

Localization of the filled and virtual orbitals in the nucleotide bases

J. Cizek*, W. Förner, and J. Ladik

Chair for Theoretical Chemistry, Friedrich-Alexander University Erlangen-Nürnberg,
D-8520 Erlangen, Federal Republic of Germany

For the four nucleotide bases cytosine, uracil, adenine and guanine both Boys (B) and Edminston–Ruedenberg (ER) localization procedures of the *ab initio* canonical orbitals have been performed. The results obtained for both σ - π separation and by treating all electrons together show a very good localization for all electrons (one-center lone-pairs and two-center localized orbitals even for π -electrons) and a rather good localization for the virtuals applying both B and ER criteria. The results of the two methods are essentially identical. These results suggest that the application of localized orbitals will open new possibilities for the calculation of correlation in extended systems.

Key words: Localization — Orbitals of nucleotide bases.

1. Introduction

Despite all the large progress achieved in the last fifteen years the calculation of correlation energy in extended systems [1] (molecules with more than 50 electrons, polymers and solids) presents a formidable problem. If we have an extended system the number of interconfiguration matrix elements (even if we restrict ourselves to singly and doubly excited configurations in CI calculations) increases very rapidly with the number of basis functions m . At the same time, because of the size consistency problem [2] with the increasing number of electrons n , the inclusion of only singly and doubly excited configurations provides smaller and smaller parts of the correlation energy. For these two reasons even the

* *Permanent address:* Department of Applied Mathematics, Department of Chemistry, University of Waterloo, Guelph-Waterloo Center for Graduate Work in Chemistry, University of Guelph and University of Waterloo.

application of the most advanced techniques for the generation (like the unitary group approach [3] and storage (like the direct CI method [4]) of the inter-configuration matrix elements there are foreseeable limits to the applicability of the CI and related methods.

Therefore one has to look for alternative methods which are size consistent and provide the bulk part of the correlation energy with a manageable computer time even for extended systems. As is well-known, perturbation theory [5], electron pair theories [6] and, as next step coupled cluster theory [7] are size consistent. On the other hand these methods also require an unacceptable large number of matrix elements if the number of electrons is large and one starts with delocalized canonical HF orbitals. The only prespective way to handle this problem seems to be to start correlation calculations with localized orbitals (LO-s) which are obtained from a unitary transformation of the original canonical orbitals. In this way we can subdivide a large molecule into a number of smaller regions. Using these LO-s in conjugation with one of the above mentioned size consistent methods and taking into account as a first approximation only excitations between filled and virtual orbitals localized into the same region of space, one can most probably reduce by several orders of magnitude the number of necessary matrix elements. Of course, the reduction achieved depends extremely sensitively on the extent of the localization.

For the localization of the filled orbitals in a molecule of low symmetry there are two widely used techniques, the method of Boys [8] and that of Edminton and Ruedenberg (ER) [9]. The construction of suitably localized orbitals is much more difficult for virtual orbitals, and especially for a molecule without a high degree of symmetry it is very much open to discussion. One would like to have virtuals which are localized as much as possible in the same regions as the filled orbitals. For planar molecules this means that the orbitals would be localized both horizontally (in the plane of the molecule) and vertically. To achieve this, one can follow the suggestion of Boys [10] and multiply the filled localized orbitals by powers of x , y and z (his so-called oscillatory orbitals). The virtual orbitals obtained in this way are not orthogonal to each other, but they can be orthogonalized without loosing the localization with the help of a newly developed procedure [11].

There are different localization studies in the literature. England et al. [12] investigated the localization of π orbitals in aromatic hydrocarbons using the Edminton–Ruedenberg procedure. In the case of larger aromatic hydrocarbons they have found that the π orbitals could be transformed mostly to 3- and 4-center LO-s but not to 2-center ones. Lipscomb et al. [13] have compared the B and ER localization procedure in the case of simple heteroatomic molecules (CO_2 and CO_3^{2-}) applying once $\sigma - \pi$ separation and once without it. We are not aware, however, of localization studies on any larger heterocyclic π electron molecule.

In a recent paper [14] also the application of the coupled cluster theory [7] with Del Re's localized orbitals to CH_4 (taking into account only singly and doubly

excited states) has been reported. The results obtained for the correlation energy seem to be rather promising for this highly symmetric molecule.

In this paper, we have investigated the localization of the filled and unfilled $HF\sigma$ - and π -orbitals of the four planar nucleotide bases, cytosine (C), uracil (U), adenine (A) and guanine (G) using both the Boys and Edminston–Ruedenberg procedure. In a subsequent paper we shall describe the construction of localized “oscillatory” virtual orbitals for the same molecules which are appropriate for these studies because besides being planar, they do not possess any other symmetry. After this we plan to perform correlation calculations for these molecules using localized orbitals as a basis for any type of size consistent method.

To conclude this introduction let us mention that these localized orbitals are useful not only for the calculation of correlation energy but they also facilitate the interpretation of the correlated wave function in chemical terms. In this way the results obtained should be much closer to the usual chemical description of molecules than those based on delocalized orbitals.

2. Localization of the Hartree–Fock orbitals of the nucleotide bases

For the localization of the canonical HF orbitals of the closed shell nucleotide bases both the B [8] and ER [9] procedure has been applied. Both methods in their original formulation refer to closed-shell systems. In the B method [7] the expression

$$I_1 = \sum_{i=1}^{n^*} \langle \varphi_i \varphi_i | r_{12}^2 | \varphi_i \varphi_i \rangle \quad (1)$$

has to be minimized (in (1) the summation has to be extended over all the n^* doubly occupied orbitals φ_i). In the method of ER [9] the sum of Coulomb integrals (see Eq. (1)) has to be maximized.

$$I_2 = \sum_{i=1}^{n^*} \langle \varphi_i \varphi_i | \frac{1}{r_{12}} | \varphi_i \varphi_i \rangle. \quad (2)$$

The criterion of ER has the conceptual advantage that I_2 is part of the total energy and that it is applicable to every case. In the criterion (1) (though it cannot be applied to the Be atom [15]) due to the fact that r_{12} occurs on the second power, the integration according to the coordinates of electrons one and two can be separated and therefore the calculation is much faster.

The *ab initio* calculation of the canonical orbitals has been performed by means of the minimal STO-3G basis. For the calculation of the nucleotide bases the geometry established on the basis of X-ray diffraction of B-DNA [16] has been used. In the few cases when some atoms (first of all the atoms of the amino groups) were not completely in the plane they have been projected into the plane of the rings. This procedure is of course not exact, but since the largest deviation from the plane was 0.5 a.u., the error in the bond distances caused in this way

(maximum 0.1 Å) certainly does not significantly influence the results of the localization procedure.

The ratio of the computer time necessary for the ER and B procedures increased rapidly with the number of orbitals (for the 5π orbitals of C this ratio was 9, for the 24 non- π orbitals of U it was 120 and for all the 29 orbitals of C it was 212) in accordance with the findings of Kleier et al. [17] who found this ratio to be 50 for the 24 orbitals of B_4H_4 .

In Tables 1 and 2 we present the orbital coefficients and populations, respectively, obtained from the localization of the π -orbitals of C using both the B (for occupied and virtual orbitals) and ER procedures (only for filled π -orbitals). Fig. 1a shows the numbering of the C molecule used to construct the Tables.

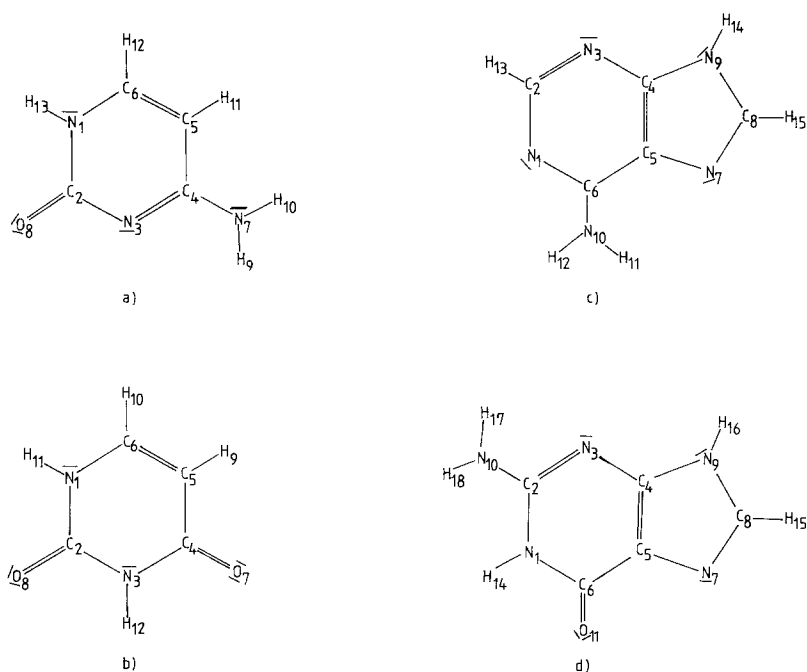
Tables 1 and 2 show that the localization in the ground state of the electrons is very pronounced in a way which very nearly resembles the classical expression given in Fig. 1a. One obtains a localized orbital for all the three classical double bonds polarized because of the different electronegativities of the constituent atoms and the additional two localized π -orbitals basically correspond to the lone pairs of nitrogen atoms 1 and 7, respectively. Further comparing the results of the B and ER procedures one finds that they are almost identical. Actually if one uses localized B orbitals instead of ER orbitals for the calculation of expression (2), one only finds the very small difference of 3.3×10^{-3} a.u. between this value and the value calculated by the ER orbitals. This means that criterion (2) is nearly perfectly fulfilled also with B orbitals. On the other hand if HF

Table 1. Coefficient of the π -atomic orbitals in the 5 filled and 3 virtual B and ER, respectively, localized π -orbitals of cytosine. The first number refers to B, the second one to the ER-orbital (the latter one only for occupied orbitals)

Atom	LO1	LO2	LO3	LO4	LO5	LO6	LO7	LO8
N ₁	-0.889	-0.035	-0.122	-0.093	-0.024	0.359	-0.355	0.094
	-0.893	-0.034	-0.113	-0.068	-0.017			
C ₂	-0.176	0.033	0.532	0.090	0.196	-0.020	0.872	0.052
	-0.172	0.038	0.530	0.084	0.208			
N ₃	-0.010	0.113	-0.058	0.082	0.736	-0.121	-0.356	-0.611
	-0.006	0.136	-0.072	0.079	0.731			
C ₄	0.075	-0.257	-0.085	-0.156	0.466	-0.106	-0.052	0.897
	0.082	-0.235	-0.096	-0.160	0.470			
C ₅	0.125	0.035	0.052	-0.672	0.029	0.743	0.078	-0.290
	0.144	0.045	0.049	-0.668	0.023			
C ₆	-0.200	0.035	-0.008	-0.547	-0.132	-0.860	0.050	-0.067
	-0.186	0.039	-0.005	-0.551	-0.136			
N ₇	-0.032	-0.915	0.035	0.073	-0.152	-0.015	0.058	-0.423
	-0.037	-0.920	0.036	0.060	-0.120			
O ₈	0.121	-0.032	0.756	-0.059	-0.128	-0.020	-0.652	0.076
	0.128	-0.038	0.758	-0.057	-0.128			

Table 2. Mulliken populations of the localized occupied and virtual π -orbitals in cytosine: (The first number is obtained with the aid of the B, the second one in the case of filled orbitals with the aid of the ER procedure)

Atom	LO1	LO2	LO3	LO4	LO5	LO6	LO7	LO8
N ₁	1.70	0.00	0.00	0.00	0.00	0.15	0.13	0.02
	1.71	0.00	0.00	0.00	0.00			
C ₂	0.12	0.00	0.69	0.02	0.12	0.00	1.06	0.00
	0.11	0.00	0.68	0.01	0.14			
N ₃	0.00	0.01	0.00	0.01	1.26	0.03	0.15	0.54
	0.00	0.02	0.00	0.01	1.25			
C ₄	0.01	0.21	0.01	0.09	0.54	0.00	0.00	1.13
	0.01	0.19	0.02	0.09	0.56			
C ₅	0.02	0.00	0.00	1.12	0.00	0.76	0.01	0.08
	0.03	0.00	0.00	1.11	0.00			
C ₆	0.14	0.00	0.00	0.76	0.03	1.06	0.00	0.01
	0.12	0.00	0.00	0.78	0.03			
N ₇	0.00	1.77	0.00	0.00	0.01	0.00	0.00	0.21
	0.00	1.78	0.00	0.00	0.00			
O ₈	0.02	0.00	1.30	0.00	0.03	0.00	0.64	0.01
	0.02	0.00	1.30	0.00	0.02			

**Fig. 1a-d.** Numbering for the molecules cytosine (a), uracil (b), adenine (c) and guanine (d)

orbitals are used to calculate expression (2), one obtains a value smaller by 1.2 a.u. One should also point out that due to the role of the hetero atoms in C the localization of the π -orbitals is much stronger than in aromatic hydrocarbons [12]. Looking finally at the three B localized virtual π -orbitals of C one finds that they are again nearly as well localized to two centers (corresponding again to the double bonds in the classical chemical expression) as the occupied ones having smaller contributions from one or two additional atoms.

For U, A and G, the localization of the π -orbitals has given very similar results as for C. The only slight difference is the six-membered ring of A in which because of its pseudoaromatic structure one obtains polarized (with a majority of the populations on the nitrogen atoms) three-center localized filled orbitals. To demonstrate this we show in Table 3 the π -AO coefficients of the B and ER, respectively, localized seven filled π MO-s of guanine.

From the Table one sees that the localization is also in this case excellent, leading again essentially to two-center bonds and to lone pairs (for the numbering of the atoms see Fig. 1d). In Table 4 we present the Mulliken populations for the four virtual π^* orbitals of adenine (for numbering of the atoms see Fig. 1c) obtained after B localization. The localization is again rather good having the major

Table 3. The π -AO coefficients of the occupied localized π MO-s of guanine; the first number corresponds to B, the second one to ER, localization

Atom	LO1	LO1	LO3	LO4	LO5	LO6	LO7
N ₁	0.074	0.031	0.893	-0.078	0.113	-0.025	0.015
	0.072	0.029	0.895	-0.073	0.106	-0.020	0.013
C ₂	-0.227	-0.050	0.214	0.451	0.015	0.075	0.149
	-0.212	-0.059	0.211	0.457	0.009	0.083	0.150
N ₃	0.102	-0.036	-0.095	0.743	-0.055	0.007	0.023
	0.124	-0.052	-0.101	0.738	-0.052	0.014	0.021
C ₄	0.030	0.261	-0.064	0.205	0.078	-0.224	-0.472
	0.031	0.252	-0.065	0.211	0.097	-0.221	-0.472
C ₅	-0.037	-0.101	0.004	-0.088	0.036	0.096	-0.684
	-0.045	-0.110	0.002	-0.092	0.055	0.093	-0.681
C ₆	0.005	-0.013	0.158	-0.081	-0.549	-0.032	-0.173
	0.000	-0.015	0.154	-0.077	-0.545	-0.040	-0.191
N ₇	0.004	-0.150	-0.015	0.022	0.056	0.641	-0.155
	0.006	-0.157	-0.020	0.010	0.052	0.641	-0.153
C ₈	0.018	0.200	0.003	0.033	-0.037	0.610	0.217
	0.029	0.198	0.001	0.030	-0.052	0.612	0.211
N ₉	-0.024	0.857	0.022	-0.093	-0.014	-0.054	0.063
	-0.025	0.860	0.027	-0.073	-0.016	-0.048	0.052
N ₁₀	0.932	0.018	-0.072	0.129	-0.006	-0.029	-0.048
	0.935	0.022	-0.071	0.101	-0.004	-0.026	-0.041
O ₁₁	-0.003	0.017	-0.117	0.060	-0.745	0.022	0.145
	-0.001	0.018	-0.122	0.062	-0.748	0.012	0.122

Table 4. Mulliken populations of the virtual B-localized π^* -orbitals of adenine (B)

Atom	LO1	LO	LO3	LO4
N ₁	0.31	0.39	0.00	0.07
C ₂	1.03	0.00	0.00	0.01
N ₃	0.57	0.05	0.00	0.20
C ₄	0.02	0.01	0.01	0.96
C ₅	0.05	0.22	0.09	0.49
C ₆	0.00	1.10	0.00	0.00
N ₇	0.00	0.00	0.78	0.06
C ₈	0.01	0.01	0.92	0.02
N ₉	0.00	0.00	0.19	0.18
N ₁₀	0.01	0.22	0.00	0.00

contributions from two-centers with minor contributions from one or two additional centers.

Turning now to the localization of σ -orbitals in Table 5 we show the dominant σ -populations belonging to the 16 filled LO-s of U (for numbering of the atoms see Fig. 1b), if one does not take into account the π -electrons and the 1s electrons of the non-hydrogen atoms (the populations of a given LO of the non-listed atoms in Table 3 are smaller than 10^{-2}). In Table 6 we present the dominating coefficients of five selected LO-s (for the LO which corresponds to one of the lone pairs of O₇ and for the LO-s which describe the σ -bonds between atoms C₄-O₇, N₃-C₄, C₅-C₆ and C₆-H₁₀, respectively).

Table 5 shows that the σ -electrons are extremely well localized either on one of the oxygen atoms (lone pairs) or on a two-center σ -bond. For the two-center

Table 5. The dominant populations corresponding to the different localized σ -orbitals of uracil. The first number corresponds to the B, the second one to the ER LO

LO	N ₁	C ₂	LO	C ₅	H ₉
1 ^a	1.21	1.20	9 ^a	1.09	1.09
	C ₂	N ₃		C ₆	H ₁₀
2 ^a	0.95	0.86	10 ^a	1.10	1.10
	N ₃	C ₄		N ₁	H ₁₁
3 ^a	1.23	1.22	11 ^a	1.25	1.25
	C ₄	C ₅		N ₃	H ₁₂
4 ^a	1.04	1.05	12 ^a	1.25	1.25
	C ₅	C ₆		C ₄	O ₇
5 ^a	0.96	0.96	13 ^b	0.01	0.00
	C ₆	N ₁		C ₄	O ₇
6 ^a	0.84	0.84	14 ^b	0.00	0.00
	C ₄	O ₇		C ₂	O ₈
7 ^a	1.02	1.01	15 ^b	0.02	0.01
	C ₂	O ₈		C ₂	O ₈
8 ^a	1.01	1.01	16 ^b	0.02	0.01

^a σ bonds.

^b Lone pairs.

Table 6. The dominating coefficients of five selected LO-s of uracil. The first number refers to B, the second one to ER LO-s

AO	Atom	O ₇ lone pair	Atom	C ₄ -O ₇ σ bond	Atom	C ₄ -N ₃ σ bond	Atom	C ₆ -C ₅ σ bond	Atom	C ₆ -H ₁₀ σ bond
1s	O ₇	0.110	C ₄	0.101	C ₄	0.088	C ₆	-0.106	C ₆	0.097
2s		-0.711		-0.400		-0.289		0.379		-0.380
2P _x		-0.364		-0.425		0.253		0.443		0.197
2P _y		-0.637		0.092		0.280		-0.077		-0.428
1s	C ₄	-0.020	O ₇	0.062	N ₃	0.101	C ₅	-0.101	H ₁₀	-0.515
2s		0.114		-0.174		-0.437		0.330		-0.513
2P _x		0.059		0.535		-0.321		-0.410		-0.409
2P _y		-0.124		-0.085		-0.387		0.075		0.072

bonds one should note that the sum of the two populations is slightly larger than 2. This is because small positive and negative contributions of the populations belong also to the other atoms.

Table 6 shows that the populations localized on one or on two centers are made up always of the four orbitals $1s$, $2s$, $2p_x$ and $2p_y$, the contribution of the $1s$ orbital being in all cases much smaller than those of the other three AO-s. Furthermore Tables 5 and 6 again show, that the results of the B and ER procedures, respectively, are practically the same. Finally, it may be mentioned that similar results are obtained for the other three nucleotide bases C, A and G.

As a second procedure we did not introduce a $\sigma - \pi$ separation and have localized like Lipscomb et al. for (O_2 and CO_3^{2-} [12]) all the σ - and π -electrons together. In this case because of the formation of so-called τ ($\pi + \sigma$ and $\pi - \sigma$ orbitals, respectively) bonds, one also has a vertical localization, but one would expect that the extent of horizontal localization would decrease. Surprisingly this was not the case in both the B and ER procedures: for all the four nucleotide bases we obtained well localized two-center and reasonably well localized lone pair LO-s. The localization studies performed for the four nucleotide bases on the σ^* and π^* virtual orbitals have given similar results which will be published together with the construction of "oscillatory" virtual orbitals in a subsequent paper.

3. Conclusions

From the above described results we can conclude that: (i) the Boys procedure gives very similar results as the Edminston-Ruedenberg one. Therefore for a larger molecule with a low symmetry, one should prefer to use the method of Boys, since it is much faster than the ER method (ii) these localization studies have shown that for a planar molecule with heteroatoms one obtains a very good (one- or two-center) localization for the occupied and virtual orbitals even for the π electrons (if one uses σ - π separation), (iii) if one localizes all electrons together besides retaining nearly as good a horizontal localization as in the case of the $\sigma - \pi$ separation, one obtains the additional vertical localization because of the formation of τ bonds. Therefore, in correlation calculations this procedure is to be preferred even for planar molecules.

Acknowledgments. We should like to express our gratitude to Professor W. Forbes for critical reading of the manuscript and to Professor P. Otto for putting at our disposal his Boys localization program. One of us (J.C.) expresses his special thanks to the "Deutsche Forschungsgemeinschaft" for making his stay in Erlangen as a Visiting Professor and in this way this work possible. The financial support of the "Fond der Chemischen Industrie" is gratefully acknowledged.

References

1. Petrongolo, C., Buenker, R. J., Peyerimhoff, S. D.: J. Chem. Phys. **76**, 3655 (1982); Lie, G. C., Peyerimhoff, S. D., Buenker, R. J.: J. Chem. Phys. **75**, 2892 (1981)

2. Brueckner, K. A.: *Phys. Rev.* **100**, 36 (1955); Goldstone, J.: *Proc. Roy. Soc. (London)* **A 239**, 267 (1957); Primas, H., in: *Modern quantum chemistry*, Vol. 2, p. 45; ed. Sinanoglu, O., New York: Academic Press 1965; Pople, J. A., Seeger, R., Krishnan, R.: *Int. J. Quantum Chem.* **S11**, 149 (1977); Pople, J. A., Krishnan, R., Schlegel, H. B., Binkley, J. S.: *Int. J. Quantum Chem.* **14**, 545 (1978)
3. Paldus, J.: *J. Chem. Phys.* **61**, 5321 (1974); Paldus, J. in: *Theoretical chemistry: Advances and perspectives*, Vol. 2, p. 131; eds. Eyring, H., Henderson, D., New York: Academic 1976; Paldus, J., in: *The unitary group for the evaluation of electronic energy matrix elements*, *Lecture Notes in Chemistry*, Vol. 22, p. 1, ed. Hinze, J., Berlin: Springer 1981; Shavitt, I.: *Int. J. Quantum Chem.* **S11**, 131 (1977), *ibid* **S12**, 5 (1978); Shavitt, I., in: *The unitary group for the evaluation of electronic energy matrix elements*, *Lecture Notes in Chemistry*, Vol. 22, p. 51, ed. Hinze, J., Berlin: Springer 1981
4. Roos, B. O., Siegbahn, P. E. M., in: *Modern theoretical chemistry, methods of electronic structure theory*, Vol. III, ed. Schaefer III, H. F., New York: Plenum 1977
5. Goldstone, J.: *Proc. Roy. Soc. (London)* **A 239**, 267 (1957)
6. Mayer, W.: *J. Chem. Phys.* **58**, 1017 (1973); Kutzelnigg, W., in: *Modern theoretical chemistry*, Vol. III, ed. Schaefer III, H. F., New York: Plenum 1977; Zirz, C., Ahlrichs, R.: *J. Chem. Phys.* **75**, 4980 (1981)
7. Cizek, J.: *J. Chem. Phys.* **45**, 5456 (1966); Cizek, J.: *Adv. Chem. Phys.* **14**, 35 (1969)
8. Foster, J. M., Boys, S. F.: *Rev. Mod. Phys.* **32**, 300 (1960)
9. Edminton, C., Ruedenberg, K.: *Rev. Mod. Phys.* **34**, 457 (1963)
10. Boys, S. F., in: *Quantum theory of atoms, molecules and solid state*, p. 253, ed. Löwdin, P.-O. New York: Academic Press, 1966
11. Pipek, J.: *J. Comp. Chem.*, submitted for publication
12. England, W., Salomon, L. S., Ruedenberg, K., in: *Topics in current chemistry*, Vol. 23, p. 31, eds. Davison, A., Dewar, M. J. S., Hafner, K., Heilbronner, E., Hofmann, U., Niedenzu, K., Schäfer, Kl., Wittig, G., Berlin-New York: Springer 1971
13. Brown, L. D., Kleier, D. A., Lipscomb, W. N.: *J. Am. Chem. Soc.* **99**, 6787 (1977)
14. Ladig, W. D., Purvis, G. D. S., Bartlett, R. G.: *Int. J. Quantum Chem.* **S16**, 561 (1982)
15. Edminton, C., Ruedenberg, K.: *J. Chem. Phys.* **43**, 597 (1965)
16. Arnott, S., Dover, S. D., Wonacott, A. J.: *Acta Cryst.* **B25**, 2192 (1966)
17. Kleier, D. A., Halgren, T. A., Hall, J. H. Jr., Lipscomb, W. N.: *J. Chem. Phys.* **61**, 3905 (1974)

Received June 9, 1983