721

# A new scaling procedure to correct semiempirical MEP and MEP-derived properties

Carlos Alemán[a], F.J. Luque[b],* and Modesto Orozco[c],*

[a]Dept. Enginyería Química, ETSEIB, U. Politécnica de Catalunya, Avgda Diagonal 647, 08028 Barcelona, Spain
[b]Dept. Farmacia, Unitat Fisicoquímica, Fac. Farmacia, U. Barcelona, Avgda Diagonal sn, 08028 Barcelona, Spain
[c]Dept. Bioquímica, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

## SUMMARY

The MNDO, AM1, PM3, and ab initio 6–31G* and 6–31+G* MEPs for 21 neutral and 12 charged molecules were computed in layers ranging from 1.2 to 2.0 times the van der Waals radii of atoms. Semiempirical and ab initio MEPs for each layer and two groups of layers were compared to gain insight into the relationships between semiempirical and ab initio MEPs. A detailed statistical study allowed us to obtain a new set of scaling coefficients able to correct the semiempirical MEPs to provide better representations of the ab initio values. The corrected semiempirical MEPs were used to obtain electrostatic charges, whose quality was tested by the comparison between semiempirical Coulombic MEPs and ab initio quantum mechanical MEPs.

## INTRODUCTION

The ability of the Molecular Electrostatic Potential (MEP; Eq. 1) to describe and predict the chemical behavior of complex molecules has been well established during the past 20 years [1–25]. The MEP is widely used in very different fields, ranging from the prediction of organic reactivity [2–12] to the study of complex bioorganic systems [13–19], solvation phenomena [20,21], or crystalline systems [22–25]. Nowadays, the relevance of the MEP has been reinforced owing to its use as a source of atomic charges, which are obtained by fitting quantum mechanical and Coulombic MEPs ([26–28]; Eqs. 2–3) in regions outside the van der Waals radii of the atoms. These atomic charges, known as potential-derived or electrostatic charges, are used in a large number of molecular mechanics, dynamics or Monte Carlo studies (see for example Refs. 29–35), since they give reliable estimates of the electrostatic interaction between molecules.

---

*To whom correspondence should be addressed.

$$V^{QM}(r_l) = \sum_A \frac{Z_A}{|r_l - R_A|} - \sum_\mu \sum_\nu P_{\mu\nu} \int \frac{\chi_\mu(r)\,\chi_\nu(r)}{|r_l - r|}\,dr \tag{1}$$

where $Z_A$ is the nuclear charge, $\chi$ is the set of the atomic orbitals, and $P_{\mu\nu}$ is the first-order density matrix.

$$V^C(r_l) = \sum_i \frac{q_i}{|r_l - R_i|} \tag{2}$$

$$\sum_k \left[ V^{QM}(r_k) - V^C(r_k) \right]^2 = minimum \tag{3}$$

The quality of the MEP strongly depends on the quality of the wave function used to determine the density matrix [36–46]. Previous studies determined that, for simple molecules, an SCF 6–31G* calculation describes the essential features of the MEP, and that the increase of the basis set, or the incorporation of electron correlation effects does not change the main characteristics of the MEP [43,45,46]. Use of mixed basis sets provides good approximation to the 6–31G* MEPs for some molecules [42], at a more reduced computational cost. Use of split-valence or minimal basis sets leads to qualitatively correct MEPs, but large numerical differences with respect to the 6–31G* MEPs are found.

In spite of the increase in computer performance, there are still a very large number of molecules which cannot be considered at the ab initio level, even using a very small basis set. In these cases, semiempirical methods are an excellent alternative, since they reliably evaluate molecular properties at a reasonable computational cost. In this respect, the use of Dewar's [47,48] and Stewart's [49] methods as a source of MEPs for large molecules has been proposed by several groups [43,44,50–53].

Several studies stressed the reliability of the semiempirical MEPs to describe the most relevant features of the ab initio 6–31G* MEPs, but some failures also became clear [43,44,50,54,55]. For instance, semiempirical methods underestimate the ab initio MEP in outer regions, and accordingly the semiempirically derived electrostatic charges for neutral molecules are smaller than the 6–31G* charges [50,51,55]. This defect may be corrected by means of a scaling procedure, which was first suggested by Cox and Williams [56]. The procedure fits the 6–31G* and semiempirical charges in a series of molecules to an equation of the type $y = cx$, (y: 6–31G*; x: semiempirical), where c is the scaling coefficient.

The procedure improves the quality of semiempirical charges [50,51,55,57,58]. However, it still presents some shortcomings: (1) it cannot be used for charged molecules, (2) it assumes that the ratio MEP ab initio/MEP semiempirical in outer regions is independent of the distance to the nuclei, (3) it is based on the assumption that scaled semiempirical charges may reproduce the ab initio QM MEPs, and (4) it is not applicable in cases like SCRF calculations [59,60], where it can be of interest to scale the MEP in areas much closer to the molecule than those considered in the fitting of MEPs.

In this paper, we try to overcome these defects by reformulating the scaling procedure [56]. Our approach is based on the direct scaling of the MEP instead of on the scaling of charges. In order to test this method, the suitability of semiempirical QM MEPs to reproduce high-level ab initio QM MEPs at different layers around the molecule (from 1.2 to 2.0 times the van der Waals radii

of atoms) was explored. From this study, which was performed with 21 neutral and 12 charged molecules, the validity of a single scaling factor to correct the semiempirical MEP in all the regions of the space is discussed. An extended statistical analysis allowed us to determine scale factors for the semiempirical MEPs for neutral and charged molecules. Finally, these scaling factors were directly applied to the QM semiempirical MEPs, which in turn give corrected semiempirical electrostatic charges. The ability of these corrected charges to represent the ab initio QM MEP was examined by comparing the corresponding semiempirical Coulombic MEPs and the high level ab initio QM MEPs.

## METHODS

MEPs were computed for all the molecules by using four Connolly layers [61] placed at distances of 1.2, 1.3, 1.4, 1.6, 1.8, and 2 times the van der Waals radii. This set of layers covers most of the regions where electrostatics is expected to be the most important molecular interaction. A density of five points/$\text{Å}^2$ was considered in order to have a complete representation of the MEP distribution on the layer. Electrostatic charges were determined using the MEP in layers from 1.4 to 2.0 times the van der Waals radii following Kollman's suggestions [29].

The ab initio MEPs were computed by using standard procedures [2]. The 6–31G* and 6–31+G* basis sets were used to describe the neutral and charged molecules, respectively. The semiempirical MEPs were computed at the MNDO [47], AM1 [48], and PM3 [49] levels, following the deorthogonalization procedure developed in our laboratory [43,44,50]. This procedure is based on: (1) the deorthogonalization of the semiempirical wave function by means of the Löwdin inverse transformation, (2) the expansion of each Slater-type orbital in four Gaussian functions to facilitate the calculation of integrals in Eq. 1, and (3) the use of 'core' effective charges to evaluate the nuclear electrostatic term. Experimental geometries were used in all cases to make possible the direct comparison between the MEPs computed at the different computational levels.

The analysis of the huge amount of data collected was carefully performed by using a variety of tests which provide information on the similarity between semiempirical and ab initio MEPs. Thus, for each molecule (i) and each layer (j) the following calculations were performed:

*Root-mean-square deviation*    (Eq. 4), which gives an average estimate of the total deviation between semiempirical and ab initio MEPs.

$$RMS_i^j = \sqrt{\frac{\sum\limits_{\kappa=1}^{N} (MEP_{ab\ initio} - MEP_{semiempirical})^2}{N}} \tag{4}$$

where N is the total number of points at the layer, and the MEP is always computed rigorously from the wavefunction (QM MEP).

*Relative root-mean-square deviation*    (Eq. 5), which gives an average estimation of the relative deviation between semiempirical and ab initio MEPs.

$$\text{Rel RMS}^j_i = \sqrt{\frac{\sum\limits_{\kappa=1}^{N} (MEP_{ab\ initio} - MEP_{semiempirical})^2}{(MEP_{ab\ initio})^2}} \tag{5}$$

*Linear regression fitting*  Here the ab initio (y) and the semiempirical (x) MEPs for each molecule (i) and for each layer (j) were fitted to a regression equation of the type: y = c x. The Pearson correlation coefficient (r) gives the quality of fitting, and hence an estimation of the consistency of the semiempirical MEPs for this molecule at this layer. The slope of the equation (scaling factor) gives an estimation of the relative deviation between semiempirical and ab initio MEPs.

Averaged information for the whole of molecules was obtained by grouping the layers in two categories: 1.2 and 1.3 (inner layers) and from 1.4 to 2.0 (outer layers). The statistical test described above was performed for the MEPs in both inner and outer regions. The values of c, r, RMS, and relRMS for the 'inner' and 'outer' regions for each molecule (data not shown in the paper) were then averaged for all the molecules (in the paper these values are denoted as 'average'). In addition to these 'average' values, we also computed the 'global' values, which were determined from the comparison between the semiempirical and ab initio MEPs in each set of layers for all the molecules simultaneously. Comparison between 'global' and 'average' values gives an estimation of the regularity of the semiempirical MEP.

The 'global' scaling coefficients were used to scale the semiempirical MEPs, which were used to compute corrected electrostatic charges. The suitability of these charges was determined from their ability to reproduce the 6–31G* (or 6–31+G*) QM MEPs. The statistical test used was the RMS(s), defined as shown in Eq. 6.

$$RMS(s) = \sum_{\kappa=1}^{N} \frac{\left( \sum\limits_{\kappa=1}^{N} \sum\limits_{j=1}^{L} \frac{(MEP_{ab\ initio}^{QM} - MEP_{s}^{coulombic})}{N * L} \right)^{1/2}}{M} \tag{6}$$

where s means the method to be considered. In most cases s = AM1, MNDO, or PM3, but for comparison the ab initio Coulombic MEPs were fitted to the ab initio QM MEPs (see Results). The RMS(s) values give a magnitude of the ability of the Coulombic potential obtained by corrected semiempirical charges to reproduce ab initio QM MEPs.

Semiempirical wave functions were determined by using our version of the MOPAC program [62–64]. Ab initio wave functions were determined by using a locally modified version of the HONDO-77 program [65]. MEPs and electrostatic charges were determined with the MOPETE/MOPFIT program [66]. Calculations were performed on an HP-730 and on the IBM-3090 of the Centre de Supercomputació de Catalunya.

RESULTS AND DISCUSSION

In order to clarify the discussion, the results are presented in two different sections, the first corresponding to the analysis of the 21 neutral molecules and the second to charged molecules.

Each section is presented in two blocks: the first deals with the analysis of the similarity between semiempirical and ab initio MEPs in different regions of space, and the second with the scaling procedures presented.

## NEUTRAL MOLECULES

*Analysis of the MEPs in different regions of space*

Twenty-one molecules of small or medium size which contain diverse relevant functional groups were selected. Figure 1 shows histogram representations of the correlation coefficients obtained when the semiempirical and ab initio MEPs were fitted. The high quality of the fittings, especially for MNDO, is very remarkable. In general, the worst correlations appear for the inner layers, mainly for those located at 1.2 times the van der Waals radii. Nevertheless, there is not a clear relationship in the range of distance considered between the quality of the fitting for semiempirical and 6–31G* results and the layer considered in the calculation of the MEP. The ability of the semiempirical MEPs to fit 6–31G* values depends mainly on the method and on the molecule, but not on the distance to the molecule.

The molecules that exhibit a poor correlation (r < 0.90) between semiempirical and ab initio MEPs were considered as 'outliers' and then removed from the statistical studies performed to obtain scaling coefficients and averaged values in order to avoid bias in the results. For MNDO, only $CH_4$ is an outlier for layers from 1.2 to 1.8 times the van der Waals radii, and $C_2H_4$ is also an outlier for some layers. It should be pointed out that for 14 molecules (almost 70% of the molecules) MNDO and 6–31G* MEPs correlated almost exactly (r > 0.98) in all the external layers. The number of outliers for AM1 increases to four for most of the layers: $CH_4$, $CH_2CO$, furan, and pyridine. The number of molecules where the correlation between AM1 and 6–31G* is excellent changes depending on the layer considered. Thus, it is only 4–6 for inner layers and rises to 10 for external layers. The large number of molecules with correlation coefficients in the range of 0.92 < r < 0.98 is also noticeable. These results are in agreement with the known 'irregularity' of the MEP derived from AM1 calculations that was previously noted in mediocre correlations between AM1 and 6–31G* electrostatic charges [50,51]. Finally, PM3 shows a number of 'outliers' depending on the layer considered. For the most internal layer, four outliers appear: $CH_4$, $CH_2CO$, $NH_3$, and furan. The $NH_3$ is not an outlier in the remaining layers, and the $CH_2CO$ and $CH_4$ are well fitted for most external layers (up to 1.4 for the first, and 1.6 for the second molecule). The number of molecules where the fit between PM3 and 6–31G* MEPs is almost perfect (r > 0.98) is 7 for the 1.2 layers, and 10–12 for the remaining layers.

The results in Fig. 1 illustrate good correlations between semiempirical and 6–31G* MEPs, especially in the layers located at up to 1.2 times the van der Waals radii. The correlations are especially remarkable for the MNDO method, which also exhibits the lower number of 'outliers'. PM3 and AM1 MEPs correlate poorly to 6–31G* values, but the quality of the fitting is enough to guarantee in general the suitability of PM3- and AM1-derived MEPs.

Figure 2 displays the scaling coefficients obtained for MNDO, AM1 and PM3 considering all the molecules (except the 'outliers') and the different layers. (The scaling coefficients for $CO_2$ were initially considered in the AM1 plot in Fig. 2. Nevertheless, it exhibits an anomalously large scaling factor ranging from 2.4 (1.2) to 1.8 (2.0), even if it is not an outlier (r > 0.96). Accordingly, in order to maintain a reasonable scale in Fig. 2 (AM1) we kept it out of the plot.) In general,
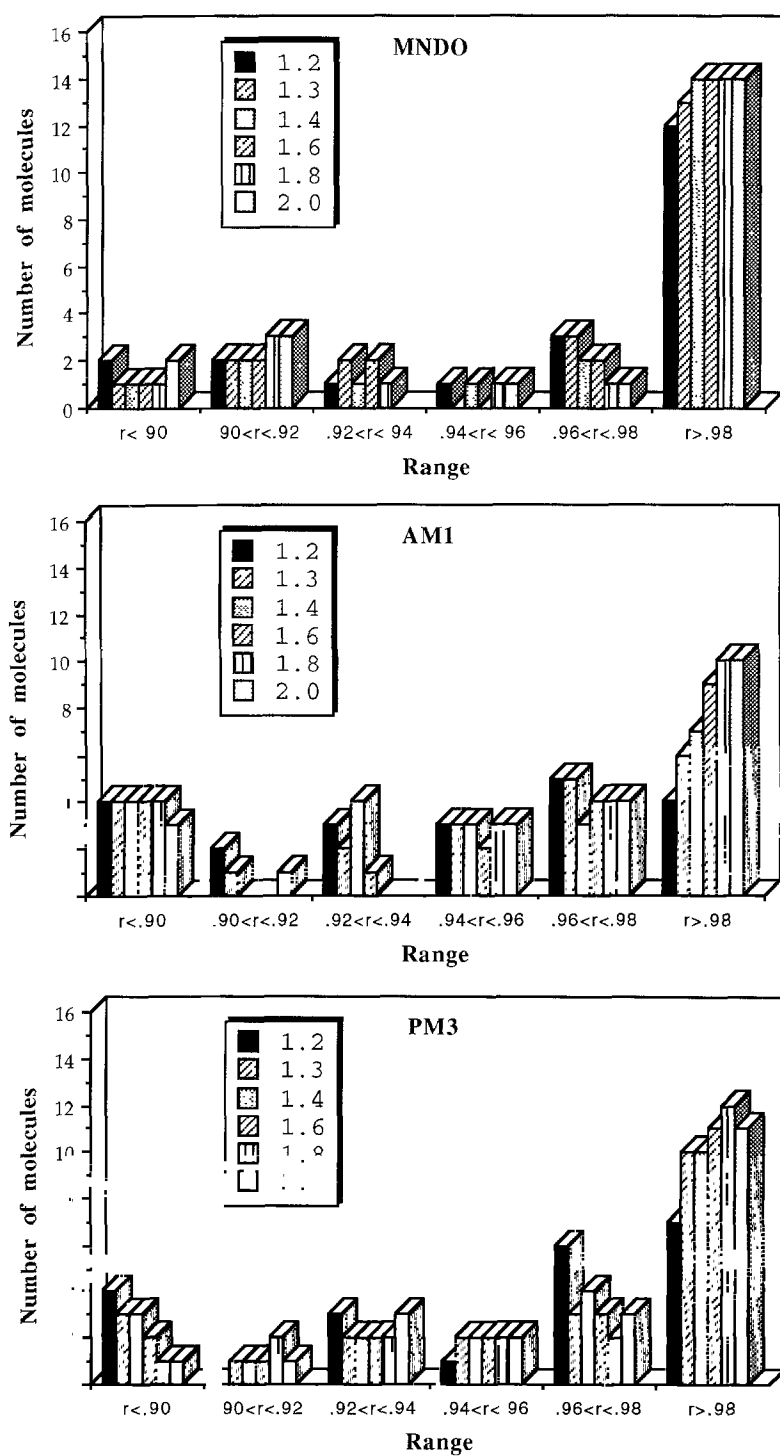
Fig. 1. Histogram representations of the Pearson correlation coefficients between semiempirical and 6–31G* MEPs fc neutral molecules. The number of outliers for each layer is shown in the bars corresponding to r < 0.90.
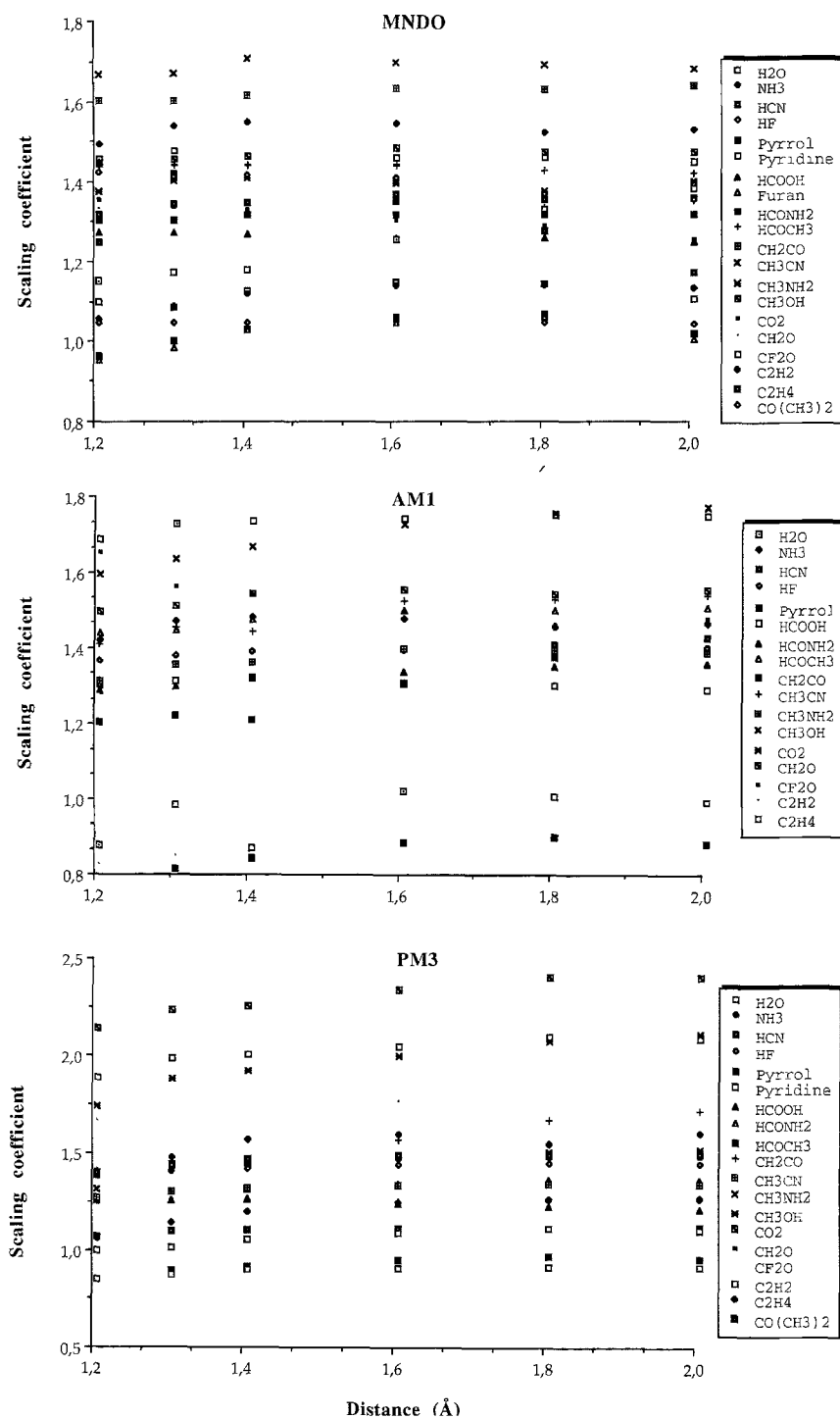
Fig. 2. Values of the scaling coefficients obtained between semiempirical and 6–31G* MEPs for neutral molecules. Outliers were omitted from the figure. For the AM1 plot, $CO_2$ was considered, but the scaling coefficient (about 2.6) was kept out of scale to clarify the figure.
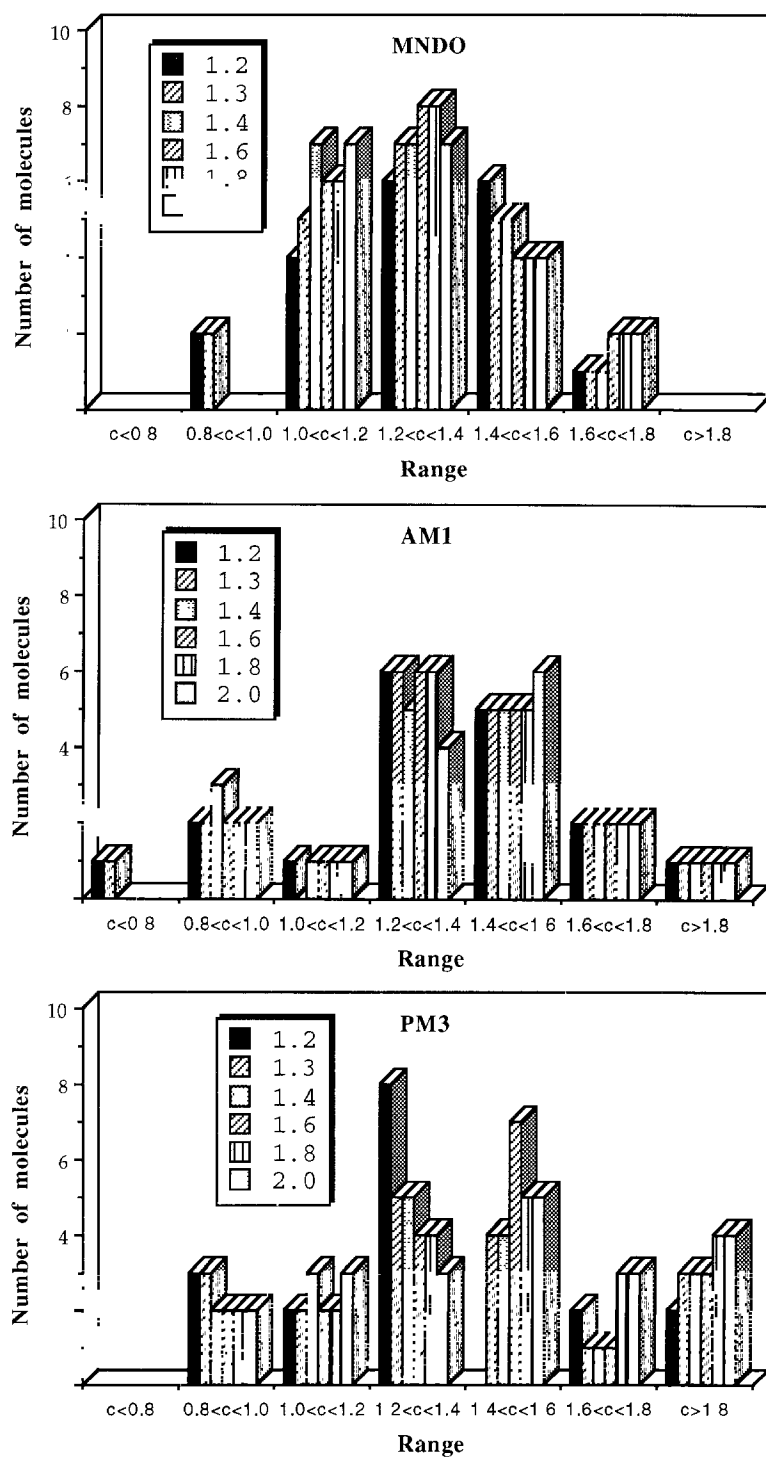
Fig. 3. Histogram representations of the scaling coefficients between semiempirical and 6–31G* MEPs for neutra molecules. Outliers were omitted from this figure.

semiempirical methods underestimate the magnitude of the MEP, as stated by the scaling coefficients greater than one. Indeed, the dependence of the scaling coefficients on the layer is very small. Thus, the scaling coefficients increase for the layers from 1.2 to 1.4, but remain nearly unchanged for the layers from 1.4 to 2.0. Keeping in mind the excellent correlation coefficients between semiempirical and ab initio MEPs, this finding strongly supports the use of a single scaling coefficient to correct the semiempirical MEPs in all the outer molecular regions.

An averaged and more compact representation of the scaling coefficients obtained from semiempirical calculations is shown in Fig. 3 (where, as usual, outliers are removed). This plot clearly points out the superiority of MNDO with regard to the other two semiempirical methods, as is shown in the 'Gaussian' shape of the histogram, which indicates (see also Fig. 2) that MNDO gives similar scaling coefficients for almost all molecules. Scaling coefficients for AM1 exhibit a larger variation, but in general most of the molecules take values in the range $1.2 < c < 1.6$. Finally, PM3 shows the worst behavior (see also Fig. 1), as indicated by the large variation in the scaling coefficient, and by the large number of molecules with scaling coefficients greater than 1.8. Therefore, despite the excellent correlations between PM3 and 6–31G* MEPs, this method gives a very irregular profile when different molecules are considered. As a consequence, the possibility to develop algorithms to correct the PM3 MEP is strongly handicapped.

*Scaling*

For each molecule, MEPs were grouped in two sets: layers 1.2 and 1.3, and layers 1.4 to 2.0. Semiempirical and 6–31G* MEPs in these two groups of layers were fitted for each molecule. Data obtained for nonoutliers were used to obtain 'average' values for the whole set of molecules (see Methods). Furthermore, semiempirical and ab initio MEPs were compared for the two sets

TABLE 1
STATISTICAL RESULTS OF THE COMPARISON BETWEEN SEMIEMPIRICAL AND 6–31G* MEPs IN INNER (1.2, 1.3) AND OUTER (1.4, 1.6, 1.8, AND 2.0) LAYERS FOR NEUTRAL MOLECULES[a]

| | Method | Correlation coefficient | Scaling coefficient | RMSD | Relative RMSD |
|---|---|---|---|---|---|
| Inner layers | MNDO | 0.97 (0.03) | 1.31 (0.20) | 5.71 (1.88) | 0.32 (0.2) |
| | | 0.96 | 1.26 | 5.9 | 0.33 |
| | AM1 | 0.96 (0.03) | 1.38 (0.38) | 7.46 (2.32) | 0.39 (0.09) |
| | | 0.93 | 1.25 | 4.42 | 0.40 |
| | PM3 | 0.97 (0.03) | 1.34 (0.36) | 6.46 (2.46) | 0.35 (0.11) |
| | | 0.95 | 1.21 | 6.68 | 0.36 |
| Outer layers | MNDO | 0.98 (0.02) | 1.32 (0.19) | 3.30 (1.22) | 0.31 (0.09) |
| | | 0.97 | 1.29 | 3.60 | 0.32 |
| | AM1 | 0.98 (0.02) | 1.44 (0.41) | 4.24 (1.51) | 0.37 (0.11) |
| | | 0.95 | 1.30 | 4.42 | 0.38 |
| | PM3 | 0.97 (0.03) | 1.43 (0.37) | 3.93 (1.61) | 0.36 (0.13) |
| | | 0.96 | 1.27 | 4.05 | 0.36 |

[a] The plain numbers correspond to the 'average' values and their standard deviations (in parentheses); the 'global' values are in italics (see Methods). Outliers (molecules for which the semiempirical/6–31G* correlations were of insufficient quality $(r < 0.90)$) were eliminated in order to avoid bias in the results. This implies that the following molecules were not considered in the table: $CH_4$ (for all the methods except PM3 in outer layers), furan (AM1 and PM3), $CH_2CO$ (AM1 and inner layers of PM3), pyridine (AM1) and $NH_3$ (inner layers of PM3). RMSDs are in kcal/mol.

of layers, thus providing 'global' values. The 'global' and 'average' statistical parameters are shown in Table 1.

Inspection of Table 1 confirms the similar behavior of semiempirical MEPs in inner and outer regions, although the ability of semiempirical methods to represent the 6–31G* MEPs is slightly smaller in inner regions. Both average and global parameters demonstrate the suitability of the three semiempirical methods to give suitable approximations to 6–31G* MEPs. Comparison of the 'average' and 'global' parameters allows us to check the consistency of semiempirical methods. Thus, a large difference between 'average' and 'global' values means that the method produces irregular results for the different molecules. Bearing this in mind, the superiority of the MNDO method over the other two semiempirical methods becomes evident. The better performance of MNDO is also demonstrated by the smaller values of the RMS, relative RMS, and scaling coefficients, as well as in the excellent correlation coefficients and the small standard deviations of the 'average' values.

It is worth noting that the scaling factors obtained by fitting the 6–31G* and semiempirical MEPs at layers from 1.4 to 2 times the van der Waals radii would be close to those previously obtained by simple fitting of charges [50,51,55,67]. The latter were 1.31 (MNDO), 1.33 (AM1) and 1.44 (PM3). The fitting of MEPs leads to values of 1.29 (MNDO), 1.30 (AM1) and 1.27 (PM3). That is, almost the same scaling coefficients are detected for MNDO and AM1, but different values are found for PM3. Note also that the 'global' correlation coefficient for the MNDO MEPs in the region 1.4–2.0 is 0.97, which matches the value previously obtained by fitting charges. Nevertheless, the correlation coefficients obtained by AM1 and PM3 in the region 1.4–2.0 are much better, 0.95 (AM1) and 0.96 (PM3), than those reported for charges (0.88 in both cases). These discrepancies indicate that AM1 and PM3 electrostatic charges for neutral molecules should not be used to investigate the fine details of electron distributions in molecules. Nevertheless, these charges might be useful for MM, MD, or MC calculations, since they are able to reproduce properly the ab initio QM MEP.

The 'global' scaling factors reported in Table 1 were used to scale semiempirical MEPs, which in turn were used to fit electrostatic charges for all the molecules ('outliers' were included). The results (the complete list of charges is available from the authors upon request) are summarized in Fig. 4, where the charges obtained from the scaled semiempirical MEPs (corrected charges) are displayed against the 6–31G* values. A straight line corresponding to c = 1 is also plotted as reference. The results demonstrate that the electrostatic charges obtained from scaled semiempirical MEPs are very similar to the 6–31G* values, the similarity being excellent when MNDO is used.

Finally, the ability of the corrected semiempirical charges to give reasonable estimates of the 6–31G* electrostatic energy for the whole set of molecules was tested (outliers included). The RMS(s) (see Methods) between the Coulombic MEPs obtained from standard and corrected semiempirical charges, and the 6–31G* QM MEP were used. The results are displayed in Table 2, where the RMS(s) between the 6–31G* Coulombic and QM MEPs is also included for comparison. The results clearly show the great similarity between Coulombic and QM MEPs, which supports the suitability of its use for general representations of the MEP. As expected, the smallest RMS(s) was detected when 6–31G* Coulombic and QM MEPs are compared (1.1 kcal/mol). The standard semiempirical charges lead to moderate deviations ranging from 3.2 kcal/mol (MNDO) to 4.0 kcal/mol (AM1). Use of corrected charges leads to a reduction in the
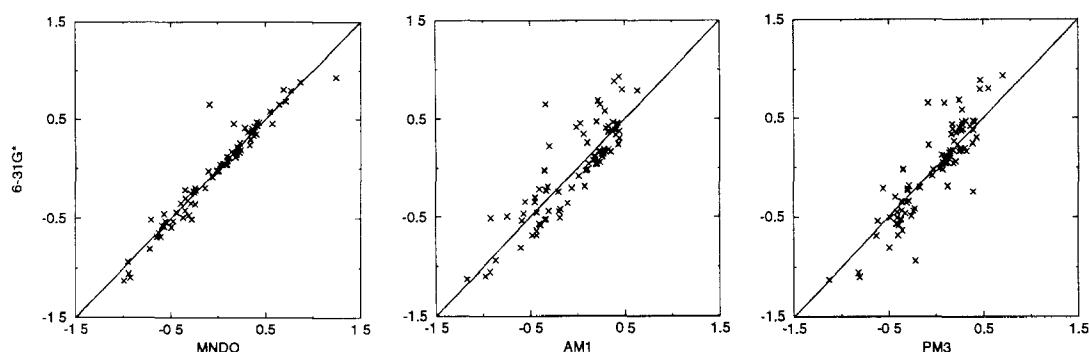
Fig. 4. Graphical representations of the relationship between electrostatic charges obtained by fitting scaled semiempirical MEPs and 6–31G* electrostatic charges for neutral molecules. The ideal correlation (y = cx with c = 1.00) is plotted as a reference. All the molecules are included in the plots. Charges are in fractions of electron.

RMS(s), which in the case of MNDO reaches values of only 1.9 kcal/mol. It is worth noting that such a value compares very well with the optimum value of 1.1 kcal/mol obtained when 6–31G* Coulombic and QM MEPs were compared, which strongly supports the suitability of the corrected set of MNDO charges.

## CHARGED MOLECULES

### Analysis of the MEPs in different regions of space

The ability of the semiempirical MEPs to represent high-level ab initio MEPs for charged molecules (six anions and six cations) was explored using the 6–31+G* basis set as reference. Choice of this basis set observes the need for proper representation of the diffuse electron density in anions. Figures 5 and 6 display histogram plots corresponding to the Pearson correlation coefficients between semiempirical and 6–31+G* MEPs for cations (Fig. 5) and anions (Fig. 6) at the different layers. Figure 5 shows excellent correlations between MNDO and 6–31+G* in all regions of space. AM1 also shows a satisfactory behavior, with only one molecule (pyridine$^+$)

TABLE 2
RMS(S) BETWEEN COULOMBIC AND QUANTUM MECHANICAL MEPs IN OUTER LAYERS FOR NEUTRAL MOLECULES[a]

| Method | RMS(s) |
|---|---|
| 6–31G* | 1.1 |
| MNDO | 3.2 |
| MNDO (cor) | 1.9 |
| AM1 | 4.0 |
| AM1 (cor) | 3.2 |
| PM3 | 3.9 |
| PM3 (cor) | 3.1 |

[a] Coulombic MEPs were computed from the semiempirical charges (obtained after scaling of the MEP) and the 6–31G* charges. The QM MEP was computed at the 6–31G* level. All the molecules were included in the analysis shown in this table. All values are in kcal/mol.
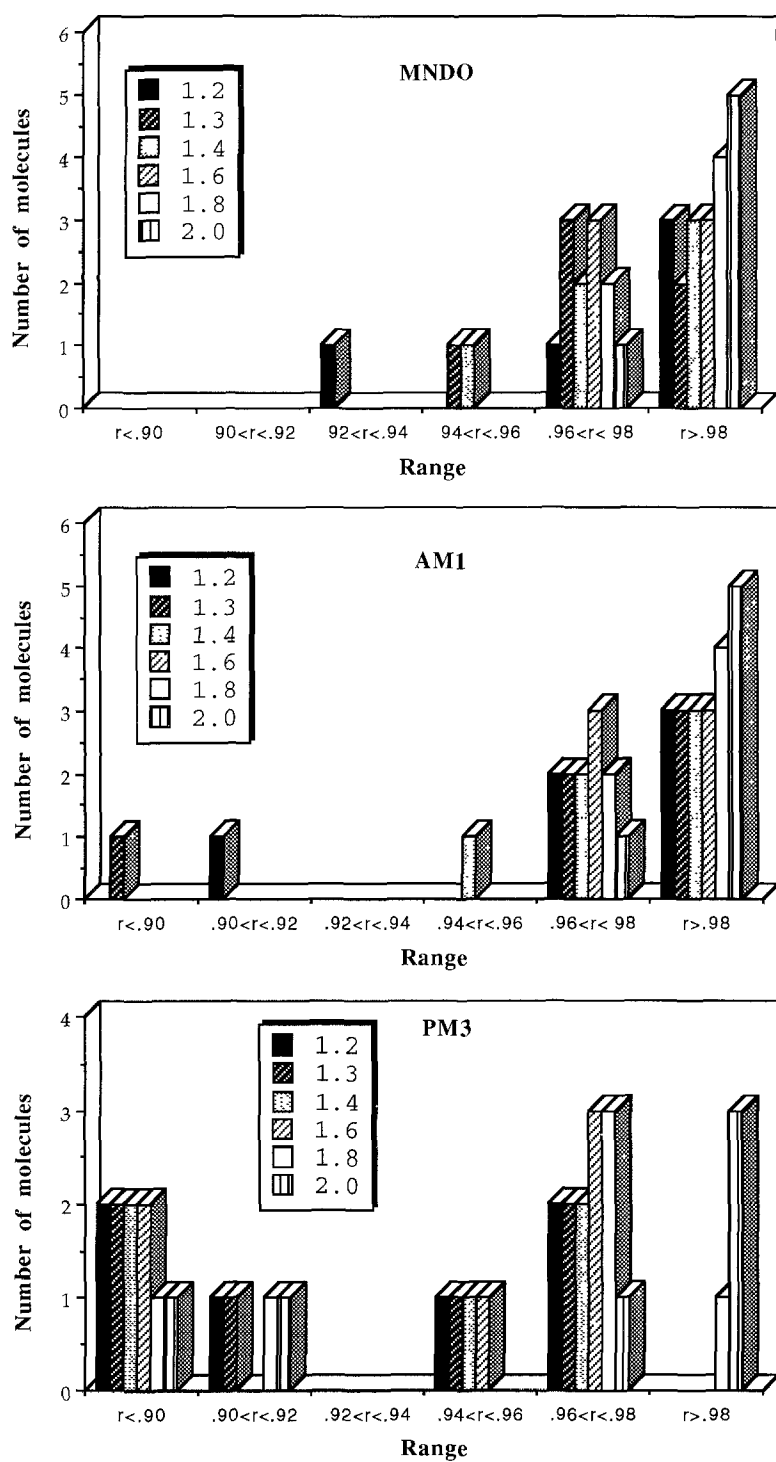
Fig. 5. Histogram representations of the Pearson correlation coefficients between semiempirical and 6–31+G* MEPs fo
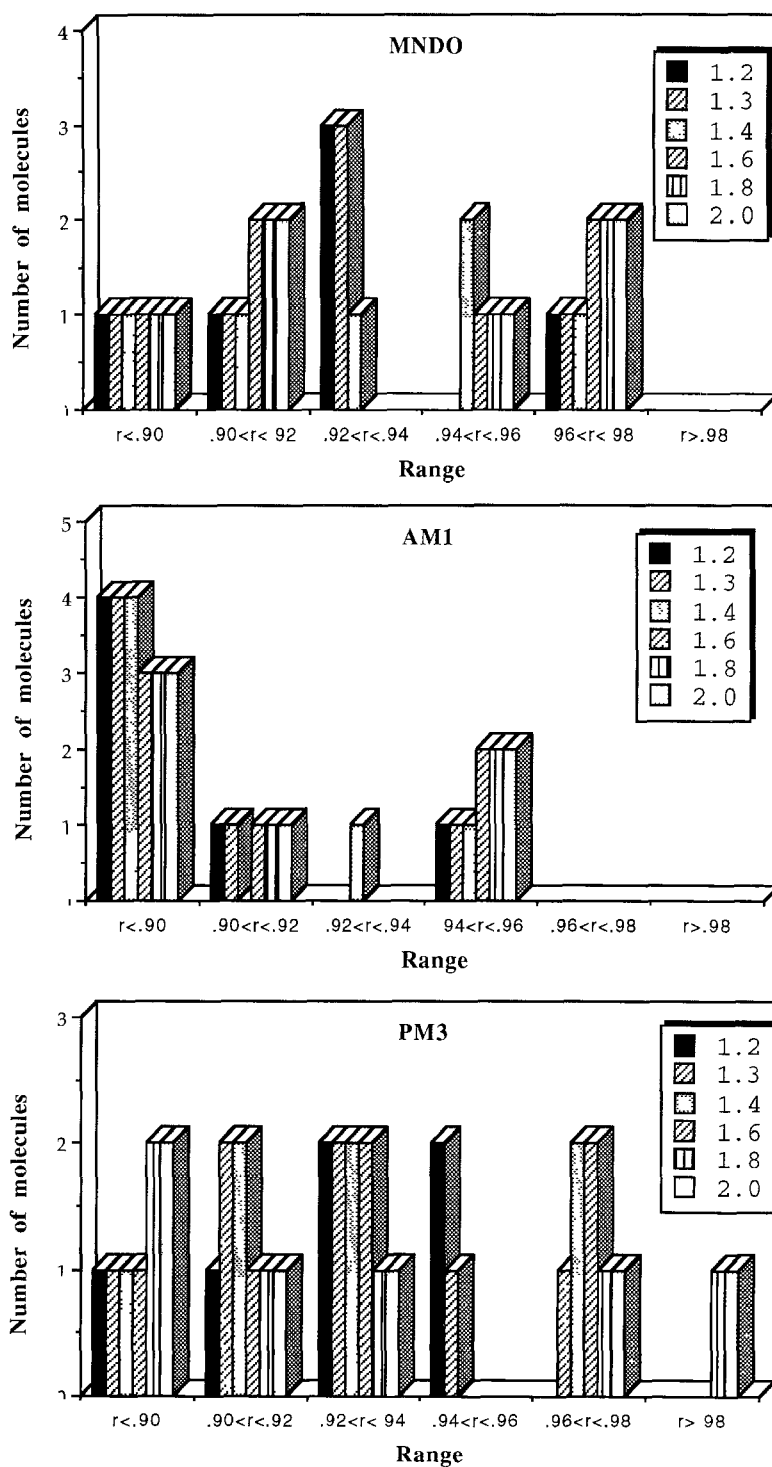cations. The number of outliers for each layer is shown in the bars corresponding to r < 0.90.

Fig. 6. Histogram representations of the Pearson correlation coefficients between semiempirical and 6–31+G* MEPs for anions. The number of outliers for each layer is shown in the bars corresponding to r < 0.90.
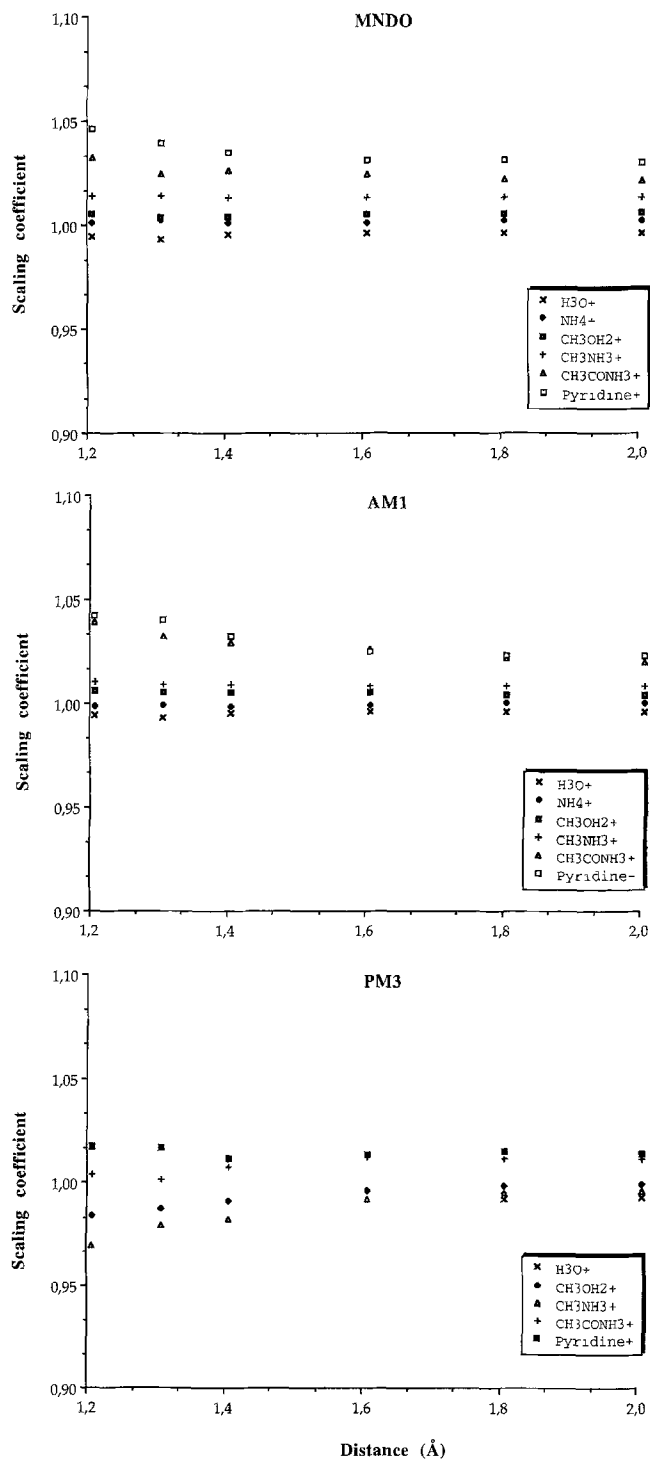
Fig. 7. Values of the scaling coefficients obtained between semiempirical and 6–31+G* MEPs for cations. Outliers were omitted from the figure.

having r < 0.90 for one of the layers (1.3). PM3 shows the poorest behavior, since $H_3O^+$ and $NH_4^+$ are outliers for inner layers, and $NH_4^+$ is also an outlier in the outer regions. This finding confirms the failure of PM3 to deal with the proper charge definition for the latter molecule [67–70]. Furthermore, few molecules achieve perfect correlations (r > 0.98) for all the layers in the PM3 calculations.

Figure 6 demonstrates the poorer correlation of semiempirical MEPs with 6–31+G* MEPs for anions. Nevertheless, when MNDO is considered, only $CN^-$ correlates poorly with the 6–31+G* data. PM3 also shows good behavior, with the $CN^-$ being an outlier for all the layers and $OH^-$ having poor correlations only for the outer layers (1.8 and 2). Finally, AM1 shows the worst behavior for anions, as shown by the large number of 'outliers'.

In summary, MEPs obtained from MNDO calculations show the best correlations with high-level ab initio calculations for both cations and anions. AM1 gives MEPs that correlate very poorly for cations. PM3 gives MEPs for anions that correlate reasonably well, but its behavior is worse for cations. No clear relationship between the correlation coefficients and the distance from the nuclei to the layer was found. Nevertheless, in general, the MEP in the inner layers is the most difficult to reproduce in semiempirical calculations.

The scaling coefficients (c) obtained from the fitting (y = cx) between 6–31+G* MEP (y) and semiempirical data (x) are shown in Fig. 7 (cations) and 8 (anions) (outliers not included). Note that scaling coefficients for all the methods, and for all the molecules, are close to the ideal value of 1, both for cations and anions. No variation in the scaling coefficient of the distance to the nuclei was detected. However, for anions (see Fig. 8) the semiempirical MEP in inner layers reached larger values than 6–31+G* did, whereas for outer layers (distance to the nuclei greater than 1.4 Å) the scaling coefficient was very constant and close to 1.

TABLE 3

STATISTICAL RESULTS OF THE COMPARISON BETWEEN SEMIEMPIRICAL AND 6–31+G* MEPs IN INNER (1.2, 1.3) AND OUTER (1.4, 1.6, 1.8, AND 2.0) LAYERS FOR CATIONS[a]

|  | Method | Correlation coefficient | Scaling coefficient | RMSD | Relative RMSD |
|---|---|---|---|---|---|
| Inner layers | MNDO | 0.98 (0.01) | 1.01 (0.01) | 4.49 (2.11) | 0.03 (0.02) |
|  |  | 0.99 | 1.01 | 4.62 | 0.02 |
|  | AM1 | 0.97 (0.03) | 1.01 (0.01) | 3.93 (1.12) | 0.03 (0.02) |
|  |  | 0.98 | 1.01 | 5.34 | 0.04 |
|  | PM3 | 0.95 (0.02) | 0.99 (0.01) | 6.07 (1.53) | 0.05 (0.01) |
|  |  | 0.98 | 0.98 | 6.72 | 0.05 |
| Outer layers | MNDO | 0.99 (0.01) | 1.01 (0.01) | 2.24 (1.46) | 0.02 (0.01) |
|  |  | 0.99 | 1.01 | 2.88 | 0.02 |
|  | AM1 | 0.99 (0.01) | 1.01 (0.01) | 2.16 (1.72) | 0.04 (0.03) |
|  |  | 0.99 | 1.01 | 3.00 | 0.03 |
|  | PM3 | 0.99 (0.01) | 0.99 (0.01) | 3.10 (0.74) | 0.03 (0.00) |
|  |  | 0.99 | 0.99 | 3.03 | 0.03 |

[a] The plain numbers correspond to the 'average' values and their standard deviations (in parentheses); the 'global' values are in italics (see Methods). Outliers (molecules for which the semiempirical/6–31G* correlations were of insufficient quality (r < 0.90)) were eliminated in order to avoid bias in the results. This implies that for the PM3 method $NH_4^+$ and the inner layers of $H_3O^+$ were not considered.
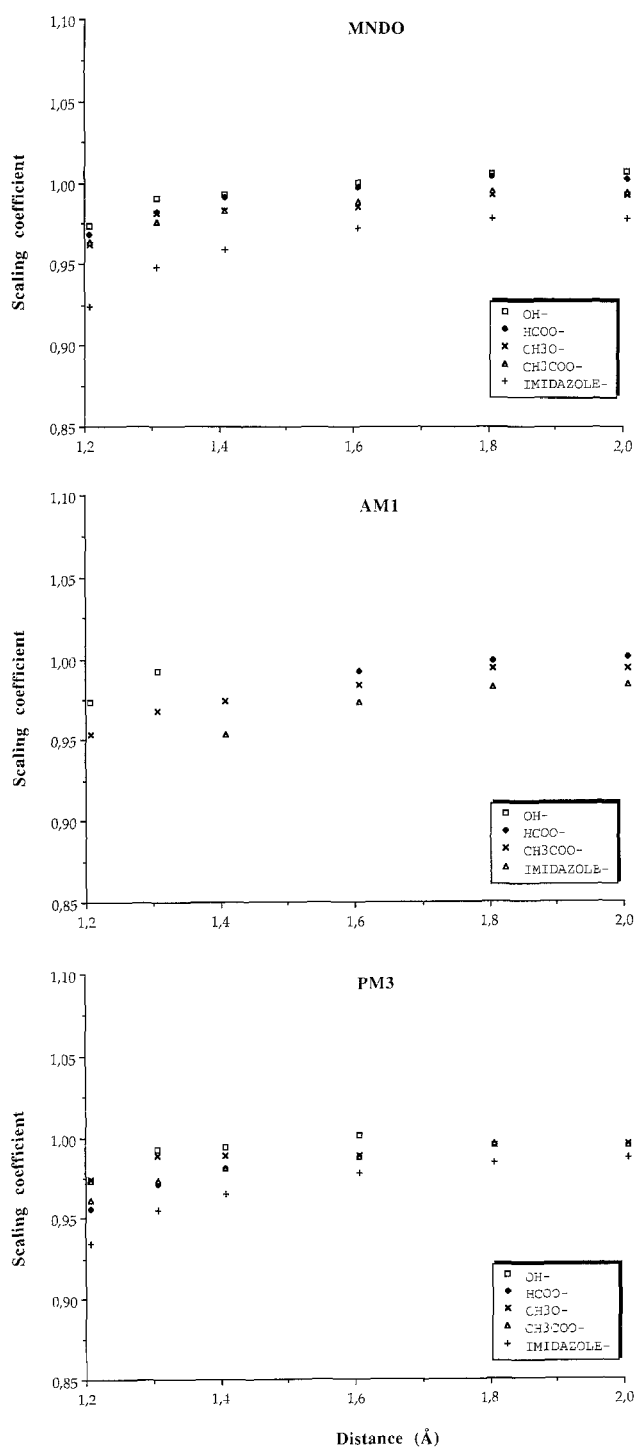
Fig. 8. Values of the scaling coefficients obtained between semiempirical and 6–31+G* MEPs for anions. Outliers were omitted from the figure.

TABLE 4

STATISTICAL RESULTS OF THE COMPARISON BETWEEN SEMIEMPIRICAL AND 6-31+G* MEPs IN INNER (1.2, 1.3) AND OUTER (1.4, 1.6, 1.8 AND 2.0) LAYERS FOR ANIONS[a]

|  | Method | Correlation coefficient | Scaling coefficient | RMSD | Relative RMSD |
|---|---|---|---|---|---|
| Inner layers | MNDO | 0.94 (0.02) | 0.96 (0.02) | 10.9 (2.0) | 0.08 (0.02) |
|  |  | 0.96 | 0.96 | 10.9 | 0.05 |
|  | AM1 | 0.92 (0.03) | 0.95 (0.03) | 13.4 (1.3) | 0.09 (0.01) |
|  |  | 0.95 | 0.97 | 13.1 | 0.09 |
|  | PM3 | 0.93 (0.02) | 0.96 (0.02) | 11.2 (1.0) | 0.08 (0.01) |
|  |  | 0.95 | 0.96 | 11.1 | 0.08 |
| Outer layers | MNDO | 0.96 (0.02) | 0.99 (0 01) | 5.48 (2.00) | 0.05 (0.01) |
|  |  | 0.97 | 0.99 | 5.64 | 0.05 |
|  | AM1 | 0.94 (0.02) | 0.99 (0.01) | 6.82 (2.01) | 0.06 (0.02) |
|  |  | 0.95 | 0.99 | 6.69 | 0.06 |
|  | PM3 | 0.96 (0.02) | 0.99 (0.01) | 5.56 (2.12) | 0.05 (0.02) |
|  |  | 0.97 | 0.99 | 5.70 | 0.05 |

[a] The plain numbers correspond to the 'average' values and their standard deviations (in parentheses); the 'global' values are in italics (see Methods). Outliers (molecules for which the semiempirical/6-31G* correlations were of insufficient quality (r < 0.90)) were eliminated in order to avoid bias in the results. This implies that the following molecules were not considered in the table: $CN^-$ for all the methods, $OH^-$ (outer layers of AM1 and PM3), $CH_3O^-$ (AM1), $HCOO^-$ (inner layers of AM1), and imidazole (inner layers of AM1).

*Scaling procedure*

Tables 3 and 4 show the statistical data corresponding to the comparison between semiempirical and 6-31+G* MEPs in the inner and outer layers for cations and anions, respectively. Table 3 demonstrates the ability of semiempirical MEPs to reproduce high-level ab initio values. This is noted in the Pearson correlation coefficients, the scaling coefficients (c = 1.0), and the RMSDs in the range of 4-6 kcal/mol, which imply errors of only 3-5% with respect to the 6-31+G* value. MNDO gives the best representation of the 6-31+G* MEP, but AM1 also

TABLE 5

RMS(S) BETWEEN COULOMBIC AND QUANTUM MECHANICAL MEPs IN OUTER LAYERS FOR CHARGED MOLECULES[a]

| Method |  | RMS(s) |
|---|---|---|
| Cations | 6-31+G* | 1.1 |
|  | MNDO | 2.0 |
|  | AM1 | 2.3 |
|  | PM3 | 2.9 |
| Anions | 6-31+G* | 2.6 |
|  | MNDO | 2.9 |
|  | AM1 | 7.1 |
|  | PM3 | 6.1 |

[a] Coulombic MEPs were computed from the semiempirical and the 6-31+G* charges. The QM MEP was computed at the 6-31+G* level. All the molecules were included in the analysis shown in this table. All values are in kcal/mol.
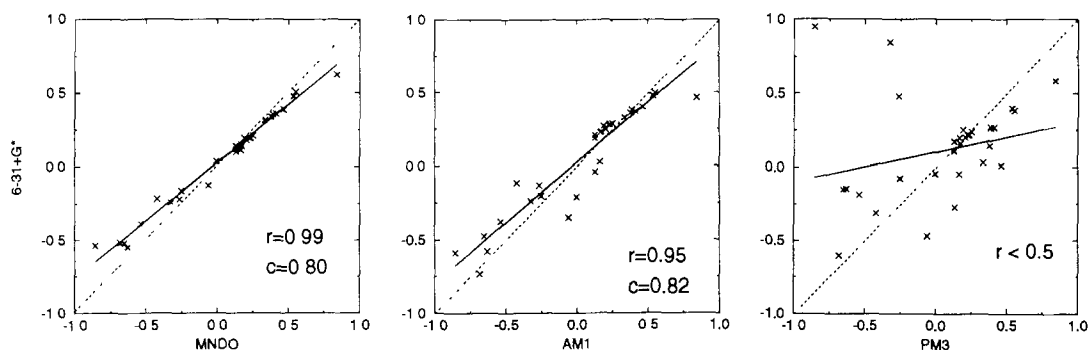
738



Fig. 9. Graphical representations of the relationship between electrostatic charges obtained by fitting semiempirical MEPs and 6–31+G* electrostatic charges for cations. The correlation equations between semiempirical and 6–31+G* charges are shown as solid lines, while the ideal correlation (c = 1.00) is displayed with a dotted line. All the molecules are included in the plots. Charges are in fractions of electron.

provides reasonable values. PM3 gives correct estimates on average, but the inability to describe a key molecule like $NH_4^+$ should be kept in mind.

The results displayed in Table 4 show that semiempirical methods are reasonably well able to describe the MEP of anions. The RMSDs are in the range of 5–13 kcal/mol, which implies percentual errors of less than 10% in the MEP. As noted before, the three methods are less able to provide the details of ab initio MEPs for anions, especially for inner layers. The scaling coefficients demonstrate that semiempirical methods slightly overestimate the MEP in the inner regions (c ≈ 0.96), while correct values of MEPs are obtained for outer layers (c ≈ 1.0). The Pearson correlation coefficients are worse than those detected for neutral molecules and cations. Once again, MNDO provides the best results for most molecules (only $CN^-$ was considered an outlier and removed from the analysis shown in Table 4).

Since the global scaling factors were almost equal to 1.0, no scaling procedure was performed to correct the semiempirical MEPs for charged molecules. The RMS(s) between semiempirical
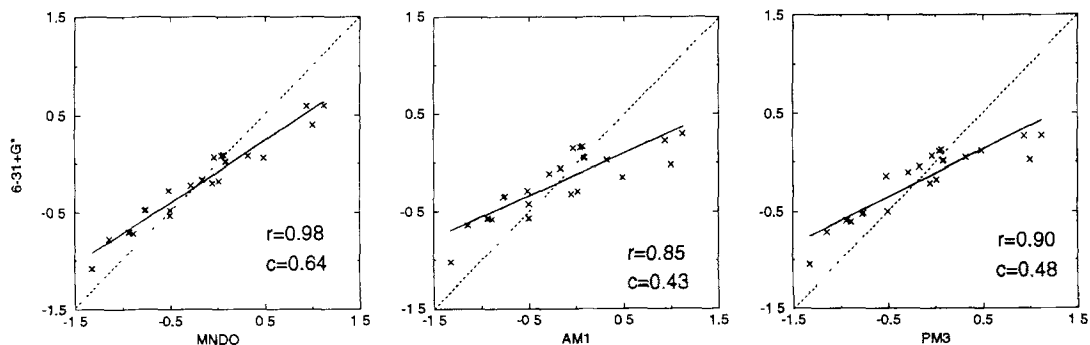


Fig. 10. Graphical representations of the relationship between electrostatic charges obtained by fitting semiempirical MEPs and 6–31+G* electrostatic charges for anions. The correlation equations between semiempirical and 6–31+G* charges are shown as solid lines, while the ideal correlation (c = 1.00) is displayed with a dotted line. All the molecules are included in the plots. Charges are in fractions of electron.

Coulombic and QM 6–31+G* MEPs was computed for all the cations and anions (outliers were included). The results in Table 5 point out the ability of Coulombic semiempirical MEPs to describe the 6–31+G* QM MEPs. For cations, both MNDO and AM1 give average deviations around 2.0–2.3 kcal/mol, which are not too different from the optimum difference of 1.1 kcal/mol found when 6–31+G* Coulombic and QM MEPs are compared. PM3 gives larger values of the RMS(s) for cations (2.9 kcal/mol), due to the presence of $NH_4^+$. The bottom part of Table 5 demonstrates the ability of the Coulombic MNDO MEPs to represent the 6–31+G* QM MEPs for anions, as noted by the small value of the RMS(s): 2.9 kcal/mol, almost identical to the optimum value of 2.6 kcal/mol obtained when 6–31+G* Coulombic and QM 6–31+G* MEPs are compared.

Figures 9 and 10 show the values of the semiempirical and 6–31+G* charges for cations and anions, respectively. The ideal correlation between semiempirical and ab initio MEPs (c = 1.0) is represented as a dotted line, while the real correlation is displayed as a solid line. Excellent correlations between MNDO and 6–31+G* charges are found for cations. AM1 also correlates well, but the presence of outliers leads to very poor correlations for PM3. For anions (Fig. 10) MNDO again correlates well, while AM1 correlates the worst.

The scaling coefficients for the charges (Figs. 9 and 10) are about 0.8 for cations (AM1, MNDO), and between 0.6 (MNDO) and 0.4 (AM1 and PM3) for anions. It is worth noting that these values are very different to those detected by the fitting of MEPs (approximately 1.0). This fact implies that semiempirical charges are uniformly overestimated with respect to the 6–31+G* values. Nevertheless, this deviation does not imply that semiempirical MEPs are equally overestimated. This is understandable considering the fact that the dominating contribution to the total MEP for charged molecules arises from the monopole. Accordingly, an approximate representation of the atomic charges is enough to reproduce high-level ab initio MEPs satisfactorily. It should be noted that the scaling of charges by factors as shown in Figs. 9 and 10 must be avoided, since they lead to a decrease in the total molecular monopole, and subsequently to a large error in the MEP.

## FINAL REMARKS

The electrostatic strategy as suggested by Momany [26], Kollman and co-workers [29], and others [56,71,72] provides atomic charges able to represent the electrostatic distribution of molecules properly. It must be stressed that when these charges are going to be used in MM, MD, or MC studies, the relevant magnitude is not the set of charges by themselves, but the Coulombic MEP generated by these charges. This view of the problem leads to the conclusion that when semiempirical methods are used to derive charges, the scaling procedure must be performed so that Coulombic semiempirical MEPs will be able to reproduce ab initio QM MEPs. The results presented in this paper demonstrate that the standard scaling procedure [50,51,56], which fits semiempirical and ab initio charges, is equivalent to the scaling of the MEP in some cases. Nevertheless, in other cases, especially for charged molecules, or when we are interested in the description of the electrostatic interactions in regions close to the nuclei, it is not valid.

Comparison between semiempirical and ab initio MEPs suggests that, in the range considered, which covers most of the areas of interest in MM, MC, and MD studies, the suitability of semiempirical methods is approximately constant, irrespective of the distance from the nuclei.

Nevertheless, a slight decrease in the scaling coefficients occurs when inner layers are considered, which confirms our previous suggestions [43,44,67] about the tendency of semiempirical methods to overestimate the electron density in inner regions.

The uniform behavior in the space of semiempirical MEPs allowed us to define two sets of scaling coefficients, corresponding to inner and outer MEPs. Thus for neutral molecules the following values were obtained: MNDO 1.26 (inner), 1.28 (outer), AM1 1.25 (inner), 1.30 (outer), and PM3 1.21 (inner), 1.27 (outer). For cations a scaling coefficient of 1.00 is recommended for both inner and outer MEPs. Finally, for anions a scaling coefficient of 0.96 is recommended for inner MEPs, while 1.00 should be used to scale outer MEPs.

Comparison between the three methods indicates that MNDO is the best strategy to obtain MEPs, at least within the deorthogonalization strategy used here [50,51]. The scaling of the MNDO MEP according to the values noted in the above paragraph gives deviations between the MNDO Coulombic MEP and the QM high-level ab initio MEPs of 1.9 kcal/mol (neutral), 2.0 kcal/mol (cations), and 2.9 kcal/mol (anions). Bearing in mind that the minimum possible differences (those obtained when ab initio QM and Coulombic MEPs were fitted) are 1.1 kcal/mol (neutral), 1.1 kcal/mol (cations), and 2.6 kcal/mol (anions), the general quality of the MNDO-derived semiempirical MEPs is clearly demonstrated.

This new scaling procedure should be much more powerful than the simple scaling of charges in two important cases: (1) to obtain semiempirically corrrected charges for charged molecules whose size makes ab initio calculations difficult; and (2) to correct the MEP in inner regions in selfconsistent reaction field calculations of solvated systems.

## ACKNOWLEDGEMENTS

## REFERENCES

1 Scrocco, E. and Tomasi, J., Topics Curr. Chem., 42 (1973) 95.
2 Kollman, P.A., McKelvey, J., Johansson, A. and Rothenberg, S., J. Am. Chem. Soc., 97 (1975) 955.
3 Kollman, P.A. and Rothenberg, S., J. Am. Chem. Soc., 99 (1977) 1333.
4 Orozco, M., Canela, E.I. and Franco, R., J. Org. Chem., 55 (1990) 2630.
5 Goldblum, A. and Pullman, B., Theor. Chim. Acta, 47 (1978) 345.
6 Kollman, P., J. Am. Chem. Soc., 99 (1977) 4875.
7 Tomasi, J. In Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum, New York, 1981, pp. 257–294.
8 Náray-Szabó, G., Acta. Phys. Acad. Sci. Hung., 51 (1981) 65.
9 Scrocco, E. and Tomasi. J., Adv. Quantum. Chem., 11 (1978) 115.
10 Politzer, P., Landry, S.J. and Wärnheim, T., J. Phys. Chem., 86 (1982) 4767.
11 Sjoberg, P. and Politzer, P., J. Phys. Chem., 94 (1990) 3959.
12 Murray, J.S. and Politzer, P., J. Org. Chem., 56 (1991) 6715.
13 Hayes, D.M. and Kollman, P.A., J. Am. Chem. Soc., 98 (1976) 3335.
14 Weinstein, H., Osman, R., Green, J.P. and Topiol, S., In Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum, New York, 1981, pp. 309–334.

15 Weiner, P.K., Langridge, R., Blaney, J.M., Schaefer, R. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 79 (1982) 3754.

16 Orozco, M., Canela, E.I. and Franco, R., Mol. Pharmacol., 35 (1989) 257.

17 Orozco, M., Canela, E.I. and Franco, R., Eur. J. Biochem., 188 (1990) 155.

18 Weiner, P.K., Langridge, R., Blaney, J.M., Schaefer, R. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 79 (1982) 3754.

19 Hayes, D.M. and Kollman, P.A., J Am. Chem. Soc., 98 (1976) 7811.

20 Miertus, S., Scrocco, E. and Tomasi, J., Chem. Phys., 55 (1981) 117.

21 De Oliveira Neto, M., J. Comp. Chem., 7 (1986) 617.

22 Spackman, M.A. and Stewart, R.F., In Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum, New York, 1981, pp. 407–423.

23 Moss, G. and Coppens, P., In Politzer, P. and Truhlar, D.G. (Eds.) Chemical Applications of Atomic and Molecular Electrostatic Potentials, Plenum, New York, 1981, pp. 427–443.

24 Politzer, P., J. Chem. Phys., 72 (1980) 3027.

25 Gadre, S.R. and Bendale, R.D., Chem. Phys. Lett., 130 (1986) 515.

26 Momany, F.A., J. Phys. Chem., 82 (1978) 592.

27 Williams, D.E. and Yan, J.M., Adv. At. Mol. Phys., 23 (1988) 87.

28 Tomasi, J., Bonaccorsi, R. and Cammi, R., In Maksic, Z. (Ed.) Theoretical Models in Chemical Bonding, Vol III, Springer, Berlin, 1991, p. 231.

29 Singh, U.C. and Kollman, P.A., J. Comp. Chem., 5 (1984) 129.

30 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D., J. Comp. Chem., 7 (1986) 230.

31 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.

32 Laughton, C.A. and Neidle, S., J. Mol. Biol., 223 (1992) 519.

33 Lane, A.N., Jenkins, T.C., Brown, D.J.S. and Brown, T., Biochem. J., 279 (1991) 269.

34 Hoops, S.C., Anderson, K.W. and Merz, K., J. Am. Chem. Soc., 113 (1991) 8262.

35 Kollman, P.A., Annu. Rev. Phys. Chem., 38 (1987) 303 and references cited therein.

36 Petrongolo, C., Gazz. Chim. Ital., 108 (1978) 445.

37 Culberson, J.C. and Zerner, M.C., Chem. Phys. Lett., 122 (1985) 436.

38 Edwards, W.D. and Weinstein, H., Chem. Phys. Lett., 56 (1978) 582.

39 Mó, O. and Yáñez, M., Theor. Chim. Acta., 47 (1978) 263.

40 Ramos, M.J. and Yáñez, M., J. Chem. Soc., Faraday Trans 2, 79 (1983) 1389.

41 Giessner-Prettre, C. and Pullman, A., Theor. Chim. Acta, 37 (1975) 335.

42 Orozco, M. and Luque, F.J., Chem Phys. Lett., 160 (1989) 305.

43 Luque, F.J., Illas, F. and Orozco, M., J. Comp. Chem., 11 (1990) 416.

44 Luque, F.J. and Orozco, M., Chem. Phys. Lett., 168 (1990) 269.

45 Luque, F.J., Orozco, M., Illas, F. and Rubio, J., J. Am. Chem. Soc., 113 (1991) 5203.

46 Price, S.L., Andrews, J.S., Murray, C.W. and Amos, R.D., J. Am. Chem. Soc., 114 (1992) 8268.

47 Dewar, M.J.S. and Thiel, W., J. Am. Chem. Soc., 99 (1977) 4899.

48 Dewar, M.J.S., Zoebisch, E.G., Horsley, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.

49 Stewart, J.J.P., J. Comp. Chem., 10 (1989) 209.

50 Orozco, M. and Luque, F.J., J. Comp. Chem., 11 (1990) 909.

51 Besler, B.H., Merz, K.M. and Kollman, P.A., J. Comp. Chem., 11 (1990) 431.

52 Ferenczy, G.G., Reynolds, C.A. and Richards, C.W.G., J. Comp. Chem., 11 (1990) 341.

53 Cummins, P.L. and Gready, J.E., Chem. Phys. Lett., 174 (1990) 355.

54 Bonati, L., Consentino, U., Fraschini, E., Moro, G. and Pitea, D., J. Comp. Chem., 13 (1992) 842.

55 Merz, K.M., J. Comp. Chem., 13 (1992) 749.

56 Cox, S.R. and Williams, D.E., J. Comp. Chem., 2 (1981) 304.

57 Orozco, M. and Luque, F.J., J. Mol. Struct. (Theochem), 254 (1992) 31.

58 Orozco, M. and Luque, F.J., J. Comput.-Aided Mol. Design, 4 (1992) 411.

59 Negre, M.J., Orozco, M. and Luque, F.J., Chem. Phys. Lett., 196 (1992) 27.

60 Gao, J., Luque, F.J. and Orozco, M., J. Chem. Phys., 98 (1993) 2975.

61  Connolly, M., QCPE Bull., 1 (1981) 75.

62  Luque, F.J. and Orozco, M., unpublished version of MOPAC/CMS, University of Barcelona, 1992.

63  Stewart, J.J.P., QCPE Bull., 3 (1983) 10.

64  Olivella, S., QCPE Bull., 4 (1984) 10. Extended by S. Olivella and J.M. Bofill, 1990.

65  Dupuis, M., Rys, J. and King, H.F., QCPE Bull., 58 (1977) 17. Extended by J. Rubio and F. Illas, 1986.

66  Luque, F.J. and Orozco, M., unpublished version of the MOPETE/MOPEFIT programs, University of Barcelona 1992.

67  Alemán, C., Luque, F.J. and Orozco, M., J. Comput. Chem., 14 (1993) 799.

68  Cramer, G.J. and Truhlar, D.G., J. Comp. Chem., 13 (1992) 1089.

69  Juranic, I., Rzepa, H.S. and Yi, M., J. Chem. Soc., Perkin Trans., 2, 1990 (1990) 887.

70  Rzepa, H.S. and Yi, M., J. Chem. Soc., Perkin Trans., 2, 1991 (1991) 513.

71  Chirlian, L.E. and Francl. M.M., J. Comp. Chem., 8 (1987) 894.

72  Williams, D.E., J. Comp. Chem., 9 (1988) 745.