecules. The computer used was a HITAC

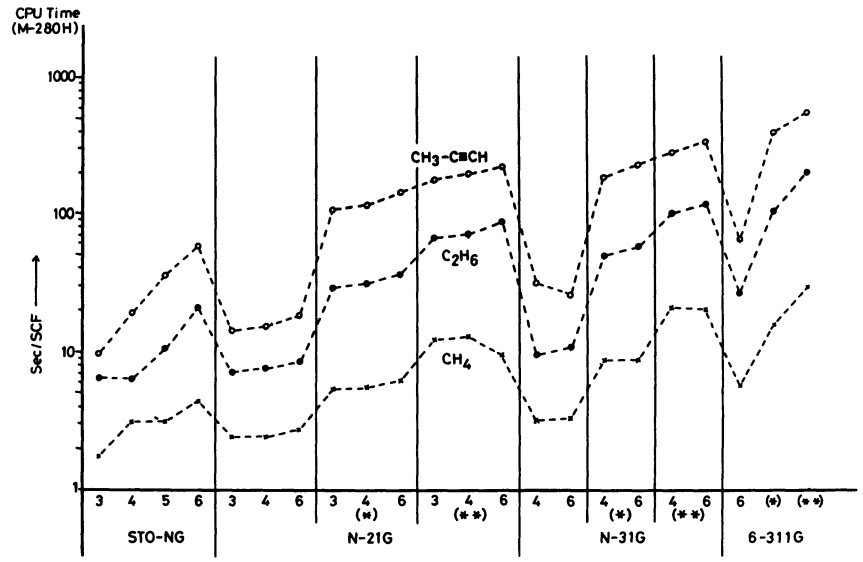

Fig. 5. Computer cpu time (HITAC M-280H) required for a single SCF calculation by 22 different MO methods.

M-280H system at the Computer Center of the University of Tokyo.

References

1) e.g. a) J. B. Moffat and H. E. Popkie, *Int. J. Quantum Chem.*, **2**, 565 (1968); b) F. Driessler and W. Kutzelnigg, *Theor. Chim. Acta (Berl.)*, **43**, 1 (1976); *ibid.*, **43**, 307 (1977); c) H. Kollmar, *ibid.*, **50**, 235 (1978).

2) H. Ichikawa, *J. Am. Chem. Soc.*, **105**, 7457 (1983); *ibid.*, **106**, 6249 (1984).

3) H. Ichikawa and Y. Ebisawa, *J. Am. Chem. Soc.*, **107**, 1161 (1985).

4) We have used GAUSSIAN 80H program (a Hitac version of GAUSSIAN 80 (QCPE No. 437, Indiana University, IN)) at the Computer Center of the University of Tokyo.

5) a) STO-3G basis set: J. B. Collins, P. v. R. Schleyer, J. S. Binkley, and J. A. Pople, *J. Chem. Phys.*, **64**, 5142 (1976); b) STO-4G basis set: W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, *ibid.*, **51**, 2657 (1969); c) 3-21G and 6-21G basis sets: J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980); M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, *ibid.*, **104**, 2797 (1982); d) 4-21G basis set: P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, *ibid.*, **101**, 2250 (1979); e) 3-21G* basis set: W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople, and J. S. Binkley, *J. Am. Chem. Soc.*, **104**, 5039 (1982); f) 4-31G basis set: R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem Phys.*, **54**, 724 (1971); W. J. Hehre and J. A. Pople, *ibid.*, **56**, 4233 (1972); J. D. Dill and J. A. Pople, *ibid.*, **62**, 2921 (1975); g) 6-31G basis set: W. J. Hehre, R. Ditchfield, and J. A. Pople, *ibid.*, **56**, 2257 (1972); J. S. Binkley and J. A. Pople, *ibid.*, **66**, 879 (1977); J. D. Dill and J. A. Pople, *ibid.*, **62**, 2921 (1975); h) 6-31G* and 6-31G** basis sets: P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **28**, 213 (1973); J. D. Dill and J. A. Pople, *J. Chem. Phys.*, **62**, 2921 (1975); i) 6-311G and 6-311G** basis sets: R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *ibid.*, **72**, 650 (1980); The GAUSSIAN-80 program includes other basis sets, whose exponents and contraction coefficients are listed in "Handbook of Gaussian Basis Sets," ed by R. Poirier, R. Kari, and I. G. Csizmadia, Elsevier Pub. Co., Amsterdam (1985).

6) The method of energy partitioning was proposed and rigorously evaluated by Reudenberg (K. Reudenberg, *Rev. Mod. Phys.*, **34**, 326 (1962)). There are some problems in partitioning the total energy by an HF SCF method into mono- and bicentric terms since a Fock operator contains the potentials from three- and four-center interactions in an averaged form. However, since there are no other means to see inside of a large molecular system, this method has been frequently applied to the analysis of chemical bonds as shown in Ref. 1.

7) G. G. Hall, *Proc. R. Soc. London, Ser. A*, **205**, 541 (1951); C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951).

8) J. A. Pople and R. K. Nesbet, *J. Chem. Soc.*, **1964**, 571.

9) R. S. Mulliken, *J. Chem. Soc.*, **23**, 1822 (1955); *ibid.*, **23**, 1841 (1955); *ibid.*, **23**, 2338 (1955); *ibid.*, **23**, 2343 (1955).