Ab Initio Calculations on Large Molecules Using Molecular Fragments. Evaluation and Extension of Initial Procedures[1]

Ralph E. Christoffersen,*[2a,b] Dale Spangler,[2b] George G. Hall,[2b] and Gerald M. Maggiora[2c]

Contribution from the Chemistry and the Biochemistry Departments, University of Kansas, Lawrence, Kansas 60644. *Received July 31, 1973*

Abstract: Characteristics of the molecular fragment procedure as initially formulated are presented by means of analysis of prototype examples and suitable theoretical analysis. The relationship of nonatomic spherical Gaussian orbital basis sets to atomic Gaussian orbital basis sets is also given. Examples are presented that illustrate how the molecular fragment basis set can be improved in a straightforward, yet computationally economical, fashion using the analysis, and recommendations regarding the use of the procedure in its initial form are given.

In a series of studies,[3-16] the initial formulation and characterization of a procedure (the molecular fragment method) that is designed to allow application of *ab initio* quantum mechanical techniques to large molecular systems have been given, along with application to a variety of problems of chemical and biological interest. The computational characteristics that have emerged indicate clearly that large molecular systems can indeed be examined in a computationally convenient manner using this technique and that results of a qualitative, and frequently quantitative, nature are obtainable for a variety of properties. It is therefore particularly important and appropriate at this time to analyze these results in a uniform manner, so that general characteristics of the procedure in its initial form, both good and bad, can be identified.

The following sections provide such an analysis, by introducing a general theoretical framework in which all of the results can be considered. Also, specific examples will be used to show how improvements in the initial formulation can be carried out easily. Furthermore, the relationship of floating spherical Gaussian orbitals (FSGO) to atomic Gaussian orbital basis sets will be developed and the relative advantages and disadvantages of each discussed.

The Molecular Fragment Model. Initial Formulation

Since the conceptual and computational features of the molecular fragment procedure have been described previously, only a summary of points salient to the current investigation will be included here.

The motivation for the molecular fragment procedure is based upon the observation that, if the electronic and geometric structure of large molecules of arbitrary symmetry are to be examined in an *ab initio* framework, small basis sets are essential. One of the basic hypotheses of the molecular fragment method is that such basis sets can be obtained, if adequate account is taken of the anticipated molecular environment. Thus, such basis orbitals should be highly flexible, and their nonlinear parameters should be determined within an environment which closely resembles the anticipated molecular environment. Considerations such as these led to the choice of FSGO as basis orbitals,[17] which are defined as

$$G_i(r) = (2\pi/\rho_i^2)^{3/4} \exp\{-(\mathbf{r} - \mathbf{R}_i)^2/\rho_i^2\} \qquad (1)$$

where ρ_i is referred to as the orbital radius, and $\mathbf{R}_i$ defines the location[18] of the origin of the FSGO, relative to some arbitrary origin. The position and size of these FSGO are determined *via* energy minimization calculations on molecular fragments that are chosen to mimic the various anticipated bonding environments. After completion of the molecular fragment studies, the various fragments of interest and their associated FSGO are combined appropriately to form the large molecule of interest, and an SCF calculation is carried out.

In the initial formulation to be discussed here that has been used in most studies to date, each of the molecular fragments of interest is described by a single Slater determinant, whose orbitals χ are taken either as a single FSGO or, at most, a linear combination of two FSGO (in the case of π-type orbitals). A summary of the various molecular fragments to be discussed here is

(1) This work was supported in part by grants from the National Science Foundation and the Upjohn Co., Kalamazoo, Mich.

(2) (a) Alfred P. Sloan Research Fellow, 1971–1973. (b) Chemistry Department. (c) Biochemistry Department. Supported in part by a USPHS Health Sciences Advancement Award (RR-06147) to the University of Kansas, and the University of Kansas General Research Fund.

(3) (a) R. E. Christoffersen and G. M. Maggiora, *Chem. Phys. Lett.*, **3**, 419 (1969). (b) R. E. Christoffersen, D. W. Genson, and G. M. Maggiora, *J. Chem. Phys.*, **54**, 239 (1971). (c) R. E. Christoffersen, *J. Amer. Chem. Soc.*, **93**, 4104 (1971).

(4) G. M. Maggiora, R. E. Christoffersen, D. W. Genson, and B. V. Cheney, *Theor. Chim. Acta*, **22**, 337 (1971).

(5) R. E. Christoffersen, L. L. Shipman, and G. M. Maggiora, *Int. J. Quantum Chem.*, **S5**, 143 (1971).

(6) R. E. Christoffersen, *Advan. Quantum. Chem.*, **6**, 333 (1972).

(7) B. V. Cheney and R. E. Christoffersen, *J. Chem. Phys.*, **56**, 3503 (1972).

(8) D. W. Genson and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **94**, 6904 (1972).

(9) L. L. Shipman and R. E. Christoffersen, *Chem. Phys. Lett.*, **15**, 469 (1972).

(10) R. E. Christoffersen, "Energy, Structure, and Reactivity," D. W. Smith and W. B. McRae, Ed., Wiley, New York, N. Y., 1973, p 357.

(11) L. L. Shipman and R. E. Christoffersen, *Proc. Nat. Acad. Sci. U. S.*, **69**, 3301 (1972).

(12) D. W. Genson and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **95**, 362 (1973).

(13) L. L. Shipman and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **95**, 1408 (1973).

(14) L. J. Wiemann and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **95**, 2074 (1973).

(15) L. L. Shipman and R. E. Christoffersen, *J. Amer. Chem. Soc.*, **95**, 4733 (1973).

(16) R. E. Christoffersen, *Int. J. Quantum Chem.*, in press.

(17) These basis orbitals were studied extensively on small molecular systems earlier by Frost and coworkers. See, for example, S. Y. Chu and A. A. Frost, *J. Chem. Phys.*, **54**, 760, 764 (1971), and earlier references therein.

(18) Unless otherwise specified, hartree atomic units are used. See H. Shull and G. G. Hall, *Nature (London)*, **184**, 1559 (1959).

Table I. Molecular Fragments to be Discussed

Molecular fragment	Common use	"Nonsplit" inner shell description: No. of FSGO	"Nonsplit" inner shell description: No. of independent nonlinear parameters[a]
CH_4 (tetrahedral)	Saturated carbon atoms	5	3
$\cdot CH_3$ (planar)	Unsaturated carbon atoms	6[b]	5 (4)
NH_3 (tetrahedral)	Amines	5	6
NH_3 (planar)	Amides, pyrrole	6[b]	5 (4)
$\cdot NH_2$ (planar)	Pyridine	6[b]	8 (7)
NH_4^+ (tetrahedral)	Ammonium ions	5	3
H_2O ("sp^3")	Ethers, alcohols	5	7 (6)
H_2O ("sp^2")	Acids, esters, furan	6[b]	8 (7)
$\cdot OH$ ("sp")	Carbonyl groups	7[b]	10 (8)

[a] Only variation of FSGO parameters is considered here, although variation of nuclear coordinates is also possible. When two entries are given, the one in parentheses represents the number of parameters actually varied in the fragment calculation. [b] These fragments employ the π-orbital description of eq 2.

FSGO is employed, *i.e.*

$$\chi_\pi = [2(1 - \Delta_{ud})]^{-1/2}(G_u - G_d) \quad (2)$$

where G_u and G_d are FSGO that are placed symmetrically above and below the central atom, on a line perpendicular to the plane of atoms, and Δ_{ud} is the overlap integral between G_u and G_d.

The total fragment wave function is then written as a single Slater determinant of orbitals for closed shell systems and for systems containing one unpaired electron. The size and position of the individual FSGO are then determined by minimizing the expectation value of the total energy of the fragment under consideration. Hence, one obtains a set of FSGO which are tailored to a particular bonding environment.

These functions are now used as basis functions for an SCF–MO calculation on a large molecule, where each molecular orbital (φ_i) is given by

$$\varphi_i = \sum_{A=1}^{P} \sum_{k=1}^{N_A} c_{ki}^A \chi_k^A \quad (3)$$

where the χ_k^A are the previously determined fragment orbitals, and the c_{ki}^A are the molecular orbital coefficients that arise from the solution of the well-known SCF equations.[19,20] The sums are taken over all fragments (P) and orbitals within a fragment (N_A). These calculations are carried out using double precision arithmetic (15 significant figures), and convergence of the SCF procedure is typically monitored using the charge

Table II. Optimized Fragment Data[a]

Fragment	Orbital radius (ρ)	Distance from the heavy atom	Molecular parameters
CH_4 (tetrahedral, sp^3)[d]	ρ_{CH} = 1.67251562	1.23379402	R_{CH} = 2.05982176
	ρ_C = 0.32784375	0.0	
$\cdot CH_3$ (planar, sp^2)[d]	ρ_{CH} = 1.51399487	1.13093139	R_{CH} = 1.78562477
	ρ_C = 0.32682735	0.0	
	ρ_π = 1.80394801	±0.1	
NH_3 (tetrahedral, sp^3)[e]	ρ_{NH} = 1.52791683	0.87735349	R_{NH} = 1.91242167
	ρ_N = 0.27732014	0.00099090[b]	
	ρ_{LP} = 1.58328000	0.25523498	
NH_3 (planar, sp^2)[e,f]	ρ_{NH} = 1.47683593	0.94031372	R_{NH} = 1.87084729
	ρ_N = 0.27814453	0.0	
	ρ_{LP} = 1.51198608	±0.1	
$\cdot NH_2$ (planar, sp^2)[e]	ρ_{NH} = 1.43795016	0.89803124	R_{NH} = 1.75153951
	ρ_N = 0.27698950	0.00088379[b]	
	ρ_{LP} = 1.51400386	0.30407714	
	ρ_π = 1.35873044	±0.1	
NH_4^+ (tetrahedral, sp^3)[g]	ρ_{NH} = 1.50046875	0.80547793	R_{NH} = 1.95021656
	ρ_N = 0.27770068	0.0	
H_2O ("sp^3")[h]	ρ_{OH} = 1.35682617	0.74365356	R_{OH} = 1.81415494
	ρ_O = 0.24053100	0.00077105	∠HOH = 104.52°
	ρ_{LP} = 1.30568359	0.43956044	∠(lone pair FSGO) = 138.579562°
H_2O ("sp^2")[h]	ρ_{OH} = 1.37684374	0.79678221	R_{OH} = 1.81415494
	ρ_O = 0.24089502	0.00083398[b]	∠HOH = 120.0°
	ρ_{LP} = 1.36888573	0.23835937	
	ρ_π = 1.13643749	±0.1	
$\cdot OH$ ("sp")[i]	ρ_{OH} = 1.23671871	0.76467773	R_{OH} = 1.54774058
	ρ_O = 0.24028227	0.00057129[c]	
	$\rho_{LP\text{-}\sigma}$ = 1.28753780	0.21614258	
	$\rho_{LP\text{-}\pi}$ = 1.19741696	±0.1	
	ρ_π = 1.12242182	±0.1	

[a] All distances and energies are given in hartree atomic units, unless otherwise specified. See ref 18. [b] This position is along a line that bisects the HXH angle (where X = N or O), displaced toward the H nuclei. [c] This position is along the OH bond, displaced in the direction of the H nucleus. [d] See ref 3b. [e] For a description of the characteristics of this fragment, see ref 8. [f] For a later, and improved, description of this fragment, see ref 13. [g] See ref 12. [h] See ref 7. [i] See ref 14.

given in Table I, and the optimized FSGO parameters associated with these molecular fragments are given in Table II. For fragments in which a π-type orbital is desired, a normalized, fixed linear combination of two

(19) G. G. Hall, *Proc. Roy. Soc., Ser. A*, **205**, 541 (1951).
(20) C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951).

and bond order matrix (**P**), defined as

$$P_{rs} = 2\sum_{i}^{occ} c_{ir}c_{is} \quad (4)$$

where the sum is taken over all occupied molecular orbitals. Final convergence is assumed in general when

$$|P_{rs}^{(i+1)} - P_{rs}^{(i)}| \leq 0.00005 \quad (5)$$

for all r and s, where $\mathbf{P}^{(i)}$ is the charge and bond order matrix at the end of the ith iteration. This convergence criteria corresponds generally to a root-mean-square error in **P** of 10^{-6} and a final total energy that is converged to ten figures.

Using this procedure, over 100 molecular systems have been investigated. The following sections summarize several characteristics of interest that have been observed regarding these systems, analyze the reasons for these results, and provide examples of how improvements can be made, where appropriate.

Results on Molecular Systems

Considering geometric characterizations first, Table III summarizes the results for several geometric properties of interest on protopype molecules. The adequacy and reasons for these results will be discussed in later sections.

Table III. Geometric Characterizations

A. Distance and Angle Predictions

Prototype molecule	Parameter predicted	Calcd value	Obsd value[a]	% error
C_2H_6	C—C distance	2.640	2.903	9.1
C_2H_4	C=C distance	2.56	2.47	2.8
C_2H_2	C≡C distance	2.3734	2.2739	4.2
CH_3NH_2	C—N distance	2.65	2.785	4.8
CH_2NH	C=N distance	2.422	2.46[b]	1.5
CH_3OH	C—O distance	2.5524	2.697	5.3
H_2CO	C=O distance	2.3735[c]	2.2843	3.8
N_2H_4	N—N distance	2.56	2.738	6.5
N_2H_2	N=N distance	2.29[d]	2.35	2.6
H_2O_2	O—O distance	2.4758	2.7873	11.2
C_3H_8	∠CCC	110.6°	112.4°	1.6
$(CH_3)_2N$	∠CNC	102°	112.2°	9.1
$(CH_3)_2O$	∠COC	94.7°	111.5°	15.1

B. Barrier to Rotation Predictions

Prototype molecule	Calcd barrier, kcal/mol	Obsd barrier, kcal/mol[a]
C_2H_6	5.38[e]	2.98
C_3H_8	5.63[f]	3.20
CH_3NH_2	5.00[g]	1.97
CH_3OH	3.1[h]	1.07
N_2H_4	*i*	3.58[k]
H_2O_2 (cis)	11.3	7.0
H_2O_2 (trans)	1.3	1.1
CH_3HCO	1.5[j]	1.16
$HCONH_2$	19.72	19.7–19.2

[a] See ref 3b, 7, 8, and 13 for references to original experimental measurements. [b] Taken from *N*-methylmethylenimine. [c] Calculated using an early "sp²" description of the ·OH fragment. See ref 7 for details. [d] Calculated for diimide in the trans conformation. [e] Calculated using tetrahedral angles and experimental distances. [f] Corresponds to *E*(staggered–staggered) − *E*(staggered–eclipsed). [g] Calculated using an axis of rotation that is collinear with the C–N bond. [h] Calculated by taking *E*(staggered) − *E*(eclipsed). [i] Wrong conformer (180° torsion angle) predicted to be most stable. [j] Barrier to methyl group rotation. [k] Value of the cis barrier in methylhydrazine. See R. P. Lattimer and M. D. Harmony, *J. Amer. Chem. Soc.*, **94**, 351 (1972). Hydrazine itself is currently being reinvestigated: M. D. Harmony, private communication.

In addition to geometric characterizations, electronic structure comparisons have also been made for prototype molecules. While many aspects of the electronic structure of molecules are of interest (*e.g.*, dipole moments, quadrupole moment, net atomic charges, etc.), it will be useful to focus primarily on the molecular orbital structure in these discussions, since a large amount of data for comparison are available on this aspect of electronic structure. Also, it will frequently be seen to be a rather sensitive measure of the ability of various procedures to describe electronic structure.

The general observation that can be made from the studies completed to date is that the ordering of "valence" molecular orbitals (*i.e.*, non-inner-shell molecular orbitals) is essentially identical with that obtained in large basis set calculations for most molecules. Several examples of this type of observation are given in Table IV for naphthalene, furan, and formamide, where comparisons are made to the more extensive basis set calculations of Buenker and Peyerimhoff,[21] Siegbahn,[22] and Christensen, *et al.*[23] As is seen from this table, the balance of the molecular fragment basis set, as exhibited by the relative ordering of molecular orbitals, is essentially identical with that of the considerably larger basis set in each case.

Another manner of observing and quantifying the ability of the molecular fragment procedure to predict molecular orbital structure can be seen by constructing plots of valence molecular orbital energies obtained from extensive basis set calculations *vs.* those obtained using the molecular fragment procedure. When this is done, a remarkable linear relationship has been observed, as shown by the data for 21 prototype systems in Table V. The data in that table correspond to a least-squares fit of the points to a straight line having the form

$$\epsilon_i^{ref} = a\epsilon_i^{MF} + b \quad (6)$$

where the ϵ_i^{MF} are valence molecular orbital energies obtained using the molecular fragment procedure, and ϵ_i^{ref} are similar quantities taken from more extensive basis set calculations.[21-39] These linear relationships

(21) R. J. Buenker and S. D. Peyerimhoff, *Chem. Phys. Lett.*, **3**, 37 (1969).
(22) P. Siegbahn, *Chem. Phys. Lett.*, **8**, 245 (1971).
(23) D. H. Christensen, R. N. Kortzeborn, B. Bak, and J. J. Led, *J. Chem. Phys.*, **53**, 3912 (1970).
(24) J. D. Petke, J. L. Whitten, and J. A. Ryan, *J. Chem. Phys.*, **48**, 953 (1968).
(25) E. Clementi and D. R. Davis, *J. Chem. Phys.*, **45**, 2593 (1966).
(26) J. W. Moskowitz and M. C. Harrison, *J. Chem. Phys.*, **42**, 1726 (1965).
(27) R. J. Buenker, J. L. Whitten, and J. D. Petke, *J. Chem. Phys.*, **49**, 2261 (1968).
(28) S. D. Peyerimhoff, private communication.
(29) W. H. Fink and L. C. Allen, *J. Chem. Phys.*, **46**, 2261 (1967).
(30) A. Veillard, *Theor. Chim. Acta*, **18**, 21 (1970).
(31) D. B. Neumann and J. W. Moskowitz, *J. Chem. Phys.*, **50**, 2216 (1969).
(32) H. Basch, M. B. Robin, and N. A. Kuebler, *J. Chem. Phys.*, **49**, 5007 (1968).
(33) P. Siegbahn, *Chem. Phys. Lett.*, **8**, 245 (1971).
(34) J. M. Lehn, B. Munsch, and P. H. Millie, *Theor. Chim. Acta*, **16**, 351 (1970).
(35) D. P. Wong, W H. Fink, and L. C. Allen, *J. Chem. Phys.*, **52**, 6291 (1970).
(36) E. Clementi, H. Clementi, and D. R. Davis, *J. Chem. Phys.*, **46**, 4725 (1967).
(37) D. H. Christensen, R. N. Kartzeborn, B. Bak, and J. J. Led, *J. Chem. Phys.*, **53**, 3912 (1970).

Table IV. Molecular Orbital Structure for Prototype Molecules

Naphthalene			Furan			Formamide		
	Orbital energies			Orbital energies			Orbital energies	
MO symmetry	Molecular fragment approach[a]	Buenker and Peyerimhoff[b]	MO symmetry	Molecular fragment approach[c]	Siegbahn[d]	MO symmetry[e]	Molecular fragment approach[f]	Christensen, *et al.*[g]
$1a_u(\pi)$	−0.1545	−0.3418	$1a_2$	−0.1596	−0.3330	$\epsilon_{12}(\pi_n)$	−0.069	−0.425
$2b_{1u}(\pi)$	−0.2002	−0.3750	$2b_2$	−0.2284	−0.4054	$\epsilon_{11}(\sigma_n)$	−0.085	−0.441
$1b_{3g}(\pi)$	−0.2658	−0.4355	$9a_1$	−0.3285	−0.5417	ϵ_{10}	−0.287	−0.578
$1b_{2g}(\pi)$	−0.3304	−0.4966	$8a_1$	−0.4043	−0.5697	ϵ_9	−0.298	−0.617
$9a_g$	−0.3755 ($6b_{1g}$)	−0.5221	$6b_1$	−0.4392	−0.5888	ϵ_8	−0.379	−0.675
$6b_{1g}$	−0.3781 ($9a_g$)	−0.5300	$5b_1$	−0.4470	−0.6068	ϵ_7	−0.466	−0.751
$7b_{3u}$	−0.4234	−0.5673	$1b_2$	−0.4987	−0.6351	ϵ_6	−0.615	−0.847
$1b_{1u}(\pi)$	−0.4261	−0.5758	$7a_1$	−0.5807	−0.7446	ϵ_5	−1.009	−1.216
$7b_{2u}$	−0.4522	−0.6094	$6a_1$	−0.6865	−0.7847	ϵ_4	−1.224	−1.140
$6b_{3u}$	−0.4686	−0.6305	$4b_1$	−0.6872	−0.8115			
$5b_{1g}$	−0.4725	−0.6347	$3b_1$	−0.9176	−1.0199	ϵ_3	−9.104	−11.369
$8a_g$	−0.5009	−0.6616	$5a_1$	−0.9923	−1.1033	ϵ_2	−12.815	−15.596
$6b_{2u}$	−0.5255	−0.6689	$4a_1$	−1.4877	−1.4734	ϵ_1	−17.071	−20.537
$7a_g$	−0.5561	−0.7178						
$5b_{3u}$	−0.5803	−0.7447	$3a_1$	−9.2367	−11.2654 ($2b_1$)			
$4b_{1g}$	−0.6255	−0.7484	$2b_1$	−9.3307	−11.2571 ($3a_1$)			
$6a_g$	−0.7453	−0.8512	$2a_1$	−9.3910	−11.3119 ($1b_1$)			
$5b_{2u}$	−0.7687	−0.8719	$1b_1$	−9.4112	−11.3119 ($2a_1$)			
$4b_{3u}$	−0.7997	−0.8866	$1a_1$	−17.4714	−20.6249			
$3b_{1g}$	−0.9246	−1.0249						
$5a_g$	−0.9654	−1.0466						
$4b_{2u}$	−1.0040	−1.0898						
$3b_{3u}$	−1.0538	−1.1455						
$4a_g$	−1.1188	−1.2133						
$3b_{2u}$	−9.2666 ($3a_g$)	−11.3598						
$2b_{1g}$	−9.3083 ($2b_{3u}$)	−11.3599						
$3a_g$	−9.3396 ($3b_{2u}$)	−11.3599						
$2b_{3u}$	−9.3607 ($2a_g$)	−11.3599						
$2b_{2u}$	−9.3817 ($2b_{1g}$)	−11.3620						
$1b_{1g}$	−9.4172 ($2b_{2u}$)	−11.3620						
$1b_{3u}$	−9.4232 ($1b_{3u}$)	−11.3628						
$2a_g$	−9.4295 ($1a_g$)	−11.3628						
$1b_{2u}$	−9.4497 ($1b_{1g}$)	−11.3916						
$1a_g$	−9.4516 ($1b_{2u}$)	−11.3926						

[a] See ref 3c. [b] See ref 21. [c] See ref 7. [d] See ref 22. [e] Although the symmetry of the molecule is too low to allow meaningful comparisons to be made using the MO irreducible representation labels, a detailed examination of the individual molecular orbitals indicates a one to one correspondence between the results of the molecular fragment procedure and those of Christensen, *et al.* [f] See ref 13. [g] See ref 23.

reveal that the molecular orbital energies from the molecular fragment procedure have larger spacings than those obtained in more extensive basis set studies, which result in calculated vertical ionization potentials that are uniformly too small. It is also of interest to observe that some of the various classes of molecules have slopes that are remarkably similar, *e.g.*, hydrocarbons. Thus, there is a quantitative, as well as qualitative, aspect of basis set balance that is identifiable in the molecular fragment basis sets.

However, not all molecular orbital structure predictions are completely satisfactory, as illustrated by the studies on pyridine and pyrazine that are summarized in Table VI. A similar difficulty is found in the case of pyrrole.[8] These more subtle difficulties, involving only one or a few molecular orbitals, do not affect the orbital energy linearity relationships indicated in Table V and will be discussed in greater detail in later sections.

In addition to the geometric and energetic properties just discussed, investigation of other aspects of electronic structure has taken place through a study of several molecular properties. Among the properties that appear to be well predicted in general are Hellmann–Feynman electric fields[40] at nuclei, an illustrative summary of which appears in Table VII. These fields (or the closely related forces on nuclei), which will be zero for the exact wave function at the equilibrium configuration, are seen to be quite small in general and reasonably independent of the kind of molecule examined.

In Table VIII, calculated dipole moments are compared with available experimentally observed values. Although these values depend rather strongly upon the assumed geometry, we see that the expected trends are well represented in general and that quantitative accuracy is also obtained in several cases (*e.g.*, amides).

Finally, in studies of various conformations of glycine polypeptides,[11,15] it was found that several aspects of intramolecular hydrogen bonding could be extracted directly from the calculations. In particular, it was found that the magnitude of intramolecular hydrogen bonding effects in the tetra- and pentapeptide of glycine was very satisfactorily predicted (6.1 kcal/mol = hydrogen bond stabilization energy). In addition, analysis of the changes in π-orbital charge distribution trends as a function of conformer allowed identification of the sites of hydrogen bonding. On the other hand, changes

(38) The molecular fragment studies on the glycine zwitterion are reported in L. L. Shipman and R. E. Christoffersen, *Theor. Chim. Acta*, **31**, 75 (1973).

(39) J. A. Ryan and J. L. Whitten, *J. Amer. Chem. Soc.*, **94**, 2396 (1972).

(40) H. H. Hellmann, "Einführung in die Quantenchemie," Franz Deuticke, Leipzig 1937; R. P. Feynman, *Phys. Rev.*, **56**, 340 (1939).

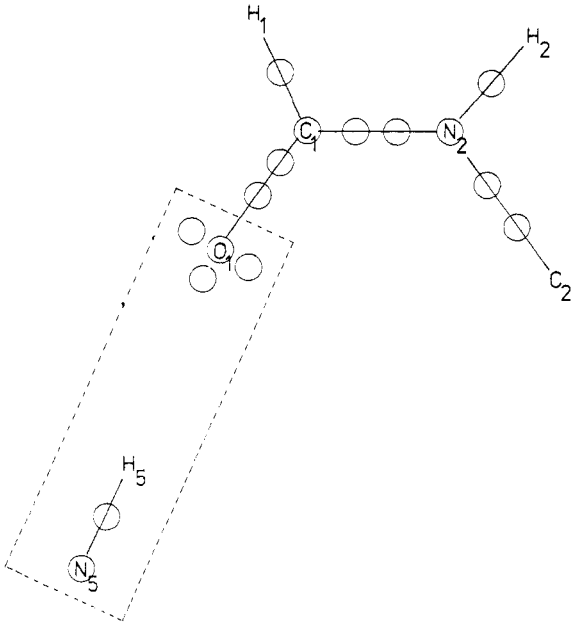

Figure 1. Approximate FSGO positions (indicated by circles) in the hydrogen bonding region of the α-helix form of glycine polypeptides of four subunits and longer. The positive z axis points up from the paper. All atoms depicted are in the same plane, except for N5 and H5, which are displaced upward by 0.5105 at 0.1197 Bohr, respectively.

Table V. Approximate Linear Relationship between Molecular Fragment and More Extensive Basis Sets[a]

Molecule	a	b	S[b]	ρ[c]	Ref[d]
C_2H_6 (staggered)	0.8847	−0.1490	0.0093	0.9995	25
C_2H_6 (eclipsed)	0.8904	−0.1486	0.0138	0.9989	25
C_2H_4	0.9096	−0.1447	0.0166	0.9983	26
C_6H_6 (benzene)	0.8927	−0.1907	0.0194	0.9977	27
C_6H_6 (fulvene)	0.8981	−0.1983	0.0171	0.9981	28
$C_{10}H_{14}$ (naphthalene)	0.8878	−0.2024	0.0150	0.9986	21
$C_{10}H_{14}$ (azulene)	0.9078	−0.1809	0.0229	0.9966	21
CH_3OH (staggered)	0.7558	−0.2917	0.0173	0.9992	29
CH_3OH (eclipsed)	0.7543	−0.2924	0.0183	0.9992	29
H_2O_2 (trans)	0.7189	−0.3984	0.0310	0.9985	30
H_2CO	0.8423	−0.3392	0.0215	0.9989	31
HCOOH	0.8004	−0.3448	0.0444	0.9961	32
C_4H_4O (furan)	0.8571	−0.2217	0.0260	0.9976	33
CH_2NH	0.8510	−0.2046	0.0291	0.9972	34
N_2H_2 (trans)	0.8441	−0.2776	0.0636	0.9910	35
N_2H_2 (cis)	0.8317	−0.2790	0.0427	0.9960	35
C_6H_5N (pyridine)	0.8729	−0.2114	0.0168	0.9986	24
$C_4H_4N_2$ (pyrazine)	0.8608	−0.2195	0.0235	0.9978	24
C_4H_5N (pyrrole)	0.8376	−0.2590	0.0188	0.9984	36
$HCONH_2$	0.8509	−0.3543	0.0199	0.9989	37
$H_3N^+CH_2COO^-$	0.8571	−0.3625	0.0201	0.9990	38, 39

[a] The coefficients "a" and "b" in this table are those of eq 6 in the text. [b] S is the root-mean-square deviation from the line, *i.e.*

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\epsilon_i^{A} - (b + a\epsilon_i^{MF})]^2}$$

[c] ρ is the correlation coefficient, *i.e.*

$$\rho = \frac{\sum_{i=1}^{n}(\epsilon_i^{MF} - \bar{\epsilon}^{MF})(\epsilon_i^{A} - \bar{\epsilon}^{A})}{\sqrt{\sum_{i=1}^{n}(\epsilon_i^{MF} - \bar{\epsilon}^{MF})^2 \sum_{i=1}^{n}(\epsilon_i^{A} - \bar{\epsilon}^{A})^2}}$$

where $\bar{\epsilon}^{MF} = \sum_{i=1}^{n}\epsilon_i^{MF}/n$ and $\bar{\epsilon}^{A} = \sum_{i=1}^{n}\epsilon_i^{A}/n$. [d] This is the reference to the extended basis set investigation that was used for comparison.

in σ-orbital charge distribution trends were not observed using the FSGO basis.

Analysis of Results

The analysis of these results within the nonrelativistic, Born–Oppenheimer, Hartree–Fock framework is aided substantially in several ways, resulting in part from the particular characteristics of the molecular fragment procedure and in part from the use of FSGO as basis orbitals. First, the use of an *ab initio* framework allows attention to be focused upon the basis set, since neglect or approximations in integral evaluations are not present. Second, the localized nature of the FSGO basis orbitals will be seen to allow a straightforward identification of basis set strengths and deficiencies. These characteristics allow an intuitively and mathematically satisfying explanation of the results to be given and suggest clear and demonstrable ways in which the procedure can be improved, without sacrificing the ability to treat large molecular systems.

The principle that appears useful in rationalizing the results obtained to date which is consistent with the previous comments is one that analyzes the results in terms of the flexibility of the virtual orbital description of the molecule in question. To illustrate this principle, several specific examples will be considered. First, it is useful to consider the analysis of the hydrogen bonding studies on polypeptides.

As noted earlier, studies on polypeptides of glycine indicate that the presence of intramolecular hydrogen bonding stabilization at specific sites in the α-helix conformation can be detected by analysis of the changes in trends of π-orbital populations as a function of conformation but that corresponding σ-orbital reorganizational effects are not observed. To rationalize these results, let us consider the σ-basis orbitals in the vicinity of the hydrogen bonding region, as depicted in Figure 1.

The localized nature of the FSGO can now be seen to be of considerable utility in the examination of the results. In particular, we note that (see Table V in ref 15) the population of each of the various symmetrically orthogonalized σ-basis orbitals in the hydrogen bonding region is very nearly equal to 2.0 in all of the conformations. This indicates that the contribution of the various FSGO within that region is localized there and that other FSGO outside the region do not contribute substantially. Within that assumption, we note that, before the rotation to the conformation in which intramolecular hydrogen bonding is a possibility, each of the five orbitals in the region is doubly occupied. In the language of Hartree–Fock theory, there are no virtual orbitals in the region. Thus, when the two moieties are brought into proximity by rotation to the α-helix conformation and the SCF calculation is carried out, no change in the total charge distribution defined by these orbitals will occur, since unitary transformations among all doubly occupied orbitals leaves the charge distribution defined by them invariant.[41] Therefore, the lack of observation of charge reorganization in the σ system is not unexpected, since no virtual orbitals are present within the region of interest.[42]

(41) See, for example, F. L. Pilar, "Elementary Quantum Chemistry," McGraw-Hill, New York, N. Y., 1968, p 345.

(42) Of course, this observation also suggests a variety of possible ways in which this situation can be remedied by the appropriate introduction of additional orbitals. Several of these are currently being investigated and will be reported at a later date.

Table VI. Molecular Orbital Structure of Pyridine and Pyrazine

Pyridine				Pyrazine			
Molecular fragment[a]		Petke, Whitten, and Ryan[b]		Molecular fragment[a]		Petke, Whitten, and Ryan[b]	
MO sym	$-\epsilon$	MO sym	$-\epsilon$	MO sym	$-\epsilon$	MO sym	$-\epsilon$
$2a_2(\pi^*)$	−0.278	$2a_2(\pi^*)$	−0.110	$1a_u(\pi^*)$	−0.248	$2b_{3u}(\pi^*)$	−0.079
$1a_2(\pi)$	0.233	$1a_2(\pi)$	0.406	$6a_g$	0.250	$1b_{1g}(\pi)$	0.440
$11a_1$	0.243	$2b_1(\pi)$	0.426	$1b_{1g}(\pi)$	0.272	$6a_g$	0.447
$2b_1(\pi)$	0.267	$11a_1$	0.443	$5b_{1u}$	0.341	$1b_{2g}(\pi)$	0.482
$7b_2$	0.406	$7b_2$	0.556	$1b_{2g}(\pi)$	0.343	$5b_{1u}$	0.541
$1b_1(\pi)$	0.436	$1b_1(\pi)$	0.582	$3b_{3g}$	0.453	$3b_{3g}$	0.602
$10a_1$	0.453	$10a_1$	0.619	$1b_{3u}(\pi)$	0.497	$1b_{3u}(\pi)$	0.634
$6b_2$	0.485	$6b_2$	0.648	$4b_{2u}$	0.519	$4b_{2u}$	0.686
$5b_2$	0.543	$5b_2$	0.691	$4b_{1u}$	0.608	$4b_{1u}$	0.739
$9a_1$	0.566	$9a_1$	0.694	$3b_{2u}$	0.612	$3b_{2u}$	0.746
$8a_1$	0.595	$8a_1$	0.760	$5a_g$	0.617	$5a_g$	0.789
$4b_2$	0.788	$4b_2$	0.884	$2b_{3g}$	0.846	$2b_{3g}$	0.932
$7a_1$	0.795	$7a_1$	0.891	$4a_g$	0.881	$4a_g$	0.957
$3b_2$	0.983	$3b_2$	1.075	$2b_{2u}$	1.041	$2b_{2u}$	1.122
$6a_1$	1.053	$6a_1$	1.128	$3b_{1u}$	1.224	$3b_{1u}$	1.279
$5a_1$	1.238	$5a_1$	1.295	$3a_g$	1.320	$3a_g$	1.371
$4a_1$	9.228	$4a_1$	11.380	$2a_g$	9.383	$2b_{1u}$	11.442
$3a_1$	9.284	$2b_2$	11.380	$1b_{2u}$	9.415	$1b_{3g}$	11.442
$2b_2$	9.307	$3a_1$	11.388	$2b_{1u}$	9.497	$2a_g$	11.443
$2a_1$	9.427	$2a_1$	11.407	$1b_{3g}$	9.502	$1b_{2u}$	11.443
$1b_2$	9.430	$1b_2$	11.407	$1a_g$	13.130	$1a_g$	15.685
$1a_1$	13.088	$1a_1$	15.639	$1b_{1u}$	13.149	$1b_{1u}$	15.685

[a] See ref 8. [b] See ref 24.

Table VII. Hellmann–Feynman Electric Fields at Nuclei in Selected Molecules[a]

Type of molecule	Electric field[b] C	N	O	H
Ethane (staggered)[c]	0.061			0.023
Propane (staggered–staggered)[c]	0.076			0.043
Ethylene[c]	0.146			0.097
Benzene[c]	0.096			0.066
Naphthalene[c]	0.102			0.066
Formamide	0.031	0.028	0.003	0.244
N-Methylacetamide	0.055	0.007	0.010	0.182[d]
Hydrogen peroxide			0.075	0.151
Methanol	0.047		0.041	0.048
Dimethyl ether	0.041		0.048	0.041
Furan	0.118		0.043	0.045
Formaldehyde	0.052		0.032	0.062
Acetaldehyde	0.074		0.087	0.056
Acetone	0.040		0.027	0.059
Formic acid	0.080		0.025	0.082
Acetic acid	0.054		0.069	0.075
Methyl formate	0.047		0.093	0.166

[a] Calculated by the use of $|\epsilon(A)| = [\epsilon_x^2(A) + \epsilon_y^2(A) + \epsilon_z^2(A)]^{1/2}$ where $\epsilon_i(A)$ is the ith component of the Hellmann–Feynman electric field at point A. See ref 4 for the further details concerning the actual calculation of these fields. Only the largest value for a particular type of atom is reported. [b] Reported in hartree atomic units, see ref 18. [c] See ref 4. [d] All other hydrogen atoms had $|\epsilon| \leq 0.027$.

Table VIII. Dipole Moment Data for Representative Molecules[a]

Molecule	Calcd dipole moment	Obsd dipole moment[b]
C_3H_8	0.0265[c]	0.083
H_2O_2	2.279[d]	2.26
CH_3OH	2.020	1.69
CH_3OCH_3	1.731	1.31
Furan	0.448	0.661
H_2CO	2.145	2.339
HCOOH	1.666	1.415
$HCOOCH_3$	1.574	1.77
$HCONH_2$	3.78	3.72
N-Methylacetamide	3.73[e]	3.71
$H_3N^+CH_2COO^-$	13.33[f]	13.3

[a] Values are given in Debyes. [b] See ref 4 and 7 for references to experimental measurements. [c] Averaged over the staggered–eclipsed and eclipsed–eclipsed forms. [d] Calculated at the minimum energy conformation. [e] The dipole moment vector forms an angle of 54.2° with the CN bond. [f] The dipole moment vector forms an angle of 29.1° with the CC bond.

Thus, the concept that appears to be of considerable importance in rationalizing the results obtained in the initial formulation of the molecular fragment procedure is that, in order to assure appropriate flexibility with respect to charge redistribution possibilities, at least one virtual orbital should be present in each region of space where charge redistribution is needed. Thus, the localized nature of the basis set plus the *ab initio* framework allow at least a qualitative, if not quantitative, definition of the notion of basis set balance[43a] for the molecular fragment procedure. To see how this notion can be used in the rationalization of the other data already available, let us consider several additional examples.

For example, the changes in charge distribution trends that *are* observed within the π orbitals of the glycine polypeptides in the α-helix conformation beginning with the tetrapeptide are also consistent with the notion just introduced. In particular, each of the amide π systems has a virtual orbital that is localized primarily in the C–O region, which is available for redistribution of charge corresponding to hydrogen bond formation in the α-helix conformation. This allows ready identification of hydrogen bonding effects in such systems directly from the *ab initio* calculations. In addition, since the primary contribution to the energetics of hydrogen bond formation is electrostatic,[43b] the slight basis set defect due to the lack of σ-orbital flexibility is not expected to cause serious difficulties.

Turning next to the question of molecular orbital

(43) (a) R. S. Mulliken, *J. Chem. Phys.*, **36**, 3428 (1972); (b) S. Bratož, *Advan. Quantum Chem.*, **3**, 209 (1967).

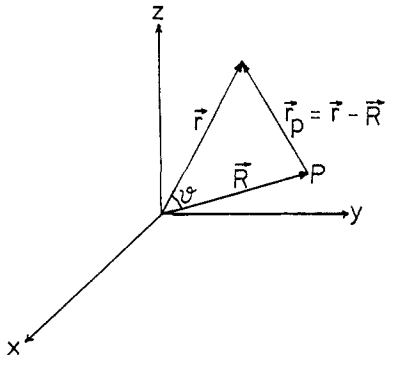

Figure 2. Depiction of an FSGO located at point P to be related to an origin (nucleus) on which GTO basis orbitals are placed.

structure predictions, let us consider the pyridine and pyrazine examples of Table VI. In the case of pyridine, it is seen that the only discrepancy in valence molecular orbital ordering compared to extensive basis set calculations occurs in the case of the $11a_1$ and $2b_1(\pi)$ molecular orbitals. Examination of the nature of the $11a_1$ orbital reveals that it is primarily a lone pair orbital, located in the vicinity of the nitrogen atom and in the plane containing the nuclei and other σ orbitals. Since this lone pair orbital can be thought of as having only one primary contributor, the single FSGO that is doubly occupied, it is seen that a lack of virtual orbitals in that region of space is again observed as a likely cause of the basis set imbalance that is reflected in a misordering of molecular orbitals.

A similar situation arises in the case of pyrazine, where both the ($5b_{1u}$, $1b_{2g}(\pi)$) pair and the ($6a_g$, $1b_{1g}(\pi)$) pair of valence molecular orbitals are incorrectly ordered, as measured by comparison with more extensive basis set calculations. As in the previous example, the $5b_{1u}$ and $6a_g$ orbitals in pyrazine are found to be primary σ lone pair orbitals in the vicinity of the two nitrogen atoms that are each described by single FSGO that are doubly occupied. Thus, the lack of virtual orbitals in the lone pair region is again identified as a likely source of the misordering of molecular orbitals.

An example that illustrates this point, where lone pairs are not present, is the case of the ethane molecule. As indicated in Table III, the C–C bond distance prediction is in error by 9.1% and the calculated barrier to rotation is too high. Each of these deficiencies can also be rationalized in terms of a lack of virtual orbitals in the appropriate regions of space. In this case, it is the C–H bond description that appears to be the source of the difficulty. In particular, each C–H bond is described by a single doubly occupied FSGO, with no virtual orbitals in the same region. Thus, the interfragment interactions among these orbitals that bear strongly upon the calculated barrier to rotation and the C–C distance prediction are not adequately described.[44]

Examples such as these illustrate several general characteristics of the molecular fragment procedure as initially formulated, as well as when satisfactory results are to be expected. They also indicate several obvious possibilities for basis set improvement at minimum additional cost. An example of an initial study related to this will be given later in the discussion. However, at this point it is important to note that the use of reasonably localized basis orbitals within an *ab initio* framework allows an ease of analysis for identification of sources of inadequacies and strengths that is not generally shared with other (*e.g.*, semiempirical) procedures that are frequently used to describe large molecular systems.

Other ways in which both the flexibility of FSGO as basis orbitals and the utility of the molecular fragment approach are reflected can also be identified. For example, determination of the position and size of the various FSGO *via* studies of molecular fragments results typically in quite low values of Hellmann–Feynman forces and fields,[45] compared to other basis sets.[4] This procedure also is related to the near satisfaction of the virial ratio typically found[46] and is likely to be an important factor in the good description of molecular properties (*e.g.*, dipole moments) that has been frequently observed.[4,13,15] It might also be noted that examination of the Hellmann–Feynman forces and fields in large molecules are frequently also sensitive measures of the suitability of the fragment description. In particular, if large magnitudes are observed at some atoms, then examination of the adequacy of the fragment description for the atom in that particular environment is usually appropriate.

Such observations also serve to emphasize the utility of molecular fragments in general as appropriate frameworks for basis orbital determination as opposed to, *e.g.*, determination of basis orbitals for molecular calculations *via* studies on atoms. While simulation of the anticipated molecular environment during basis set determination is of obvious importance when small basis sets for large molecules are desired, it also is of importance in large basis set calculations, where the use of FSGO, in addition to functions on nuclei that are used to obtain a good nuclear–electron cusp description,[47] can provide considerable economies in the description of valence electrons.

The inherent flexibility of FSGO as basis orbitals is also of interest to note. To illustrate the natue of this flexibility, it is useful to show an explicit relationship between FSGO basis orbitals and atomic Gaussian orbitals. In Figure 2 an FSGO at point P is assumed and will be related to an atomic Gaussian orbital basis at the origin (where a nucleus could be located). From this figure it is easily seen that a normalized FSGO at point P can be expressed in terms of functions at the origin[48] as follows

$$G(r_P) = Ne^{-(r_P^2/\rho_P^2)} \tag{7}$$

$$= \exp\{-(r^2 + R^2 - 2rR\cos\vartheta)/\rho^2\} \tag{8}$$

$$= N\left\{1 + \left(\frac{2R}{\rho^2}\right)r\cos\vartheta + \frac{1}{2}\left(\frac{2R}{\rho^2}\right)^2 r^2 \times \cos^2\vartheta + \ldots\right\}\exp\left\{-\frac{(r^2 + R^2)}{\rho^2}\right\} \tag{9}$$

(44) J. L. Nelson and A. A. Frost (*Theor. Chim. Acta*, **29**, 75 (1973)) have also found similar problems with the prediction of the barrier to rotation in ethane, using an FSGO model which takes each FSGO as doubly occupied. However, it should be noted that certain bond orbital descriptions containing three or four FSGO do give rise to a lower rotational barrier, indicating that an effect similar to that observed when appropriate SCF virtual orbitals are present can be obtained within the minimum basis orbital model.

(45) A. C. Hurley, *Proc. Roy. Soc., Ser. A*, **226**, 179 (1954).
(46) P. O. Löwdin, *J. Mol. Spectrosc.*, **3**, 46 (1959).
(47) J. C. Slater, *Phys. Rev.*, **36**, 57 (1930).

where N is the normalization constant for $G(r_P)$. The series expansion of $\exp[+(2rR/\rho^2)\cos\vartheta]$ in eq 9 is convergent for all r, but is of particular interest here when $(rR) < \rho^2$. This corresponds to the placement of the FSGO near the origin, and considering particularly the situation where the electron is within the orbital radius (ρ). In this case an even faster rate of convergence will be present.

Examination of eq 9 reveals that, relative to an s, p, d, ... Gaussian basis set (GTO) at the origin, the FSGO at point P contains *all* GTO components, *i.e.*, s, p, d, f, ... In other words, one of the important aspects of basis set flexibility that is included as soon as off-nuclei FSGO are introduced into the basis set is that components of all atomic GTO basis orbitals are then present in the basis. In addition, comments similar to the above can be made when FSGO are compared to other atomic basis sets (*e.g.*, Slater-type orbitals) since, in that case, all of the possible *angular* dependences encountered in an atomic orbital basis are also represented in eq 9.

Also, the four parameters (ρ, R_x, R_y, R_z) that are available for each FSGO that is added can be considered to provide greater flexibility per basis orbital than the one nonlinear parameter that is available for each atomic orbital (*e.g.*, STO) that is added. Finally, the molecular fragment procedure appears to be a particularly useful approach for determination of the FSGO nonlinear parameters, in that it does not require treatment of each molecule as a separate case and determines the FSGO parameters in environments that closely resemble the anticipated molecular environment.

However, just the presence of all angular components does not guarantee that the flexibility needed to assure proper basis set balance will be present. In particular, at least two other aspects must be considered. The first of these involves the *flexibility* with which *each* component is introduced, and the second concerns the *number* of each component that is present. In terms of the example just considered, we note that a relatively limited flexibility of each angular component is introduced with FSGO basis orbitals. This arises since only four parameters (ρ, R_x, R_y, R_z) are available to be chosen, and examination of eq 9 shows that the relative importance of the various components is fixed, instead of introducing each component independently. In addition, a single FSGO obviously contributes only one s, one p, etc., component, so that the number of each component is also limited.

Hence, the determination of whether FSGO or atomic orbitals form a better basis set (for a given basis set size) is not straightforward, since each kind of basis orbital possesses characteristics that could be advantageous, depending upon the particular molecule under consideration. However, there are some examples that indicate that the features introduced by FSGO are of perhaps greater importance than the advantageous nuclear cusp description typically obtained using s and p atomic basis orbitals.

One illustration of this point is given by the H_2O_2 molecule, where the use of an s- and p-type STO basis

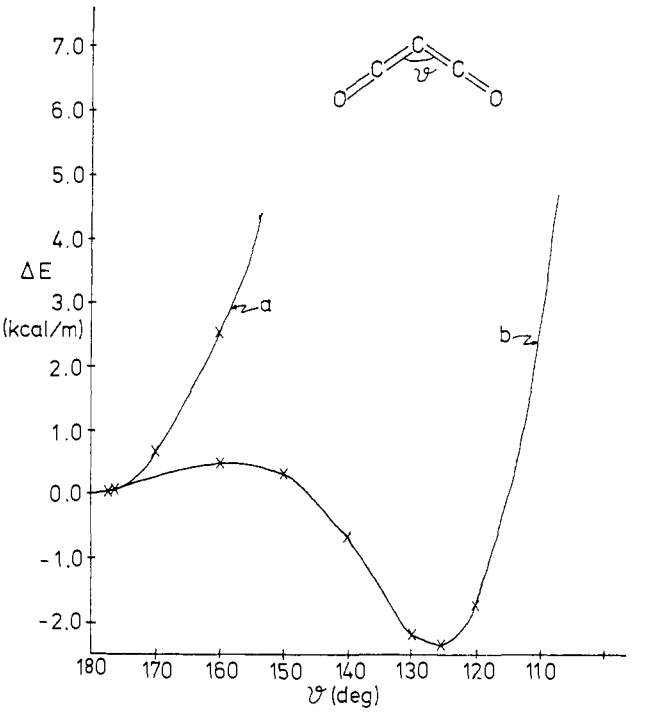

Figure 3. Variation of total energy of C_3O_2 as a function of bending at the central carbon atom. Curve "a" corresponds to an *ab initio* calculation in which s- and p-type STO basis orbitals are used (ref 51), and curve "b" is the result obtained using FSGO basis orbitals (ref 14).

is incapable[49] of reproducing the trans barrier that is known to exist from experimental measurements.[50] However, the description of H_2O_2 using only ten FSGO within the initial molecular fragment formulation does indicate a trans barrier that is duplicated only when the STO basis set contains d orbitals on the oxygen atoms and p orbitals on the hydrogen atoms. Thus, it appears that the presence of the d components introduced by the use of FSGO is of primary importance in reproducing the trans barrier, while the number and flexibility of the various components is of lesser importance.

Another example in which comparisons can be made is the carbon suboxide molecule, C_3O_2. While *ab initio* calculations in which s- and p-type STO basis orbitals are used[51] indicate that C_3O_2 should be linear, studies using FSGO basis orbitals within the molecular fragment procedure[14] indicate a bent structure. The qualitative, as well as quantitative, features of these differences are indicated in Figure 3, where the two curves have been set equal at $\vartheta = 0°$ to allow ease of comparison of the shapes of the calculated curves. The noticeable qualitative differences in the shape of the two results indicate that investigation in greater detail is of interest.

To carry this investigation further, it is helpful to consider the C_3 molecule which, similar to C_3O_2, also possesses a low-frequency bending vibration[52] about the central carbon of approximately 60 cm^{-1}. An *ab inito* study on C_3 using a variety of GTO basis sets has

(48) Similar analyses have been used in other contexts. See, for example, A. G. Turner, B. H. Honig, R. G. Parr, and J. R. Hoyland, *J. Chem. Phys.*, **40**, 3216 (1964); or M. P. Barnett and C. A. Coulson, *Phil. Trans. Roy. Soc. London*, **243**, 221 (1951).

(49) See, for example, R. M. Stevens, *J. Chem. Phys.*, **52**, 1397 (1970).

(50) R. H. Hunt, R. A. Leacock, C. W. Peters, and K. T. Hecht, *J. Chem. Phys.*, **42**, 1931 (1965).

(51) J. R. Sabin and H. Kim, *J. Chem. Phys.*, **56**, 2195 (1972).

(52) L. Gausset, G. Herzberg, A. Lagerquist, and A. Rosen, *Discuss. Faraday Soc.*, **35**, 113 (1936); *Astrophys. J.*, **142**, 45 (1965).

Table IX. Optimized Molecular Fragment Data for Fragments Used to Form the C_3 Molecule

Fragment	Orbital radius (ρ)	Distance from the heavy atom	Molecular parameters
HC≡CH	ρ_{CH} = 1.62648438	1.07226561	R_{CH} = 1.99555080
	ρ_C = 0.32678164	0.0	R_{CC} = 2.27523029
	$\rho_{C\equiv C}$ = 1.38667968	1.13761719	
	ρ_π = 1.78320312	±0.1	
HC≡C·	ρ_C = 0.32624179	0.00114062[a]	R_{CH} = 1.99555080
	$\rho_{C\equiv C}$ = 1.66757812	1.47433591	R_{CC} = 2.51000000
	ρ_{LP} = 0.25371094	1.80777341	
	ρ_π = 1.81824216	±0.1	

[a] This FSGO is displaced from the C atom along the CH line, in the direction of the H atom.

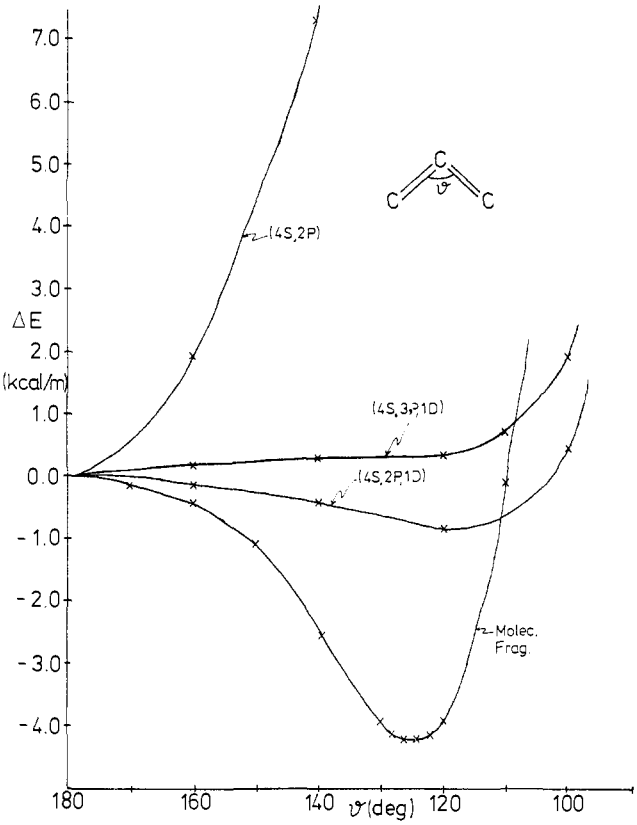

Figure 4. Variation of total energy of C_3 as a function of bending at the central carbon atom. See the text for further discussion of the basis sets used in constructing the SCF wave functions.

also been carried out,[53] allowing several comparisons to be made. Several of these studies at the SCF level are indicated in Figure 4, along with the results obtained using the molecular fragment procedure. The molecular fragment and optimized parameter values that were used to form the C_3 molecule are listed in Table IX.

In Figure 4, one curve corresponds to the use of a (4s,2p) GTO basis on each nucleus, other calculations employed a (4s,2p,1d) and (4s,3p,1d) GTO basis on each nucleus, and the lowest curve presents the results obtained using the molecular fragment procedure. Each of the curves has been drawn relative to the same energy value at $\vartheta = 180°$ for convenience of comparison. As is evident from the figure, the d-orbital components are of considerable importance and change even the qualitative features of the curve. It is significant that the d-orbital components introduced through the FSGO basis set give rise to a curve whose shape is similar to that obtained when independent d orbitals are introduced directly (using the (4s,2p,1d) basis), although the FSGO basis is a great deal smaller and considerably easier to handle computationally. However, these studies also indicate that the small FSGO basis used in the initial formulation is not sufficiently flexible to reproduce the shape of the curve obtained using the extensive (4s,3p,1d) basis and that more than simply the presence of higher orbital components is needed. Thus, the C_3 molecule appears to be an example where all of the basis set flexibilities are needed, *i.e.*, higher orbital components as well as a nonminimum number and considerable flexibility in each component.

Table X. Comparison of *ab Initio* Results for Acetylcholine

Conformer	τ_1	τ_2	τ_3	ΔE, kcal/mol[a] STO-3G[b]	ΔE, kcal/mol[a] Mol frag[c]
Trans	180	180	180	0.0	0.0
Near gauche	180	80	180	−0.8	+10
Gauche	180	60	180	+3.4	+19

[a] The trans conformer has been used as the zero of energy. [b] See ref 55. [c] See ref 12.

Finally, returning to the case of C_3O_2, it seems likely by analogy to the discussion concerning C_3 that the d-orbital components of the FSGO basis are the cause of the qualitative, as well as quantitative, differences in the results of the molecular fragment study and the (s,p) atomic Gaussian basis set studies. However, it should not necessarily be inferred from these results that the molecular fragment study provides definitive results regarding the shape of C_3O_2. In particular, since the energy differences between $\vartheta = 180°$ and $\vartheta \simeq 120°$ are so small, it is possible that correlation effects may also affect the shape of the curve substantially, as observed in the study[53] of C_3. In addition, the C_3 studies suggest that expansion of the basis may also be needed to obtain definitive results for C_3O_2. Hence, while additional studies on C_3O_2 including configuration interaction effects are appropriate before the question of the geometric structure of C_3O_2 can be settled satisfactorily, the flexibility and appropriateness of FSGO basis orbitals for the description of problems of molecular structure is apparent.

Another example of a problem involving geometric structure in which the d, f, ... orbital components of the FSGO basis set may be important is in the case of acetylcholine. In particular, the use of an STO-3G Gaussian representation of a minimum s and p atomic Slater-type orbital basis set[54] results in a gauche conformer being found to be more stable than the trans conformer,[55] as indicated in Table X. Also, given in

(53) D. H. Liskow, C. F. Bender, and H. F. Schaeffer III, *J. Chem. Phys.*, **56**, 5075 (1972).

(54) W. G. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).

this table are the results obtained using the molecular fragment procedure,[12] which are seen to predict a result which is opposite to that of the STO-3G basis set study.

Since both investigations were *ab initio* studies using the same geometries,[56] the results can be viewed as reflections of the differences in basis set characteristics. In particular, for the case of the STO-3G studies, a minimal Slater-type basis of 1s, 2s, and 2p orbitals was used. On the other hand, the molecular fragment studies introduce s, p, d, f, ... atomic orbital components (on the ether oxygen in particular) through the use of off-center FSGO that represent lone pairs and bonds. However, the STO-3G calculations involve a total of 66 independent functions in the SCF calculation, while the molecular fragment calculation employed only 50 independent functions. Thus, the question of which basis set provides the more adequate description of acetylcholine is not easily answered, since each of the basis sets has both advantages and deficiencies. On the other hand, the previous discussion of H_2O_2 and cumulenes suggests that the d, f, ... components introduced by the use of FSGO as basis orbitals may be important factors in the description of the nonbonded interactions that occur in the various conformations.[57] However, considerable additional investigations are needed (including perhaps basis set expansion and estimation of the importance of correlation effects) before definitive results on the geometric structure of acetylcholine can be obtained.

Summarizing the above results, it would appear that FSGO are particularly convenient and appropriate for the description of valence orbitals, where nuclear cusp characteristics are not as crucial. Thus, the fragments of Table I and associated parameters of Table II are expected to provide a suitable basis for the description of many qualitative and quantitative properties of a wide variety of large molecules, subject to the limitations described here.

On the basis of the analysis of virtual orbitals made earlier, it is also useful to consider how improvements in the basis set can be made. As an illustration of the possibilities, let us consider the methane fragment. For this fragment, some initial studies have been carried out that illustrate both the utility of such an analysis and the ease with which appropriate improvements in the basis can be made without significant increase in computational requirements.

In order to introduce a virtual orbital in the C–H bonding regions, as suggested by the previous analysis, a CH_4 fragment as described in Figure 5a was investigated. As can be seen in the CH_4 fragment, an additional FSGO has been added on each hydrogen nucleus. Optimization of the CH_4 fragment was then carried out, using a single FSGO to describe the inner shell and a linear combination of two FSGO to describe each C–H bond, *i.e.*

$$\chi^{CH} = N(G_{CH} + G_H) \qquad (10)$$

(55) G. N. J. Port and A. Pullman, *J. Amer. Chem. Soc.*, **95**, 4059 (1973).

(56) Contrary to the comments in ref 55, an STO-3G calculation does not represent a "standard" *ab initio* calculation but rather just one of a variety of possible calculations that happens to use a minimum Slater-type orbital basis within an *ab initio* framework.

(57) For another recent example of a case where d-orbital inclusion is essential to the description of the ground state geometry, see the discussion of CaF_2 by J. L. Gole, A. K. Q. Siu, and E. F. Hayes, *J. Chem. Phys.*, **58**, 857 (1973).

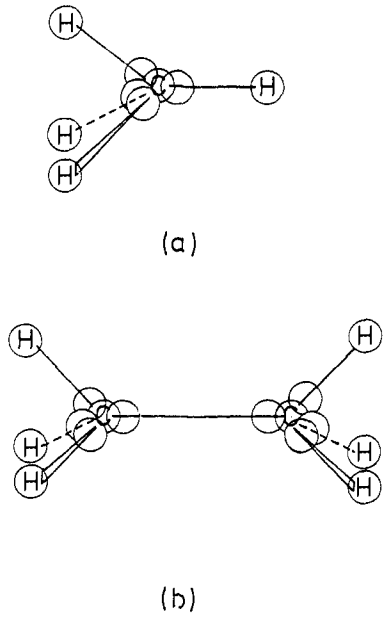

Figure 5. (a) Possible modified CH_4 fragment description. The positions of the various FSGO are approximately correct. However, the orbital radii vary widely. (See Table XI.) (b) Formation of C_2H_6 from two CH_4 fragments. Approximate FSGO positions are indicated by circles.

Table XI. Optimized CH_4 Fragment Parameters and Ethane Molecular Orbital Structure Using Modified Fragment Description

A. CH_4 Fragment[a]		
FSGO description	ρ	Distance from carbon atom
Inner shell	0.32597560	0.0
C–H	1.57260990	2.08249000
	1.68948139	0.23000000

B. C_2H_6 Molecular Orbital Structure[b]	
MO symmetry	Orbital energy (ϵ)
$3a_{2u}$	0.592583
$2e_u$	0.522722
$1e_g$	−0.403103
$3a_{1g}$	−0.414200
$1e_u$	−0.522737
$2e_{2u}$	−0.782602
$2a_{1g}$	−0.980322
$1a_{1g}$	−9.354105
$1a_{2u}$	−9.410567
E_{total} =	−67.81091360

[a] CH distance = 2.08249 Bohrs, tetrahedral symmetry. [b] The staggered conformer of C_2H_6 is used (D_{3d} symmetry), and the CC bond distance is taken to be 2.907 Bohrs, the distance corresponding to the calculated minimum energy. The CH bond distance was taken to be 2.0938 Bohrs. The calculated barrier to rotation is 5.129 kcal/mol.

Optimized values of the various FSGO parameters for this particular fragment are given in Table XI.

The formation of ethane (indicated in Figure 5b) takes place in a manner similar to that in the original molecular fragment formulation, except that the FSGO that lie on the hydrogen atoms which would lie in the C–C bonding region are removed from the basis. This provides a basis set for the C–C bond description that is now balanced to approximately the same extent as the basis for C–H bond description; *i.e.*, one virtual

orbital is present in each bonding region. The molecular orbital description of ethane, analogous to the original formulation, is obtained *via* an SCF calculation but using each of the FSGO as independent basis orbitals in the formation of each molecular orbital (*i.e.*, the linear coefficients of the two CH FSGO are discarded when C_2H_6 is formed, and the two FSGO in the CH region are treated as independent basis orbitals). Also given in Table XI is the molecular orbital structure of the ethane molecule, calculated at a C–C distance of 2.907 Bohrs, which is the distance corresponding to the calculated minimum energy.

There are several points of interest to note in Table XI. First, instead of a 9.1% error in the predicted C–C distance that occurred in the initial formulation, it is seen that the error is reduced to 0.3% using the modified fragment description. Next, comparison of the molecular orbital ordering in Table XI with the results of extensive basis set studies[25] in Table VI shows that the molecular orbital ordering remains in exact agreement with the extensive basis set studies. Also, an orbital energy linearity plot results in $a = 0.9122$, $b = -0.1141$, $S = 0.0041$, and $\rho = 0.9999$, indicating the improved balance of the modified basis set in another manner. The barrier to rotation using this fragment is calculated to be ~5.1 kcal/mol, roughly the same as in the initial formulation, where a 5.6 kcal/mol barrier was found.[58]

It should be emphasized that these are preliminary investigations of an improved molecular fragment description. However, they illustrate clearly the ease of analysis and improvement of the approach. In addition, the substantial improvement observed in geometric and electronic structure characteristics at relatively small computational expense provides considerable encouragement that an "analytical tool" that is applicable to large molecular systems and which also possesses acceptable accuracy for a variety of properties of interest to chemists and biologists can be achieved. Additional efforts are currently underway to develop a general approach for obtaining optimized fragments, similar to the one just described, that will allow for large molecules not only improved accuracy of molecular properties such as those just discussed but will also allow examination of systems in which an *a priori* choice of molecular fragment for description of the system is not obvious (*e.g.*, development of a fragment that is not restricted to description of only "sp^3" environments but is continuously "rehybridizable" into any desired mixture of "sp^3" and "sp^2" hybridization).

(58) No attempt was made in these initial studies to optimize the fragment description with respect to barrier to rotation predictions.

Acknowledgments. The authors would like to express their appreciation to the University of Kansas for partial support of the computing time required in this work, as well as to Dr. L. L. Shipman for investigating some of the orbital energy linearity relationships of Table V. In addition, several informative conversations with Dr. J.-P. Malrieu and Dr. M. G. Clark, and the calculation of Hellmann–Feynman fields for several oxygen analogs by Dr. B. V. Cheney, are gratefully acknowledged.

Calculations of Barriers to Internal Rotation in Propene and Monofluoropropenes

Alan D. English and William E. Palke*

Contribution from the Department of Chemistry, University of California, Santa Barbara, California 93106. *Received August 2, 1973*

Abstract: Barriers to internal rotation have been calculated by the Hartree–Fock method for propene, 2-fluoropropene, *cis*-1-fluoropropene, and *trans*-1-fluoropropene. The agreement of the calculated and experimental values is sufficiently close to suggest that Hartree–Fock calculations can be used quantitatively with reasonable error limits.

Propene and its fluorinated derivatives form an interesting sequence of molecules for the study of barriers to internal rotation. Only propene has undergone a complete microwave structure determination, but the barriers and some structural information have been found for all the monofluorinated propenes. We have calculated the barriers in propene, *cis*-1-fluoropropene, *trans*-1-fluoropropene, and 2-fluoropropene in order to determine whether the Hartree–Fock method is sufficiently reliable to be trusted in determinations of barriers in molecules that are either unknown or whose microwave spectrum has yet to be analyzed.

Basis Sets and Geometries

The basis sets used are those suggested by Dunning[1] in which for each carbon and fluorine atom nine s- and five p-type Gaussian functions are contracted into four s- and three p-type basis functions. For each hydrogen, four s-type Gaussians are contracted into two basis functions. The methyl hydrogen basis functions were scaled to best fit a hydrogen orbital with exponent 1.159 (the optimized orbital exponent in ethane[2]); the remaining hydrogen basis functions were scaled to best fit the optimized ethylene hydrogen exponent (= 1.227).[2] The results of this paper further show that these basis sets are adequate for calculating barriers to internal rotation of methyl groups even though more extended basis sets were required to satisfactorily describe the hydrogen peroxide internal rotation potential.[3]

(1) T. H. Dunning, *J. Chem. Phys.*, **53**, 2823 (1970).

(2) R. M. Stevens, private communication.

(3) N. W. Winter and T. H. Dunning, *Chem. Phys. Lett.*, **11**, 194 (1971).