Theor Chim Acta (1988) 74: 101–110

Theoretica Chimica Acta

On the counterpoise correction for the basis set superposition error in large systems

J. A. Sordo[1], S. Chin[1], and T. L. Sordo[2]

[1] IBM Corporation, Data Systems Division, Department 48B, MS 428, Neighborhood Road, Kingston, NY 12401, USA
[2] Departamento de Quimica Fisica y Analitica, Universidad de Oviedo, C/Calvo Sotelo s/n, E-3307 Oviedo, Spain

(Received June 16, revised December 1/Accepted December 8, 1987)

The appropriateness of the use of the counterpoise correction for the basis set superposition error in SCF calculations of the interaction energies for pairs of aliphatic amino acids is analyzed in this paper. Our results show that for this type of molecule where the magnitude of the basis set superposition error can become quite big, the use of the counterpoise method provides interaction energies in good agreement with near Hartree–Fock values. The inaccuracies associated with the counterpoise method are much less important compared with the basis set superposition error itself. It is shown that the use of a well-balanced minimal basis set together with the counterpoise method is a good compromise (quality versus computational cost) for calculating interaction energies in systems involving molecules of biological interest.

Key words: Basis set superposition error — SCF — Counter poise correction— Amino acids

Introduction

The computation of the interaction energy of a supersystem, AB, formed by two nonreactive subsystems A and B, at the SCF or higher levels, always brings with it the so-called set superposition error (BSSE). In fact, Roothaan's approximation to the Hartree–Fock equations implies the use of a *finite* number of appropriate functions to expand the molecular (or atomic) orbitals of the subsystems A and B. The larger the number of functions the more accurate Roothaan's approximation.

If we symbolize (B_a) and (B_b) as the basis sets used to represent subsystems A and B, respectively, then $(B_a \cup B_b)$ will be the basis set used to compute the energy of the AB supersystem. The interaction energy can be written as

$$\Delta E(AB) = E(AB) - E(A) - E(B) \tag{1}$$

where the first term on the right hand side is evaluated by using the $(B_a \cup B_b)$ basis set and the two remaining terms are evaluated by using the (B_a) and (B_b) basis sets, respectively.

Equation (1) is valid *if and only if* the three terms appearing on the right hand side are evaluated at the same level of approximation. Therefore, Eq. (1) should be only used when the basis sets employed to compute $E(A)$ and $E(B)$ are large enough to be considered as saturated. The error appearing in using Eq. (1) when the (B_a) and (B_b) basis sets are not saturated is the well-known BSSE. The importance of correcting for the BSSE arises from the fact that the three terms on the right hand side of Eq. (1) are usually very large but $\Delta E(AB)$ is often of the order of a few kcal/mol and therefore, the correction can become significant.

In the early seventies, Boys and Bernardi [1] proposed the so-called counterpoise (CP) method to correct for the BSSE. The philosophy of the CP method consists of the evaluation of $E(A)$ and $E(B)$ in Eq. (1) using the $(B_a \cup B_b)$ basis set. The CP method has been widely exploited; one of its more appealing features being the relatively low computational cost required to use it. In fact, the CP estimation of the BSSE only demands the re-evaluation of the one-electron integrals as well as the SCF procedure. The re-evaluation of the two-electron integrals is not necessary.

Some years later, Johansson et al. [2] concluded that the CP method overcorrected the interaction energies. This overestimate has been related to the Pauli exclusion principle preventing the electrons in subsystem A from filling occupied orbitals of subsystems B and vice versa [3]. Therefore, a means of avoiding these overcorrections could be the use of only the virtual orbitals within the CP method, the so-called polarization counterpoise correction (PCP). [4] and [5] are examples of work in this direction.

Systematic studies on the BSSE have been carried out recently [6, 7]. The main conclusion arising from these is that neither CP nor PCP methods systematically improve the accuracy obtained with small basis sets and therefore, the best thing to do is just to increase the basis set up the maximum size affordable, thus trying to avoid the BSSE [6]. However, several recent papers report successful applications of the CP method to correct the interaction energies [8–11, 16], thus creating some controversy on the subject.

With only a few exceptions [8, 16] the above-mentioned literature is concerned with relatively small (about 20 electrons) systems. Therefore, the conclusions reached should be mostly applied to such systems and one should exercise caution in extrapolating these results to larger systems. In this paper we discuss the appropriateness of the use of the CP method to correct for the BSSE when computing the interaction energies for pairs of amino acids [12, 13].

Calculations

The interaction energies for about two thousand conformations involving pairs of amino acids (alanine-alanine, alanine-serine and serine-serine) were computed in order to obtain an analytical pair potential to represent such interactions [13]. CP correction for the BSSE was evaluated in all the cases.

From the above conformations, we selected nine geometries [14] involving hydrogen-bonded complexes and computed their interaction energies with three different basis sets: Pople's STO-3G [15] 7/3 (minimal) [16] and 9/5 (double-zeta) [16]. The last two have been specially designed to be used in the generation of analytic potentials and they have been shown to be well-balanced basis sets [16]. Of course, to enable a *complete* discussion of the appropriateness of the CP correction for the BSSE, polarization and diffuse functions should be added [6, 7]. Therefore, additional calculations have been carried out using a near-Hartree–Fock 13/8** basis set. This basis set is of double-zeta quality for the core orbitals and of triple-zeta quality for the valence orbitals. One three-membered *p*-type function was used as a polarization function for the hydrogens and one six-membered *d*-type polarization function was included for the heavier atoms. The exponents of the polarization functions were taken from [20]. The number of contracted functions for this basis set is, respectively 324, 344 and 364 for the systems alanin-alanine, alanine-serine and serine-serine.

All the calculations have been performed taking advantage of the LCAP parallel computer installation at IBM-Kingston. STO-3G and 7/3(SZ) calculations were carried out on the LCAP-1 (loosely coupled array of processors) architecture, consisting of an IBM 3081 host with ten FPS-164's attached processors. Most of our single-zeta calculations have been done by using a subset of four of the FPS-164's. On the other hand, the double-zeta calculations required the use of the LCAP-2 computer system, consisting of an IBM 3084 QX host and ten FPS-264's attached processors. Again, most of these calculations have been carried out using four of the FPS-264's.

Finally, the calculations with the 13/8** basis set required the use of, at least, five, six and seven FPS-264s attached processors (within the LCAP-2 architecture) running in parallel for the systems alanine-alanine, alanine-serine and serine-serine, respectively. To provide some quantitative idea about the computational effort required, in Table 1 we collect some details about the calculations for one of the geometries of the serine-serine system.

Results and discussion

Tables 2–5 collect our results for the interaction energies of pairs of amino acids in different conformations. ΔE^{NCP} stands for the interaction energies as calculated at the SCF level using Eq. (1) with no further corrections; ΔE^{CP} stands for the interaction energies using the CP method to correct for the BSSE. $\Delta\varepsilon$ is given by

$$\Delta\varepsilon = \Delta E^{CP} - \Delta E^{NCP} \tag{2}$$

Table 1. Some computational aspects of the calculations on the first geometry of the serine-serine system with the 13/8** basis set. All the numbers are in seconds. (AP stands for Attached Processor. See text for details)

System: serine–serine	HOST	AP:	#1	#2	#3	#4	#5	#6	#7
ONE-ELECTRON integrals Ser–Ser	623								
TWO-ELECTRON integrals Ser–Ser			5163	3693	3911	4308	4326	5059	5612
ONE-ELECTRON integrals Ghost–Ser	622								
ONE-ELECTRON integrals Ser–Ghost	622								
SCF (CPU)									
Ser–Ser			4324	3367	3348	3593	3826	4495	4549
Ghost–Ser			4333	3367	3340	3593	3826	4499	4567
Ser–Ghost			3996	3127	3107	3338	3554	4169	4227
SCF (I/O)									
Ser–Ser			10380	7917	7799	8389	8879	10428	10590
Ghost–Ser			11149	8526	8405	9037	9564	11231	11389
Ser–Ghost			11157	8524	8397	9034	9568	11222	11405

Total parallel time: 54 210
Total sequential time: 317 464

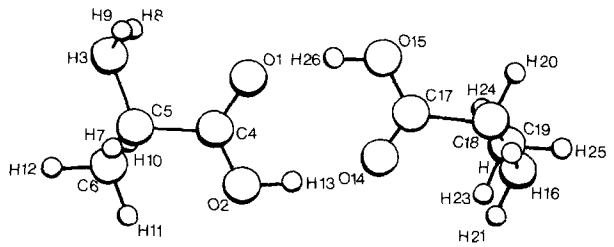
Therefore, $\Delta\varepsilon$ (which because of its definition always remains positive) directly measures the magnitude of the BSSE. All nine geometries on which the calculations have been carried out consist of complexes with a double hydrogen-bonded conformation involving the two —COOH groups, one on each interacting amino acid [13]. R(O···H) (see Fig. 1) stands for the distance between the carboxylic oxygen in one of the amino acids and the hydroxylic hydrogen in the second one.

Bearing in mind the origin of the BSSE, it seems reasonable to infer that such an error will depend mostly on three factors:

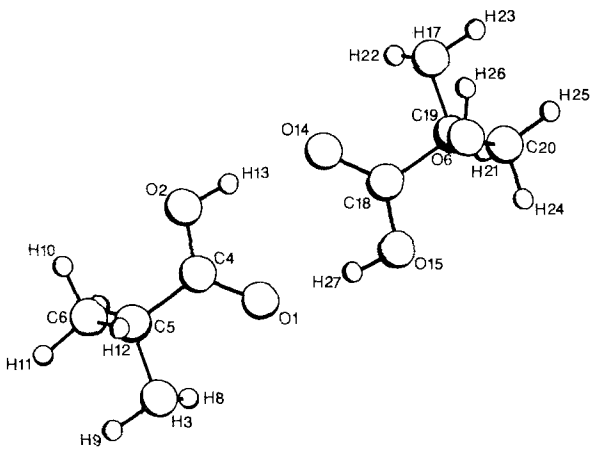
1. The number and type of atoms directly involved in the molecular association under study.
2. The proximity of the atoms directly involved.
3. The basis set used in the computation.

All three factors are important. In this study we wish to investigate the first point, and in particular the behavior of the CP method for correcting the BSSE in large systems of biological interest. Additional evidence on the relationship between the BSSE and the number (and type) of atoms directly involved in the interaction can be found in [8] and [16] (see also the discussion given below on the data plotted in Fig. 2).

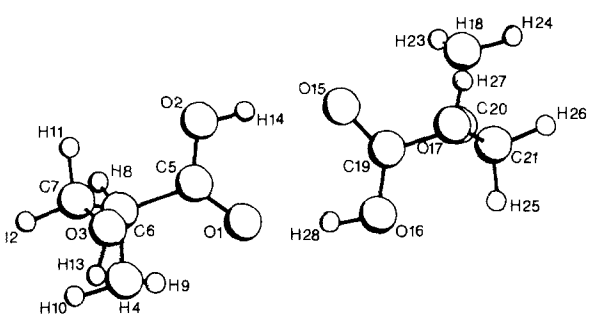
The CP correction for the BSSE has been questioned as a consequence of some recent work [6, 7] where a large variety of different basis sets were used to compute



a



b



c

Fig. 1. Configurations of the systems alanine-alanine (**a**), alanine-serine (**b**) and serine-serine (**c**) on which the calculations have been performed. For each configuration, three different distances O···H were considered (see Tables 2–5)

the interaction energies of several systems: $(HF)_2$, $(H_2O)_2$, $(NH_3)_2\cdots$. All these systems have a common property: they are relatively *small*. In considering point one above, a different behavior for the CP correction should be expected for larger systems. It now becomes quite clear why the conclusions previously reached about the appropriatenesss of the CP correction for the BSSE are contradictory in [6, 7] and [8, 17]. For small systems, the use of extended basis sets is feasible and therefore the BSSE can be notably reduced. In such a situation any anomalies associated with the CP method could become dramatically important [6, 7]. On the other hand, for larger systems where *only a minimal basis set may be used* [17], the associated large values of the BSSE must be corrected because, as we shall show below the effects of any anomalies are expected to be less important relative to BSSE itself. In this regard, Table 4 clearly shows that in spite of using double-zeta quality basis set (9/5(DZ)), the magnitude of the BSSE as calculated by means of the CP method is not negligible. The question remaining is whether, the CP correction goes in the correct direction or, on the contrary, if the anomalies in these cases are important enough to make ΔE^{NCP} more credible than ΔE^{CP}.

In order to address this question, the interaction energies for all the nine complexes were computed using a near-Hartree-Fock 13/8** basis set. Table 5 collects the results. Tables 3-5 show that the interaction energies as calculated using the CP method to correct for the BSSE are in much better agreement with 13/8** near Hartree-Fock results than the corresponding uncorrected interaction energies. This is true, independently of whether or not the near-Hartree-Fock energies are corrected for the BSSE.

Another important conclusion emerges from data collected in Tables 4 and 5. The generation of *ab initio* analytical pair potentials for the interaction between molecules of biological interest requires a computational effort such that only minimal basis sets are affordable. As stated elsewhere [13] the reliability of such calculations needs a case by case validation. In order to do so, some sample calculations using extended basis sets should be performed. For systems with, let us say, more than 30 atoms the use of near-Hartree-Fock basis sets (like the 13/8** used in this work) are impractical because of both CPU (see Table 1) and storage (the calculation reported in Table 1 required more than 8 Gbytes) requirements. Tables 4 and 5 show that the 9/5(DZ)+CP results are in good agreement with the near-Hartree-Fock calculations, thus allowing us to use the former as a good standard quality to assess the reliability of calculations performed using minimal basis sets [19].

On the other hand, Tables 2-5 also show that the generally accepted belief that the magnitude of the BSSE (as measured by means of the CP method) decreases when improving the quality of the basis set, is usually true. However, as pointed out in [6], this does not always happen. In fact, when passing from 7/3(SZ) to 9/5(DZ) basis sets for alanine-serine and serine-serine interactions at $R(O\cdots H) =$ 4.28 and 4.37 au, respectively, there is an increase of 0.5 and 0.8 kcal/mol in the CP correction for the BSSE (see Tables 3 and 4). Such anomalies have been reported by Schwenke and Truhlar [6] in their study on the $(HF)_2$ system, which led them to conclude that *the extra expense of a counterpoise correction is not*

Table 2. Interaction energies (ΔE^{NP}, ΔE^{NCP}) and CP estimate of the BSSE ($\Delta\varepsilon$) calculated with STO-3G basis set (see text for definitions). Distances are given in au and energies in kcal/mol

STO-3G basis set Pair	Geometry	R(O···H)	ΔE^{NCP}	ΔE^{CP}	$\Delta\varepsilon$
Alanine-alanine	1	2.80	−17.9	0.7	18.6
	2	3.32	−17.4	−6.1	11.3
	3	3.50	−15.6	−6.5	9.1
Alanine-serine	4	2.55	−11.4	7.2	18.6
	5	3.28	−16.8	−6.9	9.9
	6	4.28	−10.8	−5.4	5.4
Serine-serine	7	2.50	−10.2	11.8	22.0
	8	3.42	−16.1	−6.4	9.7
	9	4.37	−7.2	−5.0	2.2

Table 3. Interaction energies (ΔE^{NP}, ΔE^{NCP}) and CP estimate of the BSSE ($\Delta\varepsilon$) calculated with 7/3(SZ) basis set (see text for definitions). Distances are given in au and energies in kcal/mol

7/3(SZ) basis set					
Pair	Geometry	R(O···H)	ΔE^{NCP}	ΔE^{CP}	$\Delta\varepsilon$
Alanine-alanine	1	2.80	−19.2	−11.7	7.5
	2	3.32	−22.2	−16.4	5.8
	3	3.50	−20.9	−15.5	5.4
Alanine-serine	4	2.55	−12.3	−3.6	8.7
	5	3.28	−22.1	−16.4	5.7
	6	4.28	−12.5	−9.9	2.6
Serine-serine	7	2.50	−7.5	2.2	9.7
	8	3.42	−21.1	−15.3	5.8
	9	4.37	−11.7	−9.2	2.5

Table 4. Interaction energies (ΔE^{NP}, ΔE^{NCP}) and CP estimate of the BSSE ($\Delta\varepsilon$) calculated with 9/5(DZ) basis set (see text for definitions). Distances are given in au and energies in kcal/mol

9/5(DZ) basis set Pair	Geometry	R(O···H)	ΔE^{NCP}	ΔE^{CP}	$\Delta\varepsilon$
Alanine-alanine	1	2.80	−16.6	−9.4	7.2
	2	3.32	−20.5	−15.5	5.0
	3	3.50	−20.0	−15.4	4.6
Alanine-serine	4	2.55	−11.2	−3.0	8.2
	5	3.28	−20.9	−16.2	4.7
	6	4.28	−14.7	−11.6	3.1
Serine-serine	7	2.50	−6.2	2.4	8.6
	8	3.42	−19.8	−15.1	4.7
	9	4.37	−14.1	−10.8	3.3

Table 5. Interaction energies (ΔE^{NP}, ΔE^{NCP}) and CP estimate of the BSSE ($\Delta\varepsilon$) calculated with 13/8** (TZ plus polarization) basis set (see text for definitions). Distances are given in au and energies in kcal/mol

13/8**(TZ) basis set Pair	Geometry	R(O···H)	ΔE^{NCP}	ΔE^{CP}	$\Delta\varepsilon$
Alanine-alanine	1	2.80	−8.2	−6.6	1.6
	2	3.32	−15.5	−14.2	1.3
	3	3.50	−15.9	−14.7	1.2
Alanine-serine	4	2.55	−2.4	−0.5	1.9
	5	3.28	−16.5	−15.1	1.4
	6	4.28	−12.4	−11.7	0.7
Serine-serine	7	2.50	2.2	4.2	2.0
	8	3.42	−15.7	−14.3	1.4
	9	4.37	−12.0	−11.2	0.8

warranted, and it is better to increase the basis set to the maximum size affordable for noncounterpoise-corrected calculations. While we agree with this statement in the context where it was made (i.e. small systems with well balanced polarized basis sets), our results for large systems with well balanced unpolarized basis sets strongly support the use of the CP method. In fact, the improvement shown by ΔE^{CP} over ΔE^{NCP} (taking as a reference the results obtained with the near-Hartree-Fock 13/8** basis set with and without including the CP correction) at the double-zeta quality level (9/5(DZ) basis set) is much more significant than the magnitude of the anomalies mentioned above.

Furthermore, the extrapolation of the conclusions reported by Schwenke and Truhlar to the present work would mean the results coming from the 9/5(DZ) basis set (with no CP correction for the BSSE) being preferable to those obtained using 7/3(SZ)+CP. Taking as a reference the near-Hartree-Fock results collected in Table 5, it becomes evident that this is far from being true. The 7/3(SZ)+CP calculations provide much better interaction energies than the corresponding 9/5(DZ) without CP correction.

It is very encouraging to observe that the 7/3(SZ) basis set provides results in excellent agreement (almost quantitative), concerning BSSE ($\Delta\varepsilon$) and interaction energies (ΔE^{CP}), as compared with those provided by 9/5(DZ) basis set. This fact strongly supports the use of well-balanced minimal basis sets to compute interaction energies of biological systems [17, 18]. The choice is almost dictated by computational reasons. However, it is very gratifying to see the relatively good performance exhibited by such basis sets. The same conclusion, however, does not apply to the STO-3G basis set. Table 2 shows that the CP corrected values obtained with STO-3G basis set (ΔE^{CP}) are far from the corresponding 7/3(SZ) and 9/5(DZ) basis sets estimates. The overcorrection associated with the CP method becomes, for the STO-3G basis set, much too large. ΔE^{CP} and ΔE^{NCP} values reported in Table 2 clearly indicate the poor quality exhibited for the interaction energies computed with the STO-3G basis set. For short distances, the STO-3G basis set (including the CP correction) is not only quantitatively

inadequate but also qualitatively inadequate. In fact, geometries 1 and 4 are calculated as repulsive conformations whereas the 7/3(SZ) and 9/5(DZ) basis sets predict them as attractive. As stated elsewhere [13] potential curves for systems of biological interest calculated with the STO-3G basis set depart greatly from their expected behavior.

Figure 2 plots the values of the CP correction for the BSSE versus the distances R(O···H) for the nine conformations computed in this work. It is interesting to note that with only a few exceptions (e.g. R(O···H) = 3.28 au in Fig. 2), the CP values decrease as the distance R(O···H) increases. The smooth decrease observed for the 13/8** basis sets confirms that the magnitude of the BSSE is mostly governed by the atoms directly involved in the interaction (—COOH···HOOC— in our case) as well as by their relative positions (distances), thus being almost independent of the functional groups attached to the carboxylic groups in both interacting molecules. Again, the STO-3G basis set shows a much more irregular behavior in this regard while the 7/3(SZ) behavior approaches that exhibited by the 9/5(DZ) basis set.

Before concluding, we would like to emphasize the limitations of the present study where only three different basis sets have been compared. In two of the three cases the CP correction usually improves accuracy, using 13/8** results as the comparison. While the results presented here clearly indicate the importance of the CP method for correcting the BSSE for large systems computed with small well balanced basis sets, the fact that in one case the CP method does not improve the accuracy points to the need for further studies including other basis sets commonly referred to in the literature in order to reinforce our conclusions. In this regard, a complementary study (which includes Huzinaga's, Pople's and Clementi's geometrical basis sets) is currently being developed in this laboratory and will be presented in a forthcoming paper [21].

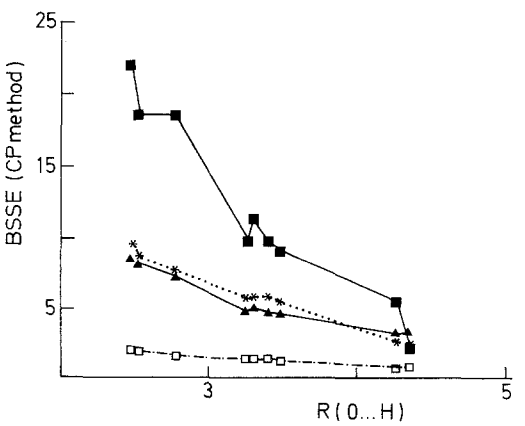


Fig. 2. Correlation between the BSSE (kcal/mol) and the distance R(O···H) (au) (see the text for definition) for the nine conformations computed with the STO-3G (■-■); 7/3(SZ) (*-*); 9/5(DZ) (▲-▲); and 13/8** (□-□) basis sets

Conclusions

The main conclusion of this work refers to the appropriateness of the use of the counterpoise method to correct for the basis set superposition error in calculations of the interaction energies for systems of biological interest. Conclusions arising from studies on smaller systems, where the interaction energies are usually small and the irregularities detected when using the counterpoise method could become significant are not applicable, at least in a straightforward manner, to larger systems involving strong hydrogen bonded interactions between molecules of biological interest. For these last kinds of systems, the use of a well-balanced minimal basis set with the associated basis set superposition error corrected by means of the counterpoise method, is strongly recommended.

Acknowledgments. One of us (J.A.S.) wishes to express his gratitude to IBM-Kingston for its very generous hospitality and financial support. The authors wish to thank Dr. R. Gomperts for his help and advice in the use of the KGNMOL package and Dr. J. Watts who kindly carried out some HONDO runs needed at the early stage of this work. Finally, the authors are indebted to Dr. E. Clementi for helpful discussions throughout the course of this work.

References and notes

1. Boys SF, Bernardi F (1970) Mol Phys 19:553 (1986) 86:681
2. Johansson A, Kollman P, Rothenberg S (1973) Theor Chim Acta 29:167
3. Morokuma K, Kitaura K (1981) In: Politzer P, Truhlar D (eds) Chemical applications of atomic and molecular electrostatic potentials. Plenum Press, New York
4. Olivares del Valle FJ, Tolosa S, Esperilla JJ, Ojalvo EA, Requena A (1986) J Chem Phys 84:5077
5. Olivares del Valle FJ, Tolosa S, Ojalvo EA, Esperilla JJ (1986) J Chem Phys 85:3448
6. Schwenke DW, Truhlar D (1985) J Chem Phys 82:2418 (1986) 84:4113(E); (1987) 87:3760(E)
7. Frisch MJ, Del Bene JE, Binkley JS, Schaefer III HF (1986) J Chem Phys 84:2279
8. Greenwood HH, Plant JS (1986) Int J Quantum Chem 30:127
9. Bonaccorsi R, Cammi R, Tomasi J (1986) Int J Quantum Chem 29:373
10. Szczensniak MM, Scheiner S (1986) J Chem Phys 84:6328
11. Sokalski WA, Lowrey AH, Roszak S, Lewchenko V, Blaisdell J, Hariharan PC, Kaufman JJ (1986) J Comput Chem 7:693
12. Sordo JA, Probst M, Chin S, Corongiu G, Clementi E (1986) In: Clementi E, Chin S (eds) Structure and dynamics of nucleic acids, proteins, and membranes. Plenum Press, New York
13. The geometries corresponding to the nine complexes studied can be found in Sordo JA, Chin S, Sordo TL: IBM Tech. Rep. KGN-130, available from the authors upon request
14. Sordo JA, Probst M, Chin S, Corongiu G, Clementi E (1987) J Am Chem Soc 109:1702
15. Hehre WH, Stewart RF, Pople JA (1969) J Chem Phys 51:2657
16. Gianolio L, Pavani R, Clementi E (1978) Gazz Chim Ital 108:181
17. Hobza P, Sandorfy C (1987) J Am Chem Soc 109:1302
18. Kolos W (1979) Theor Chim Acta 51:219; (1980) *ibid* 54:187
19. Sordo JA, Chin S and Clementi E (1987) IBM Technical Report KGN-149, August 19
20. Hehre WJ, Radom L, Schleyer PR v, Pople JA (1986) *Ab initio* molecular orbital theory. Wiley, New York
21. Sordo JA, Gomperts R, Sordo TL, Chin S, Clementi E (in preparation)