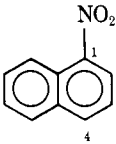
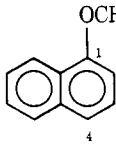
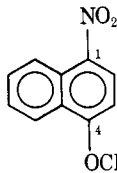


actions of 1-nitronaphthalene and 1-methoxynaphthalene is in full agreement with the substituent effect on the HO electron density. Specifically, NO<sub>2</sub> causes an increase in the coefficient of C-1 and a decrease in the coefficient of C-4, whereas methoxy operates in the opposite direction. This is illustrated by the appropriate extended Hückel HO electron densities:

			
HO electron density	C-1 (0.060) C-4 (0.036)	C-1 (0.165) C-4 (0.210)	C-1 (0.070) C-4 (0.030)

It should be pointed out that the available experimental data which was compared to the theoretical predictions has to do with *yields* rather than *rates*. The former are not necessarily related to the latter since other deactivation processes, e.g., decay of the encounter complex D...A\*, may mask the efforts of barriers and decay funnels. An analysis of the regiochemical dependence of the decay efficiency of D...A\* is complicated by the fact that the stability of a weak complex (e.g., D...A, D...A\*, etc.) depends on a number of factors.<sup>24</sup> With this in mind, the analysis of the photochemical data is offered as suggestive of a theory-experiment correlation and not as a final proof of the validity of our treatise.<sup>25</sup>

A final cautionary remark: The rules derived in this paper are based on the assumption that the lowest excited state of the aromatic involves HOMO → LUMO excitation. Depending on multiplicity, nature and number of substituents, and structure of the aromatic parent molecule, deviations may occur due to CI effects, i.e., the lowest excited state may involve subHOMO → LUMO and/or HOMO → subLUMO excitations. In such cases, the regioselection rules should be modified appropriately.

**Acknowledgment.** This work was supported by a NATO Grant (in collaboration with Professor F. Bernardi) and an A. P. Sloan Fellowship (1976–1978) to N.D.E.

## The Electronic Structure of the Low-Lying Excited States of Glyoxal

Warren B. Mueller, James F. Harrison,\* and Peter J. Wagner\*

Contribution from the Department of Chemistry, Michigan State University, East Lansing, Michigan 48824. Received May 4, 1977

**Abstract:** We have studied the electronic structure of the low-lying states of glyoxal, in particular the two lowest triplets <sup>3</sup>A<sub>u</sub> (n → π\*) and <sup>3</sup>B<sub>u</sub> (π → π\*). We find that the SCF predicted order of energies (<sup>3</sup>B<sub>u</sub> < <sup>3</sup>A<sub>u</sub>) is reversed in a configuration interaction calculation.

### Introduction

Glyoxal is the simplest of the α-dicarbonyl compounds and has been of interest to photochemists and spectroscopists for at least half a century. In this time a great deal of effort has been directed toward the identification of photoproducts and

### References and Notes

- (1) See part 1: N. D. Epiotis and S. Shaik, *J. Am. Chem. Soc.*, accompanying paper in this issue.
- (2) J. N. Murrell, S. F. Kettle, and J. M. Tedder, "Valence Theory", Wiley, New York, N.Y., 1965, p 334.
- (3) The rate of a radiationless decay process increases as the energy gap, ΔQ, becomes smaller. For discussion of this subject see (a) L. Landau, *Phys. Z. Sowjetunion*, **2**, 46 (1932); (b) C. Zener, *Proc. R. Soc. London, Ser. A*, **137**, 696 (1932); (c) G. W. Robinson and R. P. Frosch, *J. Chem. Phys.*, **37**, 1962 (1962); (d) *ibid.*, **38**, 1187 (1963); (e) J. Jortner, *Pure Appl. Chem.*, **27**, 389 (1971).
- (4) For interesting discussions of decay processes, see J. Michl, *Fortschr. Chem. Forsch.*, **46**, 1 (1974).
- (5) H. C. Longuet-Higgins and M. de V. Roberts, *Proc. R. Soc. London, Ser. A*, **224**, 336 (1954).
- (6) Data for solvent effect on phosphorescence lifetime can be found in E. J. Bowen F. R. S., "Luminescence in Chemistry", Van Nostrand, Princeton, N.J., 1968.
- (7) S. P. McGlynn, T. Azumi, and M. Kinoshita, "The Triplet State", Prentice-Hall, Englewood Cliffs, N.J., 1969, Chapter 6.
- (8) Discussion of selection rules for efficient spin-orbit coupling can be found in part 3 of this series: N. D. Epiotis and S. Shaik, *J. Am. Chem. Soc.*, preceding paper in this issue.
- (9) For an earlier treatment of thermal electrophilic aromatic substitutions see K. Fukui, T. Yonezawa, and H. Shingu, *J. Chem. Phys.*, **20**, 722 (1952).
- (10) H. C. Brown and L. M. Stock, *J. Am. Chem. Soc.*, **79**, 1421 (1957).
- (11) D. A. de Bie and E. Havinga, *Tetrahedron*, **21**, 2359 (1965).
- (12) G. Lodder and E. Havinga, *Tetrahedron*, **28**, 5583 (1972).
- (13) For earlier treatments of these reactions see (a) K. Fukui, T. Yonezawa, C. Nagata, and H. Shingu, *J. Chem. Phys.*, **22**, 1433 (1954); (b) N. D. Epiotis, *J. Am. Chem. Soc.*, **95**, 3188 (1973).
- (14) J. Cornelisse and E. Havinga, *Chem. Rev.*, **75**, 353 (1975).
- (15) J. Cornelisse, *Pure Appl. Chem.*, **41**, 433 (1975).
- (16) J. Cornelisse, G. P. De Gunst, and E. Havinga, *Adv. Phys. Org. Chem.*, **11**, 225 (1975).
- (17) E. Havinga and J. Cornelisse, *Pure Appl. Chem.*, **47**, 1 (1976).
- (18) See discussion in W. T. Borden, "Modern Molecular Orbital Theory for Organic Chemistry", Prentice-Hall, Englewood Cliffs, N.J., 1975, p 86.
- (19) R. L. Letsinger and R. R. Hautala, *Tetrahedron Lett.*, 4205 (1969).
- (20) W. C. Petersen and R. L. Letsinger, *Tetrahedron Lett.*, 2197 (1971).
- (21) C. M. Lok, J. Lugtenburg, J. Cornelisse, and E. Havinga, *Tetrahedron Lett.*, 4701 (1970).
- (22) C. M. Lok and E. Havinga, *Proc. K. Ned. Akad. Wet., Ser. B*, **77**, 15 (1974).
- (23) C. M. Lok, Thesis, University of Leiden, Leiden, 1972.
- (24) For example, see H. Umeyama and K. Morokuma, *J. Am. Chem. Soc.*, **98**, 7208 (1976), and previous papers.
- (25) Charge iterative extended Huckel calculations were performed using a program kindly provided by Professor E. R. Davidson. For the original work, see R. Hoffmann, *J. Chem. Phys.*, **39**, 1397 (1963). The CNDO/2 computations were carried out using the program described in J. A. Pople and D. L. Beveridge, "Approximate Molecular Orbital Theory", McGraw-Hill, New York, N.Y., 1970. The ab initio computations, using an STO-3G basis set, were carried out using the Gaussian 70 series of programs (W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, Quantum Chemistry Program Exchange, University of Indiana, Bloomington, Ind.) by Professor F. Bernardi of the University of Bologna.

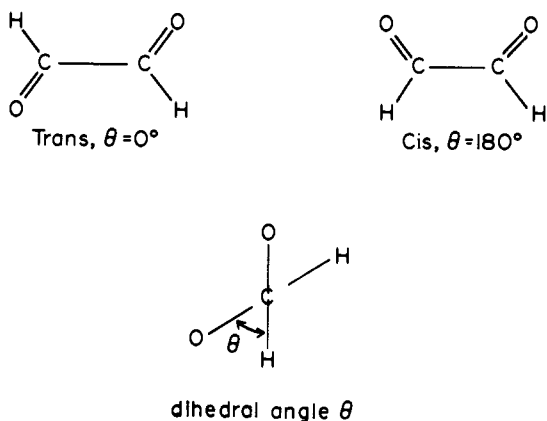


Figure 1. Torsional angle definition.

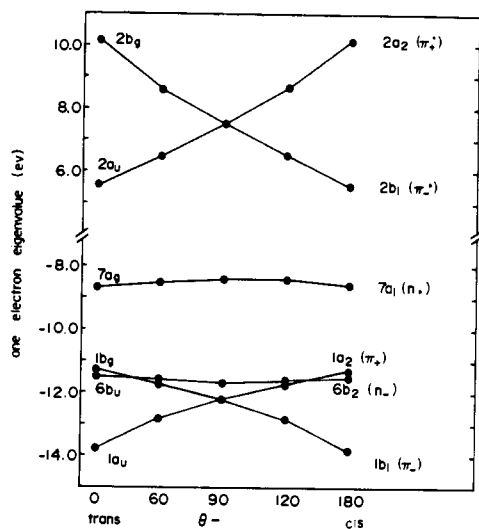


Figure 2. Hartree-Fock eigenvalues vs. torsional angle for the ground state of glyoxal.

A recent high-quality, *ab initio* calculation<sup>4</sup> places the  ${}^3A_u$  ( $n-\pi^*$ ) state 33.5 kcal above the  ${}^3B_u$  ( $\pi-\pi^*$ ) state, which, in view of the experimental evidence to the contrary, creates a serious discrepancy between theory and experiment. The purpose of this work is to reexamine the electronic structure of the low-lying excited states of glyoxal, paying particular attention to this discrepancy.

### Basis Set and Technical Details

The basis set used in all calculations reported in this study is the STO-3G set of Stewart.<sup>5</sup> All required integrals were evaluated with a local variant of the IBMOLII integral routines while the SCF solutions were obtained with standard Polyatom codes.<sup>6</sup> The configuration interaction studies were done with locally developed codes.

### Preliminary Results

In the first stage of our study we fixed the bond lengths and bond angles, with the exception of the torsional angle (defined in Figure 1), at the experimental trans values.<sup>7</sup> The torsional angle was then varied from 0° (trans) to 180° (cis). At torsional angles between 0 and 180°, the molecule does not possess the  $C_{2h}$  or  $C_{2v}$  symmetry of the trans and cis conformers, respectively. However, at any torsional angle there is a unique, angle-dependent  $C_2$  axis so that, as the molecule rotates, the molecular orbitals remain either symmetric (a) or antisymmetric (b) relative to a  $C_2$  operation about this axis. The one-electron energies of the four highest occupied and two lowest

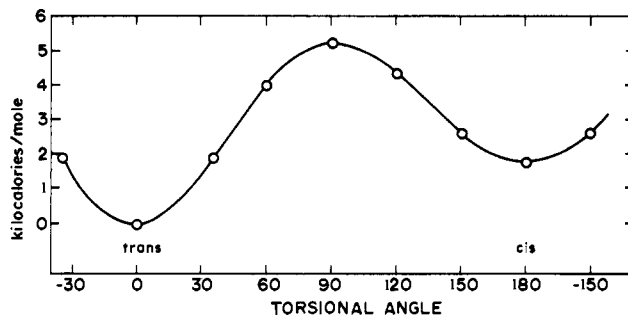


Figure 3. Self-consistent field energy vs. torsional angle for the ground state of glyoxal.

Table I. Natural Orbital Occupations for Several States of *trans*-Glyoxal at the Ground State Geometry

Natural orbitals	Occupation no.				
	${}^1A_g$	${}^1A_u$	${}^3A_u$	${}^3B_g$	${}^3B_u$
$n_+$	2.00	1.17	1.15	1.63	2.00
$\pi_+$	1.90	1.94	1.95	1.94	1.76
$\pi_-$	1.85	1.86	1.92	1.90	1.15
$n_-$	2.00	1.83	1.85	1.37	2.00
$\pi_+^*$	0.16	1.00	0.96	0.77	0.86
$\pi_-^*$	0.10	0.17	0.18	0.39	0.24

unoccupied orbitals of the ground state are shown as a function of twist in Figure 2. The variation of the total SCF energy of the ground state with twist angle is shown in Figure 3.

Our SCF calculation predicts the trans conformer to be more stable than the cis by 1.75 kcal/mol and a rotational barrier of 5.2 kcal/mol. While the rotational barrier in glyoxal has not been determined experimentally, the two closely related molecules acrolein and butadiene have experimental barriers of 5.0 and 4.9 kcal/mol, respectively. Our theoretical barrier is in reasonable agreement with most previous *ab initio* calculations.<sup>4,8</sup>

In order to improve our representation of the ground state and survey the relative positions of the excited states we constructed a modest configuration interaction function for the ground and first few excited states using the ground state molecular orbitals.

In detail we consider all singlets and triplets that arise from all possible excitations from the highest four occupied orbitals to the lowest two empty orbitals. These are the same orbitals whose energy is displayed in Figure 2. The resulting energy vs. twist curves are displayed in Figure 4. We emphasize that these curves represent the various states constrained at the ground state geometry and the energy differences relative to the ground state represent vertical transition energies rather than the minimum energy between the two states.

We can extract an orbital picture from these CI calculations by finding the natural orbitals associated with the CI function. In Table I we collect for each state in the trans configuration the occupation numbers for those natural orbitals<sup>9</sup> which are different from the molecular orbitals. The single determinant picture of the  ${}^1A_g$  state of glyoxal corresponding to the configuration  $n_+^2\pi_+^2\pi_-^2n_-^2$  is then replaced by

$$n_+^{2.00}\pi_+^{1.90}\pi_-^{1.85}n_-^{2.00}\pi_+^*^{0.16}\pi_-^*^{0.10}$$

while the  ${}^1A_u$  ( $n_+ \rightarrow \pi_+^*$ ) configuration  $n_+^1\pi_+^2\pi_-^2n_-^2\pi_+^*^1$  becomes

$$n_+^{1.17}\pi_+^{1.94}\pi_-^{1.86}n_-^{1.83}\pi_+^*^{1.00}\pi_-^*^{0.17}$$

Two previous CI studies<sup>10,11</sup> have been carried out on glyoxal using the semiempirical CNDO/S-CI<sup>12</sup> technique and are in better agreement with the experimental excitation

**Table II.** Bond Lengths and Energies for the  $^3A_u$  and  $^3B_u$  States of Glyoxal<sup>a</sup>

State	Present study						Dykstra and Schaefer		
	SCF			CI			$R_{C-C}$ , Å	$R_{C-O}$ , Å	Energy, au
$^3A_u$	1.490	1.252	-223.4666	1.439	1.341	-223.5735	1.492	1.241	-226.3959
$^3B_u$	1.342	1.421	-223.5168	1.373	1.408	-223.5661	1.353	1.356	-226.4504

<sup>a</sup>  $R_{C-C} = 1.525$ ,  $R_{C-O} = 1.206$  Å in the ground state, ref 7.

**Table III.** Largest Components of the CI Wave Function

State	Core	Configuration						Weight in CI
		$1a_u(\pi_+)$	$6b_u(n_-)$	$1b_g(\pi_-)$	$7a_g(n_+)$	$2a_u(\pi_+^*)$	$2b_g(\pi_-^*)$	
$^3B_u$	22	2	2	1	2	1	0	0.9473
	22	0	2	1	2	1	2	0.2070
	22	1	2	2	2	0	1	0.1739
	22	1	2	0	2	2	1	0.1105
$^3A_u$	22	2	2	2	1	1	0	0.8135
	22	2	1	1	2	2	0	0.3967
	22	2	1	2	2	0	1	-0.2746
	22	1	2	2	1	2	0	-0.1312
	22	1	2	1	1	2	1	0.1205
	22	1	2	2	2	0	2	-0.1085
	22	0	2	2	1	1	2	-0.1005

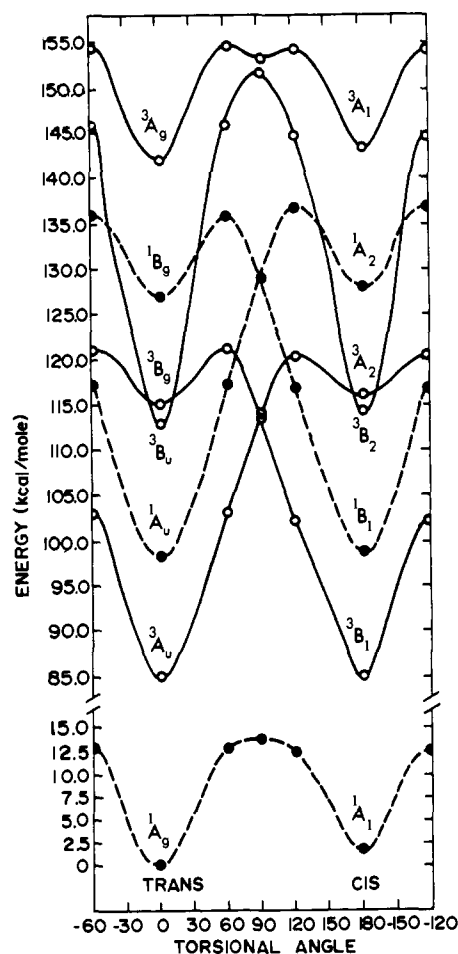
energies than are our ab initio results. Although a natural orbital analysis of these semiempirical CI wave functions was not reported, McGlynn et al.<sup>10</sup> do note the extensive configuration mixing in the various excited states and report that the *cis*-glyoxal  $^1B_2$  state is a mixture of  $\sim 80\%$   $n_+\pi_+^*$  and  $\sim 20\%$   $n_-\pi_-^*$ . This state correlates with our  $^1A_u$  state in the *trans* geometry; and, although a direct comparison is not possible, our natural orbital analysis does substantiate the presence of significant  $n_-\pi_-^*$  character. We believe that the success of the semiempirical studies is a result of two factors: the parameters used to construct the molecular orbitals and the inclusion of CI. We suspect that any reasonably parametrized one-electron theory, when coupled to a rational configuration interaction scheme, would yield a correct qualitative picture.

### The Two Lowest Triplet States

In order to improve our representation of the  $^3A_u$  ( $n_-\pi_-^*$ ) and  $^3B_u$  ( $\pi_-\pi_-^*$ ) states we first carried out restricted open-shell SCF calculations<sup>13</sup> for the *trans* configuration, optimizing the C-C and C-O bond lengths, maintaining the H-C-O angle at  $126.9^\circ$ <sup>4</sup> and the C-H bond length at  $1.080$  Å. We then constructed a configuration interaction function for both states using the open-shell orbitals appropriate to that state and configurations obtained from all excitations among the four highest occupied and the two lowest unoccupied molecular orbitals. These results are summarized in Table II where they are compared with Dykstra and Schaefer's results.<sup>4</sup>

Our SCF study predicts a  $^3A_u$ - $^3B_u$  separation of 31 kcal/mol with the  $^3B_u$  being the lowest state. This finding is in good agreement with the Dykstra and Schaefer study which predicted the  $^3B_u$  to be the lower triplet by 33.5 kcal/mol. However, our subsequent CI study of these two states inverts this order and predicts the  $^3A_u$  to lie 4.6 kcal/mol below the  $^3B_u$  state!

These results mean that while the CI calculation has improved the representation of both the  $^3B_u$  and  $^3A_u$  states it has improved the  $^3A_u$  more. This is seen by examining the CI wave function for the two states, reproduced in Table III. The large coefficients for configurations other than the SCF determinant attest to the inadequacy of the single determinant SCF solution for the  $^3A_u$  state. Perhaps more interpretable are the natural orbital occupation numbers assembled in Table IV,<sup>9</sup> and the



**Figure 4.** Energy vs. torsional angle for various low-lying states of glyoxal. Ground state bond lengths assumed.

number of electrons associated with each atom in the states of interest, shown in Table V.

We note in particular the significant depletion of  $\pi$  as well as  $n_-$  orbital occupancy in the lowest  $^3A_u$  state, which is pure

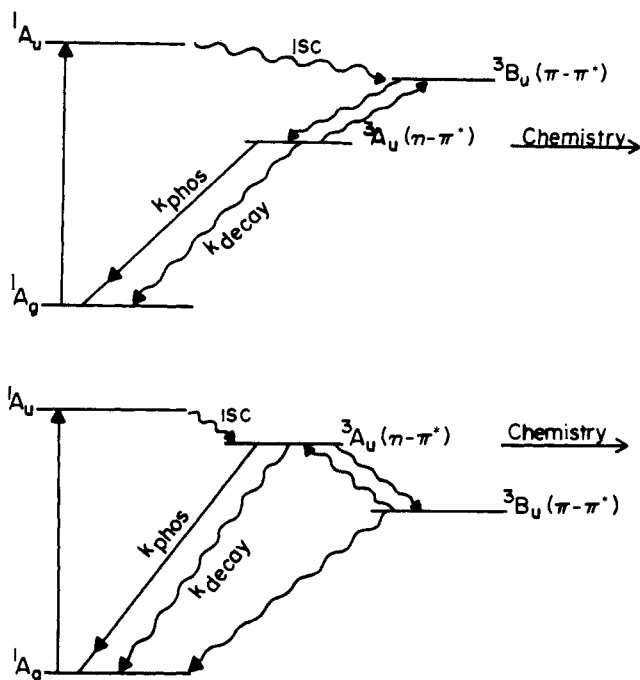


Figure 5. Two possible energy level sequences consistent with the available experimental data.

$n_+$ ,  $\pi_+^*$  in the single determinant approximation.

Both SCF and CI calculations (Table II) predict comparable bond lengths in the  $^3A_u$  state but do not agree so well on the C-O bond length in the  $^3B_u$  state. It is worth noting that inclusion of configuration interaction seems to significantly shorten the C-C bond and lengthen the C-O bond in the  $^3A_u$  state, whereas it slightly lengthens the C-C bond in the  $^3B_u$  state. The latter effect reflects loss of electron density from the totally  $\pi$ -bonding  $\pi_+$  orbital to the totally  $\pi$ -antibonding  $\pi_-^*$  orbital. Interestingly, the CI-induced shortening of the  $^3A_u$  C-C bond reflects primarily loss of electron density from  $n_-$  and  $\pi_-$ , which are respectively  $\sigma$  and  $\pi$  C-C antibonding.<sup>14</sup> The C-O lengthening in the  $^3A_u$  state appears to be due almost exclusively to the loss of electron density from the C-O  $\pi$ -bonding  $\pi_+$  and  $\pi_-$  orbitals.

## Conclusions

The uncertainties<sup>15</sup> inherent in using an STO-3G basis and our incomplete geometry optimization preclude our claiming that the  $^3A_u$  is definitely lower than the  $^3B_u$  state. We can say that it is very probable that the two states are separated by no more than a few kilocalories/mole and that the very large separation predicted by Dykstra and Schaefer<sup>4</sup> is an artifact of the single determinant SCF approximation.

Interestingly, experimental facts are not incompatible with either triplet being lower. The two possible situations are represented in Figure 5. First, if the  $^3A_u$  ( $n \rightarrow \pi^*$ ) state is lower, initial excitation into the  $^1A_u$  state followed by intersystem crossing into the  $^3A_u$  and  $^3B_u$  states with radiationless transitions from  $^3B_u$  to  $^3A_u$  will lead to most of the excitation localized in the  $^3A_u$  state from which chemistry and phosphorescence can occur.

On the other hand, if the  $^3B_u$  is only a few kilocalories/mole below the  $^3A_u$  state, thermal equilibration will significantly populate the  $^3A_u$  state. Chemistry and phosphorescence can then occur in a normal fashion. This phenomenon is well known for phenyl ketones<sup>16</sup> and several  $\alpha,\beta$ -unsaturated steroids,<sup>17</sup> and has been postulated for acrolein.<sup>18</sup> There have in fact been two independent spectroscopic observations of a second triplet just a few kilocalories above the lowest triplet of biacetyl, but the nature of this second triplet was not determined.<sup>19</sup>

Table IV. Natural Orbital Occupation Numbers of the  $^1A_g$ ,  $^3A_u$ , and  $^3B_u$  States of Glyoxal

Natural orbital	Occupation no.		
	$^1A_g$	$^3A_u$ ( $n_{\pm} \rightarrow \pi_{\pm}^*$ )	$^3B_u$ ( $\pi_{\pm} \rightarrow \pi_{\pm}^*$ )
$\pi_+$	1.90	1.90	1.85
$n_-$	2.00	1.73	2.00
$\pi_-$	1.85	1.78	1.03
$n_+$	2.00	1.27	2.00
$\pi_+^*$	0.15	1.12	0.97
$\pi_-^*$	0.10	0.20	0.15

Table V. Total Number of Electrons on Various Atoms as Defined by the Mulliken Population Analysis for the Trans Conformations of the  $^1A_g$ ,  $^3A_u$ , and  $^3B_u$  States of Glyoxal

State	C	H	O
$^1A_g$	5.8738	0.9425	8.1837
$^3A_u$	6.0237	0.9145	8.0618
$^3B_u$	5.9766	0.9336	8.0898

It should be pointed out that our calculations suggest comparable lowest  $\pi, \pi^*$  triplet energies for butadiene and glyoxal, a conclusion which certainly is not evident from a simple extrapolation from butadiene to acrolein.<sup>17,18</sup>

In summary, while our results are consistent with the  $^3A_u$  state being the lowest excited state in glyoxal, they also suggest that the first two triplets may be close enough in energy to equilibrate thermally. Although thermal equilibrium between excited states is not novel, an examination of the photochemistry and spectroscopy of glyoxal and other  $\alpha$ -dicarbonyl compounds from this vantage point may be productive.

It is noteworthy that no matter what basis set is used to construct the molecular orbitals for glyoxal<sup>4,10,11,20</sup> the SCF model is deficient in predicting the relative energies of the  $^3A_u$  and  $^3B_u$  states. The proper description of these states requires CI.

**Acknowledgment.** This work was supported in part by NSF Grants MPS 70-01653 and CHE76-11892. W.B.M. thanks the Chemistry Department for a L. L. Quill Fellowship. We also thank Professor H. F. Schaefer for preprints of related work.

## References and Notes

- (1) F. E. Blacet and R. W. Moulton, *J. Am. Chem. Soc.*, **63**, 868 (1941); J. G. Calvert and G. S. Layne, *ibid.*, **75**, 856 (1953); G. Herzberg and D. A. Ramsay, *Proc. R. Soc. London, Ser. A*, **233**, 34 (1955).
- (2) C. S. Parmenter, *J. Chem. Phys.*, **41**, 658 (1964).
- (3) M. Kasha, *Discuss. Faraday Soc.*, **9**, 14 (1950).
- (4) C. Dykstra and H. F. Schaefer, *J. Am. Chem. Soc.*, **97**, 7210 (1975); **98**, 401 (1976).
- (5) R. F. Stewart, *J. Chem. Phys.*, **52**, 431 (1967).
- (6) D. R. Davis and E. Clementi, "IBMOL", Program 92, Quantum Chemistry Program Exchange, Indiana University; D. B. Neumann, H. Basch, R. L. Kornegoy, L. C. Snyder, J. W. Moskowitz, C. Hornbach, and S. P. Liebmann, "Polyatom (Version 2)", Program 199, Quantum Chemistry Program Exchange, Indiana University.
- (7) M. Kuchitsu in "Molecular Structure and Properties", G. Allen, Ed., University Park Press, Baltimore, Md., 1972.
- (8) H. Kato, H. Konishi, H. Yamabe, and T. Yonezawa, *Bull. Chem. Soc. Jpn.*, **40**, 4761 (1967); V. Pincelli, B. Cadioli, and D. J. David, *J. Mol. Struct.*, **9**, 173 (1971); Tae-Kyu Ha, *ibid.*, **12**, 171 (1972); K. Sundberg and L. Cheung, *Chem. Phys. Lett.*, **29**, 93 (1974); M. Hackmeyer, Ph.D. Thesis, Michigan State University, 1970.
- (9) (a) If the wave function for a many-electron system is constructed as a linear combination of determinants, the electron density at a point  $R$  in space is given by

$$\rho(\vec{R}) = - \sum_{u=1}^M \psi_u^2(\vec{R}) n_u$$

where  $\psi_u$  is a natural orbital and  $n_u$ , the occupation number of the  $u$ th natural orbital, can assume any value between 0 and 2. These occupation numbers sum to the total number of electrons,  $N$ , in the system, i.e.,

$$\sum_{u=1}^M n_u = N$$

In the special case where the wave function is a single determinant the

- natural orbitals are the molecular orbitals and the occupation numbers are 0, 1, or 2. An index of the adequacy of a molecular orbital description is the closeness of the  $n_i$ 's to integer values. (b) See, for example, H. F. Schaefer, III, "The Electronic Structure of Atoms and Molecules", Addison-Wesley, Reading, Mass., 1972.
- (10) J. F. Arnett, G. Newkome, W. L. Mattice, and S. P. McGlynn, *J. Am. Chem. Soc.*, **96**, 4385 (1974).
- (11) J. M. Leclercq, C. Mijourle, and P. Yvan, *J. Chem. Phys.*, **64**, 1464 (1976).
- (12) J. Del Bene and H. H. Jaffé, *J. Chem. Phys.*, **48**, 1807 (1968).
- (13) C. C. J. Roothaan, *Rev. Mod. Phys.*, **32**, 179 (1960).
- (14) J. R. Swenson and R. Hoffman, *Helv. Chim. Acta*, **53**, 2331 (1970).
- (15) J. A. Pople, *Acc. Chem. Res.*, **3**, 217 (1970).
- (16) (a) P. J. Wagner, M. J. May, A. Haug, and D. R. Graber, *J. Am. Chem. Soc.*, **92**, 5269 (1970); (b) P. J. Wagner, A. E. Kemppainen, and H. N. Schott, *ibid.*, **95**, 5604 (1973); (c) E. Migirdicyan, *Chem. Phys. Lett.*, **12**, 473 (1972); (d) M. E. Long and E. C. Lim, *ibid.*, **20**, 413 (1973); (e) R. O. Loutfy and J. M. Morris, *ibid.*, **22**, 584 (1973).
- (17) G. Marsh, D. R. Kearns, and K. Schaffner, *Helv. Chim. Acta*, **51**, 1890 (1968).
- (18) A. Devaquet, *J. Am. Chem. Soc.*, **94**, 5160 (1972).
- (19) (a) E. Drent, R. P. vander Werf, and J. Kommandeur, *J. Chem. Phys.*, **59**, 2061 (1973); (b) K. Kaya, W. R. Harshberger, and M. B. Robin, *ibid.*, **60**, 4231 (1974).
- (20) C. E. Dykstra, R. R. Lucchese, and H. F. Schaefer III, *J. Chem. Phys.*, **67**, 2422 (1977).

## Constitutional Symmetry and Unique Descriptors of Molecules

Wolfgang Schubert and Ivar Ugi\*

Contribution from the Organisch-Chemisches Institut, Technische Universität München, 8000 München 2, West Germany. Received January 3, 1977

**Abstract:** In systems for computer-aided synthesis planning the number of redundant synthetic intermediates generated can be greatly reduced by detecting the constitutional symmetry of molecules to be processed. The chemical shift pattern of NMR spectra yields information about constitutional symmetry with respect to the observed nuclei, and constitutional symmetry is important for the interpretation of NMR spectra. An algorithm has been designed which detects constitutional symmetry by finding the atoms which are similar with respect to automorphisms of the labeled graph underlying the structural formula of a compound. The information about the constitutional equivalence of atoms is then used to canonically order the atoms of the molecule. A canonical order is necessary for treating the stereochemical features of molecules. Furthermore, the canonical order of atoms directly leads to a unique representation of molecules. This unique representation is necessary to identify molecules when generating and evaluating the reaction paths in systems for automated synthetic design. Since identifying molecules is based upon the results of the algorithm for finding constitutional symmetry the overall efficiency of such systems can be substantially increased.

### Introduction

The chemical constitution of a molecule, or an ensemble of molecules, is determined by the number and kind of atoms which it contains and those pairs of neighboring atoms which are connected by covalent bonds.

A chemical constitution is usually described by a constitutional formula. It consists of atomic symbols which are connected by lines. The atomic symbols represent atomic cores. They consist of the atomic nuclei and the inner electrons. The lines in constitutional formulas refer to covalent bonds which correspond to valence electrons in orbitals belonging to two or more cores. Furthermore, a constitutional formula may contain statements about free valence electrons at some of the cores.

In the computer-assisted solution of chemical problems and chemical documentation the chemical constitution of molecular systems is represented by connectivity lists or matrices.<sup>1</sup> Such computer-oriented representations of chemical constitutions are only unequivocal if they are based upon a canonical order of the atoms. In order to avoid ambiguities in the case of molecules which are representable by two or more resonance formulas, the same representation should result for a given molecule, regardless of the resonance structures considered.

A molecule may contain constitutionally equivalent atoms. This fact is reflected in NMR spectra, if the fine structure of the bands is neglected which is due to spin-spin coupling and stereochemical effects. An NMR spectrum then corresponds to the classes of constitutionally equivalent atoms whose nuclei are observed. As a rule, the relative areas of the peaks in the chemical shift pattern of <sup>1</sup>H NMR spectra correspond to the

number of protons in the corresponding class of constitutionally equivalent atoms.

A molecule which contains constitutionally equivalent atoms is called constitutionally symmetric. Perception of constitutional symmetry is also important in the solution of chemical problems with the aid of computers. In computer-assisted synthetic design<sup>2</sup> neglect of constitutional symmetry leads to redundancies in the tree of synthetic pathways. Furthermore, a canonical order of the atoms in a molecule is readily generated if constitutional symmetry is adequately taken into account. A canonical order of the atoms in molecules is necessary for computer-oriented descriptions of molecules.

Several algorithms for generating unique representations of molecules have been reported so far,<sup>3</sup> but they were designed for the use in retrieval systems and do not directly produce any additional information which could possibly be used in systems for synthesis planning. The sole purpose of these algorithms is to generate somehow a unique description for a molecule. For another approach to this problem see ref 4. In this paper an algorithm is presented which detects the constitutional symmetry and generates a unique description of the molecule.

### Constitutional Symmetry

Usually the symmetry of molecules is discussed in terms of point groups. Since the symmetry of three-dimensional objects is described, the point groups are based on the group of all orthogonal transformations in three-dimensional space, like rotations, reflections, and translations. A symmetry operation representing a certain symmetry property of a molecule sends