Ab Initio Molecular Orbital Theory

Warren J. Hehre

Department of Chemistry, University of California, Irvine, California 92717

Within little more than a decade, nonempirical or ab initio molecular orbital theory[1] has evolved from a specialized tool of the molecular physicist to a technique of general utility to the practicing chemist. Thus, it is now possible for one who is moderately well versed in theory, but who is by no means an expert, to use ab initio molecular orbital methods much as he would any one of a number of currently available experimental instrumental techniques (e.g., NMR or mass spectrometry). Perhaps the most attractive feature of nonempirical molecular orbital theory is its vast scope of application. It is capable of addressing itself to a variety of questions about species which are of too low thermal stability to be readily observed and characterized experimentally, or, in fact, about reaction transition states which, in the usual sense of the word, have no lifetime at all. Like even the most routine experimental technique, ab initio molecular orbital heory must be applied with some care in order that its results be of any value. This being the case, the findings of the theory—while not to be taken as experimental fact—should display a degree of internal consistency in which we might place a certain level of confidence.

In this Account we should attempt to convey to the reader our view of the current limits of application of ab initio molecular orbital theory to problems of chemical interest. We shall approach the topic by delineating a logical series of steps designed to take us from the simplest possible (and, hence, most widely applicable) level of theory, to the (Hartree–Fock) limit of the formalism. Each of these steps has associated with it its own particular "chemistry", and we shall take the time to point out what we believe to be its characteristic strengths and weaknesses and to indicate briefly what types of problems are readily approachable given present computer technology and which are beyond current range. We shall focus on two basic areas of application—structure and stability—domains which we feel have been central to the development of chemical theory. We shall concentrate heavily on the work of Pople and his past and present collaborators. This latter restriction is not meant to slight the efforts of others, for it is clear that the research described herein rests strongly on the work of previous generations of electronic structure theorists. Rather the efforts of the Pople school have been emphasized in order that the thread of logic, connecting the various levels of ab initio theory, be clearly exposed.

General Molecular Orbital Theory[2]

For diamagnetic species, molecular orbital theory starts by assigning electrons in pairs to spatial functions $\psi_1, \psi_2, \ldots \psi_n$, which are then used in the construction of a many-electron wavefunction as a single determinant

$$\Psi(1,2,\ldots 2n) = \frac{1}{2n!}$$

$$\times \begin{vmatrix} \psi_1(1)\alpha(1) & \psi_1(1)\beta(1)\ldots & \psi_n(1)\beta(1) \\ \psi_1(2)\alpha(2) & \cdots\cdots\cdots \\ \vdots \\ \psi_1(2n)\alpha(2n) & \psi_1(2n)\beta(2n)\ldots & \psi_n(2n)\beta(2n) \end{vmatrix}$$

where α and β are spin functions. The molecular orbitals ψ_i are so chosen as to minimize the total energy of the system, that is the expectation value of the many-electron Hamiltonian, $\mathcal{H}$

$$E = \int \ldots \int \Psi(1,2,\ldots 2n)\mathcal{H}$$
$$\Psi(1,2,\ldots 2n) \quad d\tau_1 d\tau_2 \ldots d\tau_{2n}$$

and are in practice written in terms of a linear combination of nuclear centered atomic functions, φ_μ.

$$\psi_i = \sum_\mu c_{\mu i}\varphi_\mu$$

The coefficients, $c_{\mu i}$, of the expansion are arrived at by solution of the Roothaan equations[2b]

$$\sum_\nu (F_{\mu\nu} - \epsilon_i S_{\mu\nu})c_{\mu i} = 0$$

where $S_{\mu\nu}$ is an overlap integral

$$S_{\mu\nu} = \int \varphi_\mu(1)\varphi_\nu(1)d\tau_1$$

ϵ_i is the one electron energy associated with ψ_i, and $F_{\mu\nu}$ is an element of the Fock matrix

$$F_{\mu\nu} = H_{\mu\nu} + \sum_\lambda \sum_\sigma P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - {}^1/_2 (\mu\lambda|\nu\sigma)]$$

Here $H_{\mu\nu}$ is the element

$$H_{\mu\nu} = \int \varphi_\mu(1)\left[-{}^1/_2 \nabla^2 \right] \varphi_\nu(1)d\tau_1$$

$$+ \sum_A^{\text{nuclei}} \int \varphi_\mu(1)\left[\frac{1}{r_{1A}} \right] \varphi_\nu(1)d\tau_1$$

Warren J. Hehre received his B.A. degree from Cornell University in 1967 where he studied under Roald Hoffmann, and his Ph.D. from Carnegie-Mellon University in 1971 with John Pople. After spending 2 years in Lionel Salem's lab at the University of Paris he joined the faculty of the University of California, Irvine. He is an Alfred P. Sloan Fellow for 1974–1976. His research interests include the development and application of simple molecular orbital methods, and investigations into the structures and stabilities of ions in the gas phase by ion cyclotron resonance spectroscopy.

(1) For reviews, see: (a) J. A. Pople, *Acc. Chem. Res.*, **3**, 217 (1970); (b) L. Radom and J. A. Pople, *MTP Int. Rev. Sci., Theor. Chem.*, 1972, 71 (1972); (c) H. F. Schaefer, III, "Critical Evaluation of Chemical and Physical Structural Informaion", D. R. Lide and M. A. Paul, Ed., National Academy of Sciences, Washington, D.C., 1974, p 591; (d) H. F. Schaefer, III, *Annu. Rev. Phys. Chem.*, in press.

(2) (a) J. A. Pople and D. L. Beveridge, "Approximate Molecular Orbital Theory", McGraw Hill, New York, N.Y., 1970; (b) C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951).
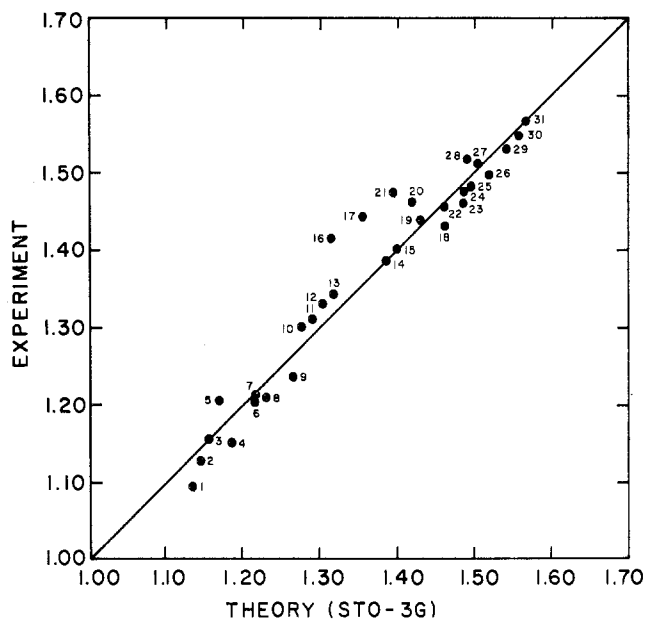
Figure 1. Experimental vs. theoretical (STO-3G) bond lengths. 1, N_2; 2, CO; 3, HCN; 4, NO; 5, C_2H_2; 6, H_2CO; 7, O_2; 8, HNO; 9, HNNH; 10, cyclopropene C=C; 11, CH_2=C=CH_2; 12, C_2H_4; 13, cyclobutene C=C; 14, CH_3F; 15, C_6H_6; 16, F_2; 17, HOF; 18, CH_2=CHC≡CH, C—C; 19, CH_3OH; 20, NH_2OH; 21, H_2O_2; 22, N_2H_4; 23, CH_3C≡CH; 24, CH_3NH_2; 25, CH_2=CHCH=CH_2 C—C; 26, CH_3CH=CH_2 C—C; 27, cyclopropane; 28, cyclopropene C—C; 29, C_2H_6; 30, cyclobutane; 31, cyclobutene C_3—C_4.

describing the motion of a single electron in a field of bare nuclei, and $(\mu\nu|\lambda\sigma)$ is a two-electron integral

$$(\mu\nu|\lambda\sigma) = \int\int \varphi_\mu(1)\varphi_\nu(1)\left[\frac{1}{r_{12}}\right]\varphi_\lambda(2)\varphi_\sigma(2)d\tau_1 d\tau_2$$

The one-electron density matrix, P, is given by

$$P_{\lambda\sigma} = 2\sum_i c_{\lambda i}c_{\sigma i}$$

where the summation is over the manifold of *occupied* molecular orbitals. Note that the Roothaan equations are not readily amenable to solution in closed form. This is simply because the quantity we seek (the $c_{\mu i}$) appears, disguised as $P_{\lambda\sigma}$, in their formulation. In practice, solution is achieved by some form of iterative procedure, for which a number of general computer programs have been developed in recent years and are now readily available.[3]

The Development of Basis Set Representations[4]

The simplest level of *ab initio* molecular orbital theory involves the use of a *minimal basis set* of nuclear centered functions. Here, each atom is described by exactly those functions which are necessary to accommodate all of its electrons, taking care to maintain an overall spherical symmetry. For hydrogen and the first row elements, boron to fluorine, this atomic orbital representation is:

H 1s
B to F 1s, 2s, $2p_x$, $2p_y$, $2p_z$

Although the number of orbitals per atom is completely determined, the nature of the functions themselves is left unspecified. For example, each atomic orbital might be represented by a single function—say, of exponential or Gaussian type—or by a linear combination of two or more of such functions. A number of criteria have been used for the selection of the functions which comprise a minimal basis set. For one, functions have been chosen so as to minimize the ground-state energy of the atom on which they are placed. Alternatively, the detailed choice of basis might represent an effort to reproduce other types of experimental data (e.g., relative energies and equilibrium geometries) for a wide spectrum of simple molecules. One minimal basis set, arrived at in this manner, has had remarkable success with regard to the calculation of a number of properties of small organic molecules. Termed STO-3G,[5] it is comprised of three-Gaussian expansions to Slater-type (exponential) functions[6]

$$\varphi_{1s}(\zeta = 1,r) = \sum_{k=1}^{3} d_{1s,k}g_{1s}(\alpha_{1k},r)$$

$$\varphi_{2s}(\zeta = 1,r) = \sum_{k=1}^{3} d_{2s,k}g_{1s}(\alpha_{2k},r)$$

$$\varphi_{2p}(\zeta = 1,r) = \sum_{k=1}^{3} d_{2p,k}g_{2p}(\alpha_{2k},r)$$

where

$$g_{1s}(\alpha,r) = (2\alpha/\pi)^{3/4}\exp(-\alpha r^2)$$

$$g_{2p_x}(\alpha,r) = (128\alpha^5/\pi^3)^{1/4}\exp(-\alpha r^2)$$

with analogous expressions for the g_{2p_y} and g_{2p_z} functions. The values of the Gaussian exponents, α, and the linear expansion coefficients, d, have been determined by least-squares condition, assuming the Slater exponent, ζ, to be 1. Best fits to Slater functions of arbitrary ζ, chosen so as to be appropriate for the description of average molecular environments, may be simply expressed.

$$\varphi(\zeta,r) = \zeta^{3/2}\varphi(\zeta = 1,r)$$

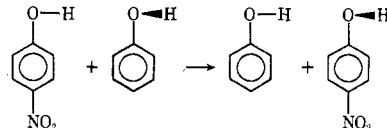Numerical values for such average ζ exponents may be found elsewhere.[5]

One area to which the STO-3G minimal basis set has been extensively applied is to the calculation of the equilibrium geometries of simple organic molecules.[7]

(3) (a) IBMOL-6, H. Popkie and E. Clementi, I. B. M. Research Laboratory, San Jose, Ca.; (b) POLYATOM 2, D. B. Newman, H. Basch, R. L. Kornegay, L. C. Snyder, J. W. Moskowitz, C. Hornback, and S. P. Liebmann, Program No. 199, Quantum Chemistry Program Exchange, Indiana University, Bloomington, Ind.; (c) GAUSSIAN 70, W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, Program No. 236, Quantum Chemistry Program Exchange, Indiana University, Bloomington, Ind.

(4) For an excellent review of basis set development previous to the work described in this Account, see: H. F. Schaefer, III, "The Electronic Structure of Atoms and Molecules", Addison Wesley, Reading, Mass., 1972, p 56.

(5) (a) W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, 51, 2657 (1969); (b) W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, *ibid.*, 52, 2769 (1970).

(6) For earlier work on Gaussian basis sets constructed to simulate exponential functions, see: (a) C. M. Reeves and R. Fletcher, *J. Chem. Phys.*, 42, 4073 (1965); (b) K. O-ohata, H. Taketa, and S. Huzinaga, *J. Phys. Soc. Japan*, 21, 2306 (1966); (c) W. J. Hehre, R. F. Stewart, and J. A. Pople, *Symp. Faraday Soc.*, No. 2, 15 (1968).

(7) Tabulations include: (a) M. D. Newton, W. A. Lathan, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, 52, 4064 (1970); (b) W. A. Lathan, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, 93, 808 (1971); (c) L. Radom, W. A. Lathan, W. J. Hehre, and J. A. Pople, *ibid.*, 93, 5339 (1971); (d) W. A. Lathan, W. J. Hehre, L. A. Curtiss, and J. A. Pople, *ibid.*, 93, 6377 (1971); (e) W. A. Lathan, L. Radom, W. J. Hehre, J. B. Lisle, and J. A. Pople, *Prog. Phys. Org. Chem.*, 11, 175 (1974); (f) W. J. Hehre and J. A. Pople, *J. Am. Chem. Soc.*, 97, 6941 (1975); (g) J. A. Pople in "Modern Theoretical Chemistry", H. F. Schaefer, III, Ed., Phenum Press, New York, N.Y., in press.

Table I
Energies of *Isodesmic* Reactions

Reaction	STO-3G ΔE^b	Exptl^a ΔH^b
(1) $CH_3CH_2CH_3 + CH_4 \rightarrow 2CH_3CH_3$	0.2	1.5
(2) $CH_3CH{=}CH_2 + CH_4 \rightarrow CH_3CH_3 + CH_2{=}CH_2$	4.1	5.0
(3) $CH_3C{\equiv}CH + CH_4 \rightarrow CH_3CH_3 + HC{\equiv}CH$	7.8	7.2
(4) $c\text{-}C_3H_6 + 3CH_4 \rightarrow 3CH_3CH_3$	−48.3	−23.5
(5) $NH_2CHO + CH_4 \rightarrow CH_3NH_2 + H_2CO$	12.5	29.8
(6) $C_6H_5NH_3{}^+ + NH_3 \rightarrow C_6H_6 + NH_4{}^+$	6.4	6.5
(7) $C_6H_5CH(Me)_2{}^+ + c\text{-}C_3H_5CH(Me)_2 \rightarrow C_6H_5CH(Me)_2 + c\text{-}C_3H_5CH(Me)_2{}^+$	2.4	−0.8
(8)	1.0	1.0

^{a} Except for reactions 6–8, experimental data have been corrected for zero-point vibrational energies. ^{b} kcal/mol.

Over a hundred detailed comparisons have now been made of complete geometrical structures both as calculated by the STO-3G method and as determined experimentally. In only a very few instances has the theory proved to be grossly in error, and in the vast majority of cases bond lengths are given to within 0.02 or 0.03 Å of experiment, and bond angles to less than 5°. An overview of the performance of STO-3G with regard to the calculation of molecular geometry is provided in Figures 1 and 2. In the first, STO-3G calculated bond lengths between heavy (nonhydrogen) atoms in a wide variety of simple molecules are plotted against experimental values. Mean deviation (31 comparisons) is 0.024 Å. It should be noted that, in addition to the reasonable overall level of agreement between theory and experiment, the calculations are quite successful in accounting for variations in bond lengths which arise due to subtle changes in molecular structure. For example, the STO-3G method succeeds in reproducing the observed progression in the lengths of carbon–carbon single bonds in the series of acyclic hydrocarbons. Thus, the

STO-3G[8]	0.0	0.025	0.029	0.061	0.082
Relative carbon–carbon bond length					
Exptl	0.0	0.028	0.052	0.070	0.092

CC linkage in propane is calculated to be 0.082 Å longer than the central bond in but-3-en-1-yne, to be compared with an experimental difference of 0.092 Å. Figure 2 shows a comparison of calculated and measured skeletal bond angles (those involving heavy atoms only) for a number of simple organic molecules. Again the mean deviation between theory and experiment is reasonably small (1.1° for 35 comparisons), and again it can be seen that only in a very few systems are the calculations seriously in error.

The STO-3G method has also met with considerable success in the calculation of the energies of those chemical reactions in which the numbers of each kind of formal chemical bond are conserved, processes in which only the immediate environment surrounding each linkage has been altered. A comparison of calculated and experimentally determined heats for a variety of such *isodesmic* reactions[9,10] is presented in Table I.

Figure 2. Experimental vs. theoretical (STO-3G) skeletal bond angles. 1, NF_3; 2, OF_2; 3, CF_2; 4, FOOF; 5, CH_2F_2; 6, CHF_3; 7, $(CH_3)_3N$; 8, $(CH_3)_2O$; 9, $i\text{-}C_4H_{10}$; 10, $(CH_3)_2NH$; 11, trans $n\text{-}C_4H_{10}$; 12, C_3H_8; 13, $CH_3N{=}O$; 14, CF_3; 15, FN${=}$NF; 16, $CH_3N{=}CH_2$; 17, O_3; 18, $\angle C_2C_1C_3$; 19, $CH_3CH{=}NH$; 20, $CH_2{=}CHCHO$, $\angle CCC$; 21, CHOCHO; 22, $(CH_3)_2C{=}O$, $\angle CCO$; 23, $(CH_3)_2C{=}CH_2$, $\angle CC{=}C$; 24, HFC${=}O$; 25, $CH_2{=}CHC{\equiv}CH$, $\angle CC{=}C$; 26, trans $CH_2{=}CHCH{=}CH_2$; 27, trans $CH_3CH{=}CHCH_3$; 28, $CH_2{=}CHCHO$, $\angle CCO$; 29, vinylcyclopropane, $\angle CC{=}C$; 30, NH_2CHO; 31, $CH_3CH{=}CH_2$; 32, CH_3CHO; 33, HCOOH; 34, F_2CO, $\angle FCO$; 35, cis $CH_3CH{=}CHCH_3$.

Processes 1 to 5 have been termed *bond separation reactions;*[10] they compare the energy of some molecule comprised of three or more heavy atoms with those of the set of simplest systems containing the same formal chemical bonds. The first provides a measure of the degree to which methyl groups, geminally substituted at carbon, interact; the next two reactions indicate the extent of interaction between a carbon–carbon single bond and simple ethylenic and acetylenic linkages, respectively. The large negative energy of reaction 4 may be taken as evidence for the strain inherent to a saturated three-membered ring. It does not accurately reproduce the experimental energy for the process, and

(8) Theoretical data for C_3 hydrocarbons from ref 7c; for C_4 hydrocarbons from ref 4f. Experimental references may be found therein.
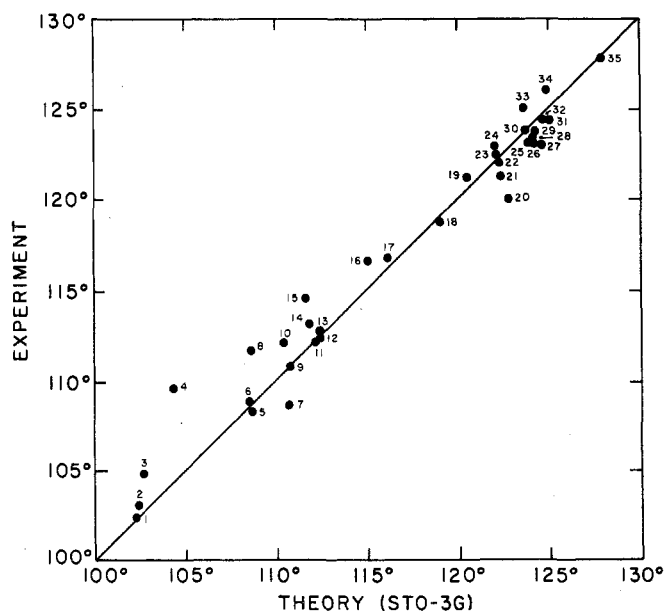(9) From the Greek, *isos* meaning equal, *desmos* meaning bond.

(10) (a) W. J. Hehre, R. D. Ditchfield, and J. A. Pople, *J. Am. Chem. Soc.*, **92**, 4796 (1970); (b) L. Radom, W. J. Hehre, and J. A. Pople, *ibid.*, **93**, 289 (1971). References to experimental data may be found within.

Reaction	STO-3G ΔE^b	Exptl^a ΔH^b
$CH_3CH_3 + H_2 \rightarrow 2CH_4$	−19.0	−18.1
$CH_3NH_2 + H_2 \rightarrow CH_4 + NH_3$	−21.0	−25.7
$CH_3OH + H_2 \rightarrow CH_4 + OH_2$	−17.4	−30.3
$CH_3F + H_2 \rightarrow CH_4 + FH$	−7.7	−29.5

^{a} Experimental data have been corrected for zero-point vibrational energies. ^{b} kcal/mol.

as we shall see in the discussion which follows, considerable basis set flexibility is necessary for adequate description. Finally, in reaction 5 we consider the energetic decomposition of formamide into its component C—N single and C=O double bonds. Here again the minimal basis set STO-3G method fares rather poorly, presumably because this system, unlike those dealt with above, is not adequately described in terms of localized single and double bonds. Rather, as is implied by classical resonance theory, the CN bond in formamide has shortened considerably, while the carbonyl linkage has elongated. Thus, this process, although formally *isodesmic*, does not in practice actually meet the bond conservation principle on which we had previously insisted. Reactions 6 and 7 are representative of another type of *isodesmic* process in which the stabilizing or destabilizing effects of a particular substituent are compared with those of some other. Thus, reaction 6 provides a measure of the effect of a phenyl substituent on the proton affinity of ammonia,[11] while process 7 indicates which of the phenyl or cyclopropyl substituents is better able to stabilize a tertiary carbonium center.[12] Finally, the *isodesmic* reaction 8 serves to indicate the relative effects of some substituent group on two conformations of the same molecule. In this case it tells us to what extent a *p*-nitro group alters the barrier to rotation in phenol.[13]

Despite the considerable success which the STO-3G minimal basis set method has achieved, particularly with regard to the calculation of molecular equilibrium geometries and of the energies of *isodesmic* reaction processes, it is a relatively easy matter to find fault with it. It is possible, however, to uncover in these failings clues essential to the development of more precise schemes. Focus on a comparison of the STO-3G calculated and experimental hydrogenation energies for the saturated 18 electron hydrides[10] (Table II). Although the theoretical energy for ethane hydrogenation is in excellent accord with the experimental value, agreement between the calculated and observed heats for methylamine, methanol, and methyl fluoride is markedly poorer. Furthermore, the error in the calculated hydrogenation energies increases with increasing electronegativity of the heteroatom bonded to the methyl.

Two reasonable explanations may be advanced to account for this behavior. First, whereas ethane, methylamine, methanol, and methyl fluoride all contain the same number of electrons, the number of atomic orbitals wich are allotted to each, within the framework

of the minimal basis set, decreases in progressing from the hydrocarbon to the halide. While each of ethane's 9 occupied molecular orbitals is constructed as a linear combination of 16 atomic basis orbitals, those describing isoelectronic methyl fluoride need to be formulated using three fewer functions. Therefore, it is entirely reasonable to expect that a minimal basis set would be better able to describe the energy of hydrogenation of a hydrocarbon than that of an alkyl halide. The obvious way to minimize this problem—other than to apportion atomic functions strictly on the basis of electron count—is to increase the number of basic functions available to all atoms. For example, we might consider doubling the number of functions used in the description of the *valence shell* of every atom. For hydrogen we would now require two atomic orbitals; for the first row atoms, boron to fluorine, nine functions would be needed.

H 1s, 1s′
B to F 1s, 2s, $2p_x$, $2p_y$, $2p_z$, 2s′, $2p_x{}'$, $2p_y{}'$, $2p_z{}'$

An alternative explanation may be given for the inability of the STO-3G minimal basis set to properly account for the hydrogenation energies of heteroatom-containing molecules. With a fair degree of precision, we might describe the electron distribution about each of the carbons in ethane as roughly *isotropic*, or spherical. This is certainly not the case for the distribution of electrons about carbon in a system such as methyl fluoride. In this instance, the charge distribution along the bond to fluorine is qite different from that perpendicular to the polar linkage. We term the electron distributions in molecules such as methyl fluoride as *anisotropic*, or of a nonspherical nature. It should be realized that the only feature in the basis representation which allows for the description of the nonspherical aspects of the molecular charge distribution is the set of p-type atomic orbitals (the s functions are, of course, only capable of handling totally spherical environments). As a minimal basis set is comprised of only a single set of p functions, and since the radial parts of all three (*x, y,* and *z*) components are constrained to be the same, we might expect that molecules in which the electron distribution is roughly *isotropic* about the individual atomic centers would be better described than those in which there is present a high degree of *anisotropy*. Thus, our insistence that the individual p_x, p_y, and p_z functions exhibit identical radial behavior forces, in the case of *anisotropic* electron distribution, a potentially severe compromise among the three cartesian directions as to what exactly that behavior should be.

There are two obvious ways in which to remedy this situation. Conceptually the simplest would be to relax the constraint that all three p functions have identical radial behavior (i.e., employ an *anisotropic* rather than an *isotropic* minimal basis set). Although such an approach would perhaps be suitable for molecules possessing a high degree of symmetry,[14] where one could clearly distinguish among the three cartesian directions, it is unsatisfactory in those situations with little or no symmetry. In such cases, the only completely unbiased way in which to choose the radial exponents of the individual p_x, p_y, and p_z components would be as the

(11) R. W. Taft in "Proton Transfer Reactions," E. F. Caldin and V. Gold, Ed., Wiley-Halstead, New York, N.Y., 1975, p 31.
(12) J. F. Wolf, P. G. Harch, R. W. Taft, and W. J. Hehre, *J. Am. Chem. Soc.*, 97, 2902 (1975).
(13) L. Radom, W. J. Hehre, J. A. Pople, C. L. Carlson, and W. G. Fateley, *J. Chem. Soc., Chem. Commun.*, 308 (1972).

(14) For applications, see: E. Switkes, R. M. Stevens, and W. N. Lipscomb, *J. Chem. Phys.*, 51, 5229 (1969).
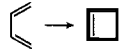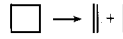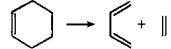
result of a costly optimization of their values with respect to the total energy of the system. A more reasonable way to surmount the difficulties inherent to an *isotropic* minimal basis set would be to to include in the atomic orbital description more than a single set of p-type functions. Thus, allowance for two sets of *isotropic* p functions in the basis representation—say, one tightly held to the nucleus, φ_{inner}, and the other relatively diffuse, φ_{outer}—permits independent adjustment of the individual radial components between the "inner" and "outer" limits.

It can be seen then that both rationalizations we have provided for the failure of the STO-3G minimal basis to adequately describe the energetics of such processes as the hydrogenation of methyl fluoride lead to the same suggestion for further development: mainly, that the most important next step to take should be the splitting of the valence shell of each heavy atom into two parts. We shall discuss only one of many possible basis sets that satisfy this criterion. In the split-valence-shell 4-31G basis,[15] the inner (1s) shell of the first row atoms, boro to fluorine, is represented as a fixed sum of 4 Gaussian-type functions, while the valence (2s,2p) shell is divided into "inner" and "outer" components represented by 3 and 1 Gaussians, respectively. Hydrogens are represented by a split 1s shell, again comprising of "inner" (3 Gaussian) and "outer" (1 Gaussian) parts. Although the Gaussian expansions used in the construction of the 4-31G basis set were chosen in order to minimize the ground-state energy of a particular atom, both "inner" and "outer" parts of the valence shell have been independently rescaled in order that they be more appropriate for the description of electron distribution in "average" molecules. These scale factors may be found elsewhere.[15]

Perhaps the most important application of the split-valence-shell 4-31G basis has been to the description of the energetics of simple chemical reactions, a number of examples of which are provided in Table III. The first five reactions are *isodesmic* bond separation processes.[10] Reactions 1 and 2 are typical of those which are well-described even at the minimal basis set STO-3G level; they provide indication of the extent to which geminally substituted methyl or fluoro groups interact. While the negative energy of reaction 3 characterizes the strain inherent to a saturated 3-membered ring, the large positive heats of bond separation for processes 4 and 5 suggest the considerable stabilization commensurate with the extensive delocalization of charge. Although these last three processes are formally *isodesmic* reactions, as commented earlier, they typify situations which are not well described using the minimal basis STO-3G method. Their energetics appear to be reasonably well handled at the split-valence-shell 4-31G level of calculations, although that for the decomposition of cyclopropane is still somewhat in error.

Reactions 6–8 establish the overall endothermicity for a number of fundamental transformations undergone by simple hydrocarbons.[16] They are not *isodesmic*

Table III
Calculated (4-31G) Reaction Energies

Reaction	4-31G ΔE [b]	Exptl[a] ΔH [b]
(1) $CH_3CH_2CH_3 + CH_4 \rightarrow 2CH_3CH_3$	1.0	1.5
(2) $CH_2F_2 + CH_4 \rightarrow 2CH_3F$	11.3	12.1
(3) $c\text{-}C_3H_6 + 3CH_4 \rightarrow 3CH_3CH_3$	−30.4	−23.5
(4) $C_6H_6 + 6CH_4 \rightarrow 3CH_3CH_3 + 3CH_2CH_2$	64.2	61.1
(5) $NH_2CHO + CH_4 \rightarrow CH_3NH_2 + H_2CO$	32.4	29.8
(6) (1,3-butadiene) → (cyclobutene)	19.9	9.7
(7) (cyclobutane) → (ethylene) + (ethylene)	14.0	13.4
(8) (cyclohexane) → (cyclobutene) + (ethylene)	34.4	32.4
(9) $CH_2O + H_2 \rightarrow CH_3OH$	−31.6	−27.0
(10) $CH_3OH + H_2 \rightarrow CH_4 + H_2O$	−31.9	−30.3
(11) $2CH_2CH_2 \rightarrow CH_3CH_3 + HCCH$	9.9	9.1

[a] Experimental data have been corrected for zero-point vibrational energies. [b] kcal/mol.

processes. We might expect, therefore, that the split-valence-shell 4-31G method would be the simplest level of theory capable of adequately describing their energetics. Reaction 6 relates the stability of 1,3-butadiene to that of its electrocyclic ring closure product, cyclobutene. Processes 7 and 8 compare the energies of the 2 + 2 and 2 + 4 cycloadducts (cyclobutane and cyclohexene, respectively) to those of their unsaturated components. Finally, reactions 9 to 11 exemplify a variety of other types of non-*isodesmic* chemical processes of some interest.[10] Reactions 9 and 10 consider the complete reduction of formaldehyde to methane and water as a two-step process. In the first a single hydrogen molecule is added to the unsaturated system to yield methanol; finally a second H_2 acts to cleave the carbon–oxygen bond. Reaction 11 measures the disproportion energy of ethylene. Although it is not an *isodesmic* process, note that the total number of bonds between carbon atoms is actully conserved. We shall see shortly that reactions of this type will be adequately described in the limit of the single-determinant Hartree–Fock formalism, while those which entail the net destruction of linkages between heavy atoms will fare more poorly. Overall, except for the transformation between 1,3-butadiene and cyclobutene, the energies of all reactions considered are reasonably well described using the 4-31G basis set.

Another area to which the split-valence-shell 4-31G basis set has been successfully applied is to the calculation of the conformations of simple organic molecules.[17] Although what is desired here is a description of a one-dimensional potential—corresponding to rotation about a single bond—some thought needs to be given to possible variations in the other geometrical coordinates of the molecule under consideration. For example, it is highly likely that the detailed geometrical structures of cis and trans planar and of orthogonal 1,3-butadiene are all different. Thus, investigation of the potential for rotation about its central carbon–

(15) (a) R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Pys.*, 54, 724 (1971); (b) W. J. Hehre and J. A. Pople, *ibid.*, 56, 4233 (1972); (c) W. J. Hehre and W. A. Lathan, *ibid.*, 56, 5255 (1972).

(16) Theoretical energy for ethylene from ref 7b; for C_4 hydrocarbons from ref 4f; for cyclohexene from R. E. Townshend, G. Ramunni, G. Segal, W. J. Hehre, and L. Salem, *J. Am. Chem. Soc.*, 98, 2190 (1976). References to experimental data may be found in the appropriate theoretical papers.

(17) Tabulations include: (a) L. Radom, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, 94, 2371 (1972); (b) L. Radom, W. A. Lathan, W. J. Hehre, and J. A. Pople, *Aust. J. Chem.*, 25, 1601 (1972); (c) L. Radom, W. A. Lathan, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, 95, 693 (1973); (d) D. Cremer, J. S. Binkley, J. A. Pople, and W. J. Hehre, *ibid.*, 96, 6900 (1974); (e) W. J. Hehre, J. A. Pople, and A. J. P. Devaquet, *ibid.*, 98, 664 (1976); (f) A. J. P. Devaquet, R. E. Townshend, and W. J. Hehre, *ibid.*, 98, 4068 (1976).

Table IV
Flexible Rotor Conformations and Energies

Molecule	Property	4-31G ΔE^a	Exptl ΔH^a
n-Butane	ΔE: trans–gauche	1.09	0.77
	Barrier: gauche → trans	2.49	2.8–3.4
	gauche → gauche	5.95	5.3–6.7
1-Butene	ΔE: skew–cis	0.90	0.2
	Barrier: skew → cis	2.17	
	skew → skew	1.97	1.74
1,3-Butadiene	ΔE: trans–gauche	3.00	1.7, >2,
	Barrier: gauche → trans	3.00	2.5
	gauche → gauche	0.50	
Acrolein	ΔE: trans–cis	0.80	2.00, 2.06
	Barrier: cis → trans	5.15	4.96
Glyoxal	ΔE: trans–cis	5.41	3.2
	Barrier: cis → trans	2.48	
Vinylcyclo-propane	ΔE: trans–gauche	1.40	1.1
	Barrier: gauche → trans	2.60	
	gauche → gauche	2.50	

^{a} kcal/mol.

Table V
Relative Energies of Hydrocarbons

Formula	Molecule	4-31G ΔE^c	6-31G*^{a} ΔE^c	Exptl^b ΔH^c
C_3H_4	Propyne	0	0	0
	Allene	0.8	1.7	2.1
	Cyclopropene	36.4	25.4	22.3
C_3H_6	Propene	0	0	0
	Cyclopropane	13.2	7.8	7.4
C_4H_6	1,3-Butadiene	0	0	0
	2-Butyne	6.9	6.8 (5.7)	8.7
	Cyclobutene	19.9	12.4 (12.3)	9.7
	Methylenecy-clopropane	25.3	20.2 (18.7)	(21.7)
	Bicyclo[1.1.0]-butane	47.2	30.4	(25.6)

^{a} Numbers in parentheses have been arrived at by assuming additivity of polarization function corrections. See text for discussion. ^{b} Except for numbers in parentheses, experimental data have been corrected for zero-point vibrational energies. ^{c} kcal/mol.

carbon linkage should make some allowance for variation in bond lengths and angles as well. In the limit, what this implies is a complete optimization of geometry for each and every point on the one-dimensional potential curve of interest, a task which, in all but the most simple of systems, is prohibitive in cost. A more reasonable approach is to assume *flexible rotation*.[18] Here one allows only for variation (as a function of conformation) in the values of the two skeletal bond angles involving the four atoms undergoing torsion. Thus, a description of torsion about the central carbon–carbon single bond in 1-butene would, for example, require optimization of both bond angles, α and β, for each choice of the dihedral angle, ω.
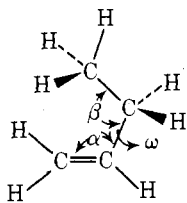
In practice, even this moderate level of geometrical optimization may well be prohibitive in cost. As alternatives, one might decide to do away with optimization altogether—that is to say, insist on *rigid rotation*—or to perform the limited geometrical variations dictated by the flexible rotor model at a level of theory less costly than 4-31G. Whereas the first of these procedures would necessarily result in the energies of cisoid conformers being badly overestimated, the latter approach, using the minimal basis set STO-3G method to account for bond angle variations during rotation, has proven itself to be of some value. A number of representative examples are presented in Table IV.[19] In all cases the 4-31G calculations are correct in their assignment of ground-state conformation. Except for acrolein, however, the theoretical model appears to underestimate the stability of the second, higher energy, rotamer, generally by 1 or 2 kcal/mol.

Although the split-valence-shell 4-31G basis represents a marked improvement over the minimal set with

regard to the calculation of relative molecular energies, situations may be found where it too performs rather poorly. For example, consider a comparison of 4-31G calculated and experimental relative isomer stabilities for each of the C_3H_4, C_3H_6, and C_4H_8 hydrocarbons (Table V).[8] Here the 4-31G basis performs admirably in comparisons involving acyclic isomers, although the relative energies of open and small-ring forms are generally in poor agreement with experiment. Furthermore, it appears that the quality of the comparisons at the 4-31G level deteriorates significantly in going to smaller and smaller ring systems. For example, the 4-31G energy of cyclopropane relative to propene is some 6 kcal/mol higher than the experimental value, while the error in the cyclobutane–isobutylene comparison is only 2 kcal/mol.

One possible explanation for the apparent deficiencies of the split-valence-shell 4-31G set in handling small-ring systems is that the electron distributions in such molecules are displaced further from the nuclear skeleton than those in their acyclic isomers. Therefore, we would expect that the restriction that the basis functions used in the construction of the system's molecular orbitals be centered exclusively on the nuclei would be more detrimental to the description of the electronic structure of small-ring molecules than to that for acyclic systems. This situation might be remedied in one of two ways. For one, we could relax the restriction of nuclear centered basis functions, that is, allow for the displacement away from the atomic positions. Although such a practice might be of use in situations where the positions of any nonnuclear centered functions might in part be determined on symmetry grounds, it is far less applicable in cases with little or no molecular symmetry.[20] In such instances, one would need to resort to a rather extensive optimization procedure in order that the resulting wavefunction be rotational invariant. A second manner exists in which to account for small displacement of the center of electronic charge away from the nuclear positions. This is simply to mix into the atomic representation basis functions of higher l quantum number (d orbitals on heavy atoms, and possibly p-type functions on hydro-

(18) L. Radom and J. A. Pople, *J. Am. Chem. Soc.*, **92**, 4786 (1970).

(19) *n*-Butane and 1-butene from ref 4f; 1,3-butadiene, acrolein, and glyoxal from ref 17f; vinylcyclopropane from W. J. Hehre, *J. Am. Chem. Soc.*, **94**, 6592 (1972). References to experimental data may be found in the appropriate theoretical papers.

(20) Such an approach has been used extensively by Kutzelnigg and co-workers in order to obtain near Hartree–Fock wavefunctions as a basis for correlation energy studies. See, for example: V. Dyczmons, V. Staemmler, and W. Kutzelnigg, *Chem. Phys. Lett.*, **5**, 361 (1970).

gens). Consider, as an example, the effect of admixing a p_x-type function into the valence s-orbital description of a hydrogen atom. The result, which a chemist would term hybridization, is a displacement of the center of

charge along the x axis and away from the hydrogen nucleus. Similar displacements of p-type atomic orbitals may be realized as the result of the addition of d-type functions to the valence description of a first row atom.
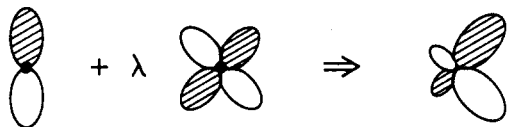

An orbital description which incorporates functions of higher l quantum numbers than are needed by the atom in its ground state is called a polarization basis. Perhaps the simplest polarization basis sets are two representations originally proposed by Hariharan and Pople.[21] The simpler of the two, termed 6-31G*, is constructed by the addition of a single set of Gaussian d functions to the split-valence-shell 6-31G basis[22] description of eac heavy atom. These added (d) functions have been uniformly rescaled so as to be suitable for use in molecules. The description given to hydrogen atoms in the 6-31G* representation is the same as that provided by the split-valence-shell 6-31G set, on which it is based. We will shortly show that, whereas the effects of polarization type functions on hydrogen atoms are by no means negligible, they may be readily accounted for in an additive manner.

The 6-31G* basis is still small enough to be readily applicable to moderately sized organic molecules (containing up four or five heavy atoms), and the results of a number of studies which have already been completed[23] are generally in good accord with experiment. In particular, calculated energies of small-ring compounds relative to those of acyclic isomers are now in reasonable agreement with experimental values, representing a marked improvement in the theory's performance over the split-valence-shell 4-31G level (Table V).

One might ask if it is possible to select for calculation at the 6-31G* level a limited set of model compounds, the energy lowerings in which could then be employed to accurately approximate the effects of polarization functions in larger molecules. For example, using the C_2 and C_3 hydrocarbons as model compounds, approximate 6-31G* energies have been obtained for the majority of the C_4 systems.[7f] Relative stabilities of the C_4H_6 hydrocarbons obtained in this manner, and given in parentheses in Table V, compare favorably with those resulting from direct calculation at the 6-31G* level. Although it is clear that the energy lowering due to the addition of d-type polarization functions on first row atoms in molecules such as the neutral hydrocarbons

may be accurately approximated using simple additivity arguments, what is not at all evident is whether such schemes would prove applicable in situations in which the molecule's electronic structure could not be properly represented in terms of a single classical valence structure. For example, it is not apparent whether the external methylene carbon in the primary cyclopropylcarbinyl cation
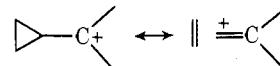
should be considered as bearing the full brunt of the molecule's positive charge, or, in the other extreme, as a perfectly normal (and uncharged) sp^2-hybridized center. The true situation is, of course, somewhere in between.[24] Another example is benzene. Without prior insight, how can we possibly be expected to know the extent to which addition of polarization functions to each of the ring carbons will lower the molecule's total energy? We are forced to concede that, while in some cases the effects of polarization functions on the heavy atom skeleton might be accurately approximated using additivity arguments, the evaluation of thei contribution to the molecule's total energy will in general be a more difficult task.

A further lowering of a molecule's total energy may be achieved by allowing for polarization functions on the hydrogen atoms as well as on the heavy-atom skeleton The 6-31G** representation is an example of a basis set of this type. It is constructed from 6-31G* by adding a single Gaussian p-type function (p_x, p_y, and p_z components) to the representation for each hydrogen atom. It should be realized that even this very simple complete polarization basis set is rather large, and its application to all but the smallest (two to three heavy atom) molecules is apt to be prohibitive in cost. It is fortunate, then, that the limited number of calculations that have been performed at such a level of theory have shown that the energy lowering resulting from the addition of polarization functions on the hydrogens may be accurately accounted for in an additive manner.[23a] Thus, it appears not to be necessary for 6-31G** calculations to be carried out in order to establish the energy lowering associated with hydrogen polarization for each and every molecule. Rather, all that seems to be required is a small number of calculations at this level on model systems. For example, a sixth of the difference between the 6-31G** and 6-31G* energies for ethane may be taken as the stability gain due to polarization for a hydrogen attached to a sp^3 carbon. Similarly, energy differences for ethylene and acetylene, for hydrazine and diazene, and for hydrogen peroxide may be used to provide corrections for hydrogens attached to carbons in other hybridization states and for the variety of NH and OH linkages. A complete listing of corrections is presented in Table VI. These values may be employed to obtain estimates of energy lowerings due to the addition of polarization functions on hydrogen for other molecules, for example, for the remaining members of the set of two heavy-atom hydrides. The mean deviation between estimated and directly calculated energy lowerings is only 0.3 kcal/mol.[25]

Even at the full polarization 6-31G** level total mo-

(21) P. C. Hariharan and J. A. Pople, *Chem. Phys. Lett.*, **16**, 217 (1972).

(22) W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).

(23) (a) P. C. Hariharan and J. A. Pople, *Theor. Chem. Acta*, **28**, 213 (1973); (b) L. Radom, W. A. Lathan, P. C. Hariharan, W. J. Hehre, and J. A. Pople, *Fortschr. Chem. Forsch.*, **40**, 1 (1973); (c) J. S. Binkley, J. A. Pople, and W. J. Hehre, *Chem. Phys. Lett.*, **36**, 1 (1975).

(24) For a discussion of the geometrical and electronic structure of the cyclopropylcarbinyl cation, see: W. J. Hehre, *Acc. Chem. Res.*, **8**, 369 (1975).

Table VI
Energy Lowerings per Hydrogen Atom due to the Addition
of p Functions to the Representation

Bond	Molecule	Energy lowering, kcal/mol
—CXY—H	C_2H_6	1.0
=CX—H	C_2H_4	1.1
≡C—H	C_2H_2	1.3
—NX—H	N_2H_2	2.0
—O—H	H_2O_2	3.9

lecular energies are considerably above their limiting values.[26] For example, energies for the first row hydrides methane, ammonia, water, and hydrogen fluoride computed using the 6-31G** basis have been estimated to lie 0.017, 0.031, 0.045, and 0.061 hartree above their respective Hartree–Fock limits. Although these remainders represent but a small fraction of a molecule's total energy (less than 0.1%) expressed in terms of their absolute values, they are by no means insignificant. Fortunately, it would appear that other molecular properties (e.g., equilibrium geometries[27] and reaction energies[23a]) have, by this time, nearly reached their limiting values.

We must, of course, at some point turn our attention to an assessment of molecular properties calculated at the Hartree–Fock limit. How closely does the "chemistry" at this limiting level of single-determinant molecular orbital theory mimic experimental observation? The response to this query is not straightforward; it depends on the particular molecular property in question. For example, the limited studies of molecular equilibrium geometry hich have been carried out near the Hartree–Fock limit of the theory indicate deviations from experiment of generally no more than 0.01 Å and 2° for bond lengths and bond angles, respectively.[28] Furthermore, the limiting energies of *isodesmic* reactions and other processes in which no net bond breaking occurs seem to be in good accord with experiment. For example, the thermochemistry of the *multiple* bond separation processes

	6-31G** ΔE, kcal/mol	Exptl ΔH, kcal/mol
HC≡CH + 4CH$_4$ → 3H$_3$C—CH$_3$	−52.7	−51.1
H$_2$C=O + CH$_4$ + H$_2$O → 2CH$_3$OH	+0.8	+3.3

is reasonably well described at the full-polarization basis, 6-31G**, level. On the other hand, the calculated energies for chemical reactions, the net result of which is bond cleavage, are generally in rather poor agreement with experimental observation. This includes not only direct bond dissociation reactions (i.e., $F_2 \rightarrow F\cdot + F\cdot$) but also such processes as those describing complete hydrogenation. For example, whereas the energy of complete hydrogenation of formaldehyde is well described by the 6-31G** calculations, that for acetylene is in rather poor accord with the experimental data:

	6-31G** kcal/mol ΔE,	Exptl kcal/mol ΔH,
$H_2C=O + 2H_2 \rightarrow CH_4 + H_2O$	−59.3	−57.3
$HC≡CH + 3H_2 \rightarrow 2CH_4$	−117.9	−105.4

Obviously, application of Hartree–Fock theory to the description of the energetics of processes of this gender—which result in the net breakage of bonds between heavy atoms—must be carried out with extreme wariness.

Conclusion

We have attempted to present our view of the sequence of steps required to progress between the simplest possible level of ab initio molecular orbital theory and the limit of the formalism. We have attempted to make clear to the reader the reasonable limits of application of the theory at each stage of its development and to indicate an overall level of confidence which he might place in its findings. Considerable progress has been made in recent years with regard to the extension of molecular orbital theory beyond the single-determinant Hartree–Fock level.[29] For the moment, calculations beyond Hartree–Fock appear to be limited to systems containing one or two heavy atoms at the most, but no doubt computations on larger molecules will be commonplace within this decade. This author strongly feels, however, that, even when such a time arrives, the single determinant Hartree–Fock method will continue to play a significant, or even dominant, role in organic chemical theory.

(25) Individual deviations (exact-approximate) are (kcal/mol); CH$_3$NH$_2$, 1.2; CH$_3$OH, 0.1; CH$_3$F, 0.2; NH$_2$OH, −0.3; NH$_2$F, 0.2; HOF, 0.2; H$_2$C=NH, 0.3; H$_2$C=O, −0.1; HN=O, −0.1; HC≡N, −0.1.

(26) For tabulations of estimated Hartree–Fock energies for simple molecules, see ref 23a, and J. A. Pople and J. S. Binkley, *Mol. Phys.*, **29**, 599 (1975).

(27) P. C. Hariharan and J. A. Pople, *Mol. Phys.*, **27**, 209 (1974).

(28) Experimental and theoretical data are summarized in ref 27.

(29) For a recent overview, see ref 1d.