Theor Chem Acc (2003) 110: 446–459
DOI 10.1007/s00214-003-0501-z

Theoretical
Chemistry Accounts

Regular article

Transferable group contributions for a variety of chemical phenomena and compounds

Giuliano Alagona[1], **Silvio Campanile**[1], **Caterina Ghio**[1], **Alessandro Giolitti**[2], **Susanna Monti**[1]

[1] CNR-IPCF, Istituto per i Processi Chimico-Fisici, Molecular Modeling Lab, Via Moruzzi 1, 56124 Pisa, Italy
[2] Menarini Ricerche S.p.A., Via Sette Santi 3, 50131 Florence, Italy

Received: 4 September 2002 / Accepted: 15 March 2003 / Published online: 25 November 2003

Abstract. An approximate method for calculating molecular electrostatic potential (MEP) maps and atomic point charge models for large molecules in a reduced computational time is proposed and tested for two widely used basis sets (STO-3G and 6–31G*). The method avoids the molecular orbital calculation of the whole system by expressing its first order electronic density matrix in terms of transferable localized orbitals (TLO), previously determined on model molecules, via a localization process followed by the cutting of the tails, and stored in two databases (one for each basis set). For systems with a canonic electronic structure TLO are made of a single vector, involving either two nuclei (to describe the covalent bond between those atoms) or one nucleus (to describe lone pairs and inner shells). Conversely, delocalized π systems require many-center TLO, formed by a suitable number of vectors. Density functions of large chemical compounds can thus be built up automatically from a code that recognizes which fragments are contained in the system of interest, extracts them from the chosen database, reorders the atoms consistently with the pertinent TLO and places them in the correct position and orientation on the relevant atoms. A great number of chemical groups were parameterized and the efficiency of the method was evaluated on different systems, including aliphatic hydrocarbons. Numerical calculations on several molecules revealed that this approximation brought no significant loss of accuracy with respect to the corresponding Hartree-Fock (HF) values for the examined properties. Although the method is specifically designed to produce approximate wavefunctions, the point charge models obtained by fitting the corresponding MEP represent a viable alternative when *ab initio* HF calculations are not affordable, and can be used in connection with any popular force field.

Keywords: Transferable localized orbitals – Chemical fragments – Electrostatic potential maps – Atomic charges

From the Proceedings of the 28th Congreso de Químicos Teóricos de Expresión Latina (QUITEL 2002)

Correspondence to: G. Alagona
e-mail: g.alagona@ipcf.cnr.it

Introduction

Combinatorial chemistry is one of the important new methodologies developed by academics and researchers in the pharmaceutical, agrochemical and biotechnological industries to reduce the time and costs associated with producing effective, marketable and competitive new drugs. Nowadays scientists use combinatorial chemistry to create large populations of molecules, or libraries, that can be screened efficiently all together. The probability of finding novel molecules of prominent therapeutic and commercial value can increase by producing larger and more various compound libraries. In order to identify ligand candidates among those enormous numbers of diverse molecules, a fast, accurate and efficient method to analyze their properties and quantify the contributions from their interactions is necessary.

Classical mechanics and dynamics simulations can successfully be used to calculate molecular properties and to describe the geometry and inter- and intramolecular energies of many different molecules such as proteins, DNA, RNA and their complexes with substrates, inhibitors, and new drugs. Those molecules are very large and still beyond the capability of full quantum mechanical methods. Force field methods ignore the electronic motions and calculate the energy of a system as a function of the nuclear positions only. Transferability is a key feature of a force field. A set of parameters developed and tested on a small sample can be used to study a series of related molecules. Force fields are

empirical; there is no unique and correct form for a force field: a wide variety of functional forms can be employed as well as many different approaches to derive the parameters. A force field is generally designed to predict certain properties and is parameterized accordingly. The simplest analytical form defining a force field is a four component function of intra- and intermolecular forces that is used by AMBER [1, 2, 3], OPLS [4], DREIDING [5], and many other force fields.

There are several important factors that determine the success of molecular mechanical models. One of the most crucial is the description of electrostatic interactions that are the dominant long-range contribution to intermolecular forces. A simple and very powerful way to characterize the electrostatic properties of electronic charge distributions for qualitative interpretation of structure and reactivity is through atom-centered partial charge models. Unfortunately, partial atomic charges cannot be unambiguously determined because they are not experimentally observable. As a result, a large number of methods have been proposed for calculating atomic charges. Some of these are purely empirical whereas many others make use of quantum mechanics. Earlier methods based on fitting charges to quantum mechanical molecular electrostatic potential (MEP) used to determine point charges [6, 7, 8, 9, 10] located on meaningful chemical entities (bonds, lone pairs) besides atoms. Subsequently they became popular for producing atomic charges (ESP) [11, 12] and can be [13, 14, 15, 16, 17, 18] regarded as a good choice for simple harmonic models [19, 20]. The restrained electrostatic potential fit model (RESP) [21] represents a refinement of the ESP method. RESP makes use of a hyperbolic restraint function on non-hydrogen atoms during the least squares fitting of the partial charges to the MEP. This restraint has the effect of reducing the charges on some atoms, and the strength of short-range intermolecular interactions is more accurately predicted [21, 22]. The RESP methodology brings an improvement over the ESP derived charges by lowering the magnitude of the charges and reducing their conformational dependency.

On the other hand, the partial charge description can break down in the case of aliphatic compounds or when a hydrocarbon chain is present in an otherwise polar system. In this respect, most of the methods in use to determine partial charges fail for *n*-alkanes [23].

Although partial charges are a fictitious concept, MEP is an experimental observable [24] and can be determined from a wavefunction. The important role played by MEP in computational chemistry is pointed out by its many applications in reactivity [25, 26], solvation [27], complementarity and similarity [28, 29]. The study of MEP is very useful in rationalizing the interaction between molecules at large separations and for molecular recognition processes, such as drug-receptor and enzyme-substrate interactions. The recognition process, which precedes formation of the complex, is believed to occur when the molecules involved are at a relatively large separation. MEP calculations with quantum chemical methods are well established, but the result is of an approximate nature. The approximation arises from the level of electron correlation and from the electronic density used for the calculation, whose quality depends on the basis set chosen. Fortunately, for most applications qualitative knowledge of MEP is sufficient, and thus the basic features of MEP can be computed with unrefined electronic densities. This is especially important in the case of aliphatic compounds that are difficult to describe with partial charge models [23].

Clearly, *ab initio* calculations cannot routinely be performed on large molecules even with small basis sets, such as STO-3G. It is therefore convenient to build up the molecules from fragments of a suitable size whose parameters have been previously determined and stored in a database. We do not claim that this is a new idea, since several groups, including the Scrocco-Tomasi one [30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40] where we began our scientific activity, have devoted much effort to cover large molecules with quantum mechanical methods. Christoffersen, in particular, used floating spherical Gaussian orbitals not restricted to lying on the nuclei [41, 42, 43, 44]. Alternative methods, making use of semi-empirical Hamiltonians and fractional models, have also been proposed [45].

Definition of transferable localized orbitals

It has been shown [30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40] that it is possible to get a sensible representation of the whole MEP without calculating the molecular wavefunction in advance. The approximation consists of expressing the first order electronic density function in terms of transferable localized orbitals (TLO), instead of actual molecular orbitals. TLO are obtained from molecular orbitals via a localization process, followed by a cutting of the tails. Although for saturated organic molecules the canonical molecular orbitals can be localized satisfactorily by standard methods [46, 47, 48, 49, 50], as discussed in [35], in the present case we preferred the Boys' method in its second version [47], because it reproduces correctly the bond dipoles [31, 33, 34]. Any electron pair in a localized orbital λ may be described by a principal part χ_p and a secondary contribution χ_s. For a bond between two atoms, A and B, the principal portion χ_p is defined as the normalized contribution to the localized orbital from all the basis functions belonging to A and B, whereas the secondary part χ_s represents the contribution to λ from the basis functions centered on all the other atoms of the system:

$$\lambda = c_p\chi_p + c_s\chi_s \tag{1}$$

Equation 1 is also applicable to lone-pair electrons whose localized orbitals have a one-center character. Of course the core orbitals are strictly localized per se. Their secondary part (χ_s) is, on average, sufficiently small to be

considered as tails and thus can be neglected. The most common TLO involve two atomic centers and describe the bonds between those atoms; one-center TLO are defined for lone pairs and many-center TLO are used for delocalized π systems.

Approximate methods for treating very large molecules containing hundreds to thousands of atoms in a drastically reduced computational time have been developed. A common feature is the partition of the whole covalent system to be treated into fragments of various sizes and the use of localized orbitals. Examples of such methods include: the Fragment Self-Consistent Field (FSCF) model [51, 52, 53, 54] and the Fragment Molecular Orbital (FMO) model [55, 56, 57, 58, 59]. The first version of the FSCF method was developed at the Complete Neglect of Differential Overlap semi-empirical level. Subsequently another semiempirical version was developed at the Neglect of Diatomic Differential Overlap level in the MNDO [60], AM1 [61], and PM3 [62] parameterizations [63, 64, 65]. It was intensively tested and successfully applied to the study of electrostatic recognition mechanisms, distribution functions and molecular geometries [66, 67, 68]. In the FMO method, recently proposed, calculations were carried out at the Hartree-Fock (HF)/STO-3G and HF/6–31G* levels and many practical applications have been considered.

Interestingly, even atom-centered electron density fragments can be used to produce whole molecules [69, 70]. The use of transferable atom equivalents however requires a library of more than 7,000 atom types [70].

We are interested in the use of chemical groups expressed in terms of TLO in order to analyze the MEP of novel compounds to be tested as tentative drugs against those of known therapeutics and to compute atomic charge models avoiding the MO calculation of the whole system. The real challenge represented by drugs, apart their delivery and adsorption fate, is due to the large variety of functional groups they may contain, much wider than that present in proteins and nucleic acids.

Computational details

All quantum mechanical calculations were performed with the Gaussian98 [71] and Gamess_US [72] programs, while the TLO calculations were performed with the Molecular Building Blocks Method (MBBM) package, coupled with Jupiter [73, 74] (the code written to prepare the data required). The potential derived charges were computed with STO-3G and 6–31G* basis sets at points on a series of molecular surfaces constructed gradually increasing the van der Waals radii for the atoms as determined using the Merz-Kollman (MK) protocol [11, 12], both at the Hartree-Fock and TLO levels. The molecular conformations were obtained from model building techniques and minimized with the Sybyl program [75], using Gasteiger-Hückel charges [76, 77, 78] and Tripos force field. All calculations were performed on Alpha Compaq workstations at the Molecular Modeling Lab of Istituto per i Processi Chimico-Fisici.

The values of the LCAO coefficients for the TLO were determined using the STO-3G and the 6–31G* basis sets and the geometries of the fragments stored in the Sybyl database to build a molecule. In order to produce most of the chemical functional groups, a great number of bonds and atoms were parameterized; their types for common atoms (C, N, O, F, S, Cl) are reported in Table 1. Using these groups, it is possible to build all the chemical compounds, involving the aforementioned atoms, without triple bonds. As a first step, the study was restricted to a series of non-conjugated compounds to set up the procedure; subsequently, it was extended to molecules containing aromatic groups with and without heteroatoms and condensed ring systems (Fig. 1). To define and parameterize those fragments, σ and π electrons were treated separately. Two-electron local groups were adopted for σ subsystems, while the bonds between carbon and hydrogen atoms, and between nitrogen and hydrogen atoms, were left apart in order to allow substitutions onto the rings.

Table 1. Parameterized bond types[a]

C.3-C.3	O.2-C.2	N.3-C.2	S.3-C.3	lpO2
C.2-C.3	O.3-C.3	N.3-C.ar	S.3-C.2	lpO3
C.ar-C.3	O.3-C.2	N.3-C.3	S.3-C.ar	lpN3
C.2-C.2(d)	O.3-C.ar	N.3-H	S.3-H	lpF
C.2-C.2(s)	O.3-H	N.2-N.am	S.o-O.2	lpS3
C.2-C.ar	O.3-N.3	N.2-C.3	S.o-H	lpSo
C.3-H	F-C.3	N.2-C.2(d)	S.o2-C.2	lpCl
C.2-H	F-C.2	N.2-N.2	S.o2-C.3	
C.ar-H	F-C.ar	N.am-C.3	S.o2-C.ar	
	Cl-C.3	N.am-C.2	S.o2-O.3	
	Cl-C.2	N.am-C.ar	S.o2-N.3	
	Cl-C.ar	N.am-H	S.o2-H	
		N.pl3-C.3	S.o2-O.2(s)	
		N.pl3-C.2		
		N.pl3-C.ar		
		N.pl3-H		

[a] definition based on the SYBYL atoms types. s Single bond, d double bond, lp lone-pair

Systems considered, results and discussion

MEP comparisons exploiting MK charges

To test the validity of the TLO approximation and to evaluate the efficiency of the method, the electrostatic potential obtained from the MK charges, sometimes supplemented with the RESP ones, was compared with the approximate MEP. The MK charges, in fact, turned out to be more reliable than other common approaches, such as CHELP [79] and CHELPG [80], to obtain potential derived charges, even though they show an appreciable dependence on the orientation of the molecule compared to the CHELPG method [81].

To match two MEPs, a least squares fitting procedure based on the underlying partial charge models can be used [82]. The sum of the squared differences between two sets of MEP values is thus minimized while changing the mutual orientation of the molecules, because that method was developed to align any system. Since in this case the molecular structures coincide, the mutual orientation is generally unaffected. In that approach, the quality of the fit is evaluated on the basis of the root mean square deviation converted to a percentage of similarity [82].

Bulky molecules Numerical and graphical comparisons to *ab initio* values as obtained from minimal STO-3G and extended 6–31G* basis sets for some selected molecules (displayed in Fig. 2), chosen among several test cases not reported here, with a variety of chemical groups, are shown in Table 2, Fig. 3 and Fig. 4. As far

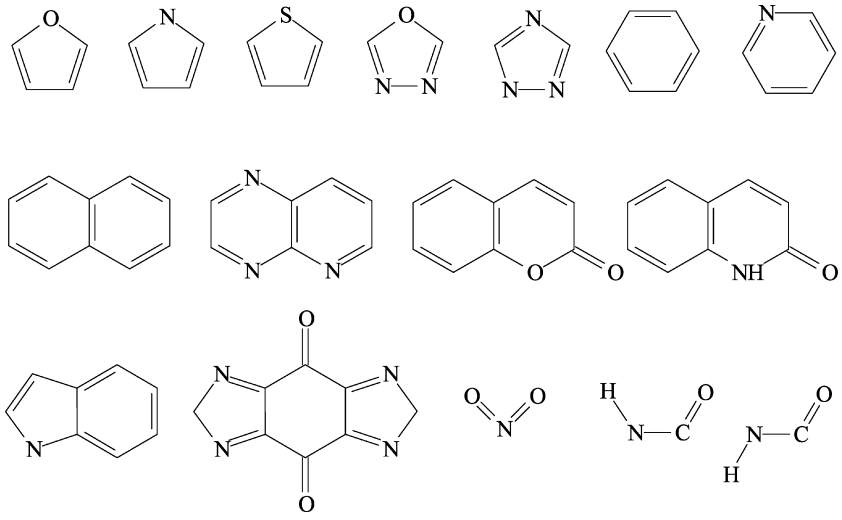

Fig. 1. Parameterized fragments

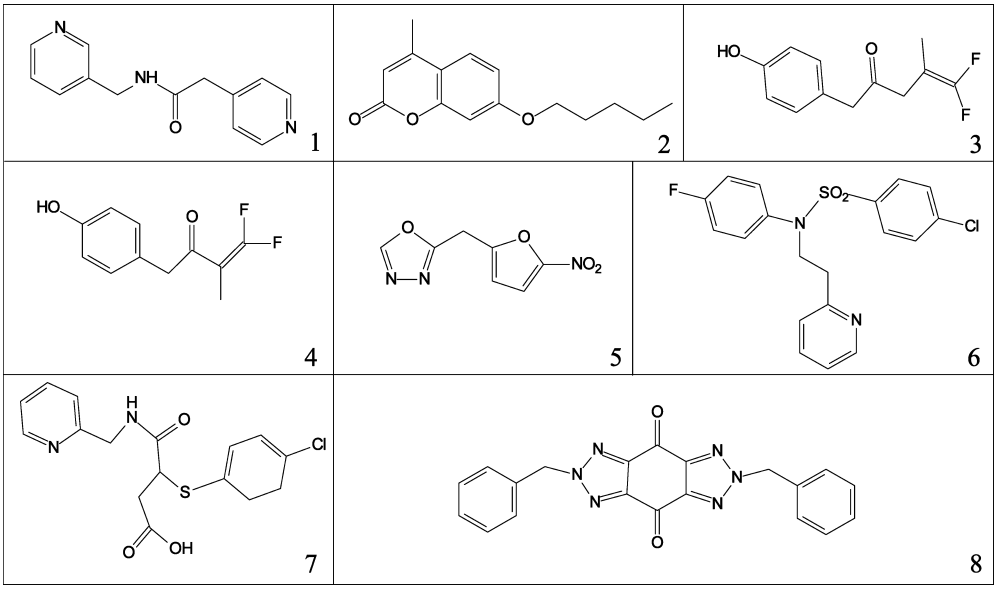

Fig. 2. The eight molecules considered for the molecular electrostatic potential (MEP) and dipole moment comparison

Table 2. Percentage of similarity obtained overlapping molecular electrostatic potential maps for the molecules shown in Fig. 2

Mol.	ESP		RESP	CM2[a]		CM1[a]
	TLO/STO-3G HF/STO–3G	TLO/6–31G* HF/6–31G*	TLO/STO–3G HF/STO–3G	AM1 HF/6–31G*	PM3 HF/6–31G*	PM3 HF/6–31G*
1	99.2	99.0	99.4	90.8	91.6	89.3
2	91.8	91.0	90.7	74.9	79.2	78.2
3	80.7	82.1	80.1	86.0	85.5	84.4
4	79.6	83.5	78.3	90.4	90.4	89.1
5	96.2	95.6	96.2	69.3	89.4	87.3
6	77.0	86.3	71.3	70.0	82.6	91.2
7	85.8	86.2	84.4	77.8	-[b]	-[b]
8	98.6	98.0	98.8	81.8	90.1	87.5

[a] [98]
[b] Point charge model not obtained, because the calculation did not converge

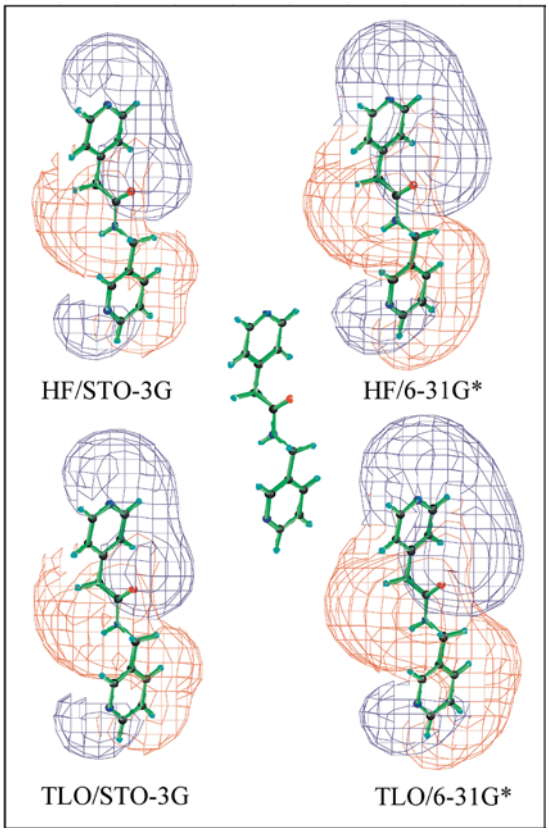

Fig. 3. MEP of molecule 1 produced by the Merz-Kollman partial charges (*blue* and *red* surfaces = −10 and + 10 kcal/mol, respectively) derived from the best fit to the transferable localized orbitals (TLO)/STO-3G, HF/STO-3G, TLO/6–31G* and HF/6–31G* MEP

as the MEP maps are concerned, a single isopotential surface drawn at ± 10 kcal/mol is less informative, though more legible than two overlaying surfaces (such as those for instance at ± 5 and ± 10 kcal/mol). However, from the comparison of the TLO three dimensional (3D) isopotential surfaces to the HF ones, it is evident that the qualitative features of the electrostatic potential are correctly reproduced by the approximate method, even though it is advisable to refer to their percentage of similarity to quantify it.

In Fig. 3 and Fig. 4, the 3D-color-coded isopotential surfaces (at ± 10 kcal/mol, with red = positive, blue = negative) for molecules 1 and 6 are shown. In Fig. 3. The TLO-MEP (lower part) and HF-MEP (upper part) for molecule 1 closely resemble each other: though both TLO descriptions slightly overestimate the corresponding MEP, the percentage of similarity is actually very high (99% for either basis set, i.e. STO-3G, left hand side, and 6–31G*, right hand side). This behavior is due to the fact that molecule 1 is mainly made of large fragments: two pyridine rings and a *trans* peptide bond group that reduce the level of approximation and, thus, the error made in describing the system. The main difference with respect to the HF description actually lies in the lack of tails. When a large fragment like pyridine or benzene is parameterized, however, molecular orbitals are localized, although the tails extending on the other atoms belonging to that fragment are maintained and thus they contribute to molecular properties.

In Fig. 4, even though the TLO surfaces of molecule 6 reproduce (with the usual slight overestimate) the MEP summary characteristics fairly well, the similarity with the HF maps is lower. This fact can be explained by considering that this molecule is made up of a small number of large fragments (a pyridine ring and two benzene rings) compared with the number of bonds and lone pairs present. In the TLO approximation the F lone pairs give rise to a large negative lobe, shown in the HF/6–31G* map, but notably reduced at the HF/STO-3G level. The effect of the Cl lone pairs is similar to the F lone pairs: a noticeable negative potential surface appears around the chlorine atom, which is reduced to a small cap at the HF/STO-3G level and seems to be absent in the HF/6–31G* map. This is, however, due to the fact that a potential of −10 kcal/mol represents a limiting value about Cl, and the use of a single isopotential surface prevents one from realizing that the HF/6–31G* potential is just few kilocalories/mole lower, in absolute value, than 10 kcal/mol, which is conversely accounted for by the percentage of similarity (87%). The similarity is worse at the STO-3G level (77%), because the TLO wide positive lobe, due to the cooperativity among ring and methylene group hydrogens, is much more extended than for the *ab initio* descriptions. On the other hand, we preferred to use a common contour value for all the maps and systems shown, to facilitate cross comparisons. Furthermore, a somewhat high MEP value (± 10 kcal/mol) was employed, since the TLO approximation seemingly gives more satisfactory results at large distances from the molecule (such as those reachable with ± 5 kcal/mol values) than at short distances.

In order to avoid artifacts linked to a single contour, MEP color-coded Connolly surfaces [83, 84] have also been attempted as an alternative view, but the various descriptions of the same molecule turned out to be almost indistinguishable from one another. Interestingly enough, resorting to color-code in turn for several limited ranges of MEP values, it was possible to visualize only the deepest regions of the MEP that, in the case of molecule 6 for instance, correspond to the contact surface about the SO_2 oxygens and the pyridine nitrogen. By gradually enlarging towards smaller potential values the window, the blue region expands and covers F, finally reaching Cl as well.

A quantitative estimate of the accuracy of the TLO aproximation can , however, be reached by taking into account the rms between the *ab initio* (V_{ref}) and TLO (V_{appr}) MEP values in each point of the Connolly surface, which has been determined as:

$$\text{rms} = \left[(1/N)\sum_{i=1,N}\left(V_{appr} - V_{ref}\right)^2\right]^{1/2} \qquad (2)$$

where N = number of surface points.

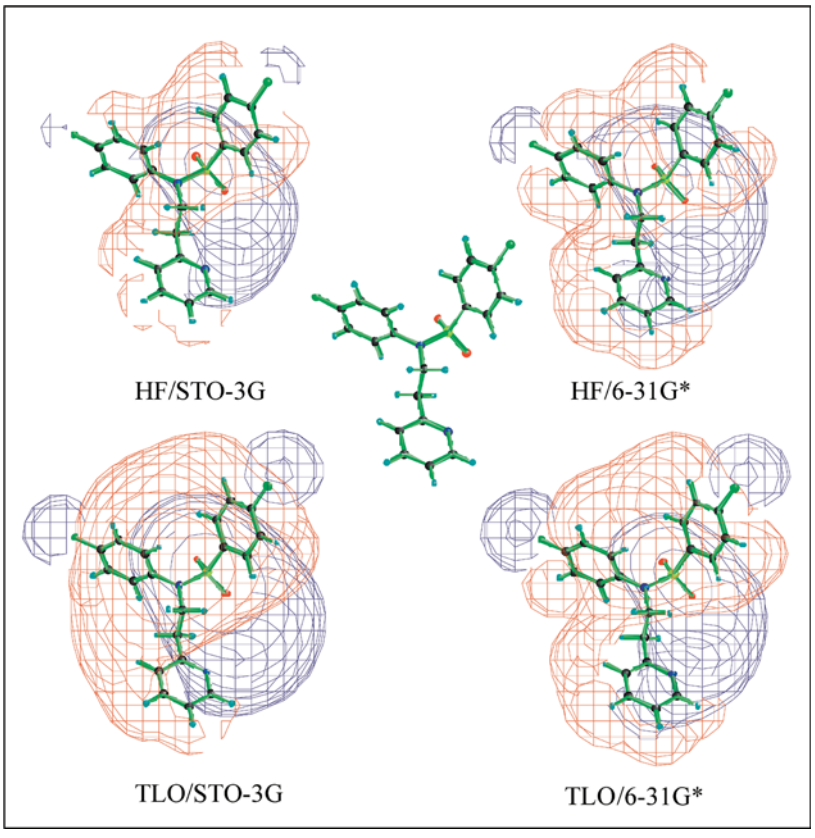

Fig. 4. MEP of molecule 6 produced by the Merz-Kollman partial charges (*blue* and *red* surfaces = −10 and +10 kcal/mol, respectively) derived from the best fit to the TLO/STO-3G, HF/STO-3G, TLO/6–31G* and HF/6–31G* MEP

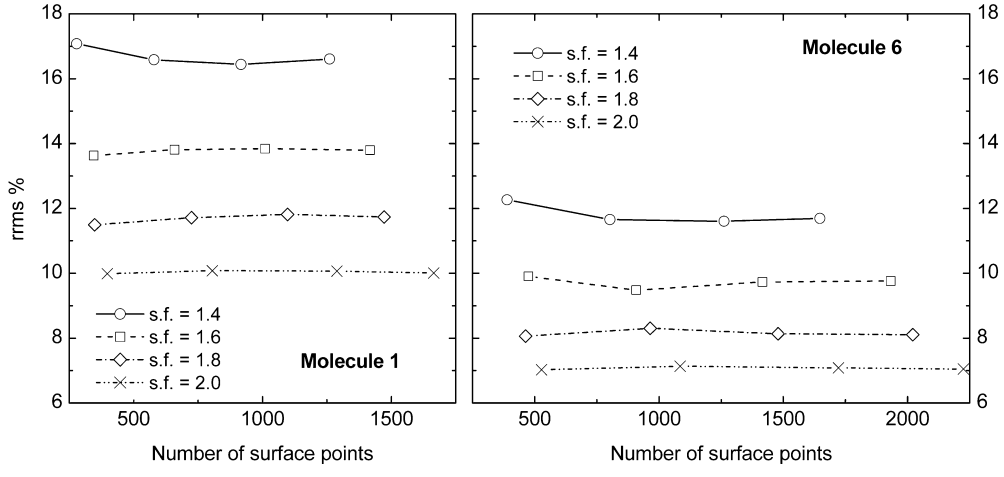

Fig. 5. Relative root mean square deviation (rrms in %) between the HF and TLO MEP at the 6–31G* level for molecules 1 and 6 (whose isocontour surfaces are shown in Fig. 3 and Fig. 4). MEP are calculated on four Connolly layers with van der Waals radii scale factors (*s.f.*) ranging from 1.4 to 2.0 and for an increasing number of surface points, which corresponds to densities of points per surface unit of 1, 2, 3 and 4 points.

The relevant indexes are defined as:

$$V_{\text{ex}} = \left[(1/N)\sum_{i=1,N}\left(V_{\text{ref}}\right)^2\right]^{1/2} \tag{3}$$

and

$$V_{\text{approx}} = \left[(1/N)\sum_{i=1,N}\left(V_{\text{appr}}\right)^2\right]^{1/2} \tag{4}$$

Therefore, the relative rms is defined as:

$$\text{rrms} = \text{rms}/V_{\text{ex}} \tag{5}$$

In Fig. 5 the rrms for the four layers, corresponding to the van der Waals radius of each atom scaled by incremental factors (s.f. = 1.4, 1.6, 1.8, and 2.0, respectively) are reported, for four different values of the density of points per surface unit ($d = 1, 2, 3, 4$, respectively). What appears evident from the trends is a better agreement between the *ab initio* and approximate pictures of the MEP at a longer distance, as expected. In contrast, there is no sensitivity advantage in increasing the density of points per surface unit. The accuracy for molecule 1 is seemingly somewhat worse than for molecule 6.

The percentage of similarity of MEP obtained from the TLO-RESP charges to that derived from the HF-RESP charges, also reported in Table 2, is slightly lower than in the ESP case. In Table 2, the results produced by the charges obtained from empirical-semi-empirical methods such as the CMx approach [85, 86] with respect to the HF/6–31G* MEP values are also shown for comparison. While in general the MEP obtained from the CMx charges (either AM1—not shown—or PM3) is much closer to the HF/6–31G* than to the STO-3G ones (also not shown), the CM2/PM3 charges do not represent a noticeable improvement with respect to the CM1/PM3 ones as far as MEP is concerned, which, at least in the case of molecule 6 among the values displayed, give a decidedly higher percentage of similarity than CM2/PM3.

Dipole moments (not shown) computed from ESP and RESP charges reproduced the quantum mechanical dipoles very accurately. Both ESP and RESP methods led to dipole moments in excellent agreement with each other despite the variations in the ESP and RESP partial charges. In contrast, dipole moments computed with the TLO methodology differed, in some cases, noticeably from the corresponding quantum mechanical values. Since the dipole moment is, by definition, very sensitive to even small charges located at large separations, it is more affected than MEP by the cutting of tails placed far away from the main component of each fragment. Consequently, the values obtained for this observable are less satisfactory.

A possible way to overcome this deficiency could be based on the intuitive idea exploited in [86] to obtain the class IV charge model (CM2), since, while the potential derived partial charges can be constrained in the MK scheme to reproduce the molecular dipole, it is however, difficult to supply a reliable value of the dipole moment, when its experimental value is not known and HF calculations cannot be afforded. Nevertheless, based on the results obtained with the CM1 and CM2 models for a large set of compounds, the MEPs do not remarkably ameliorate because of the better description of the molecular dipole moment in CM2.

Polypeptides Several polypeptides have also been employed for test calculations. In Fig. 6, the MEP values obtained from different atomic charge models for an α-helix conformation of TRP-ALA-TYR, generated using the SYBYL software, are compared with one another. The comparison between the AMBER and HF/6–31G* MEP values gives 92% of similarity, whereas at the HF/STO-3G level the match obtained is lower (85%). This is in accordance with the fact that AMBER charges were derived using the 6–31G* basis set because of the known overestimate by 10–20% of the dipole moment that produces the amount of polarization expected in aqueous solution.

The results obtained with the TLO/STO-3G approximation are not too far from those calculated using the AMBER force field charges (77%), but they are in any case closer to the HF/STO-3G (80%) MEP values. The charge sets obtained from electrostatic potential fittings can depend highly upon the basis set used to derive the wavefunction. They do not always improve when a large basis set is used. Sometimes it is possible to scale the results obtained from a calculation with a small basis set, or at a lower level of theory, in order to get results comparable to those from high level calculations. In this case, however, the MEP values computed using the TLO/6–31G* library are slightly more similar to the HF/6–31G* (82%) ones than at the STO-3G level.

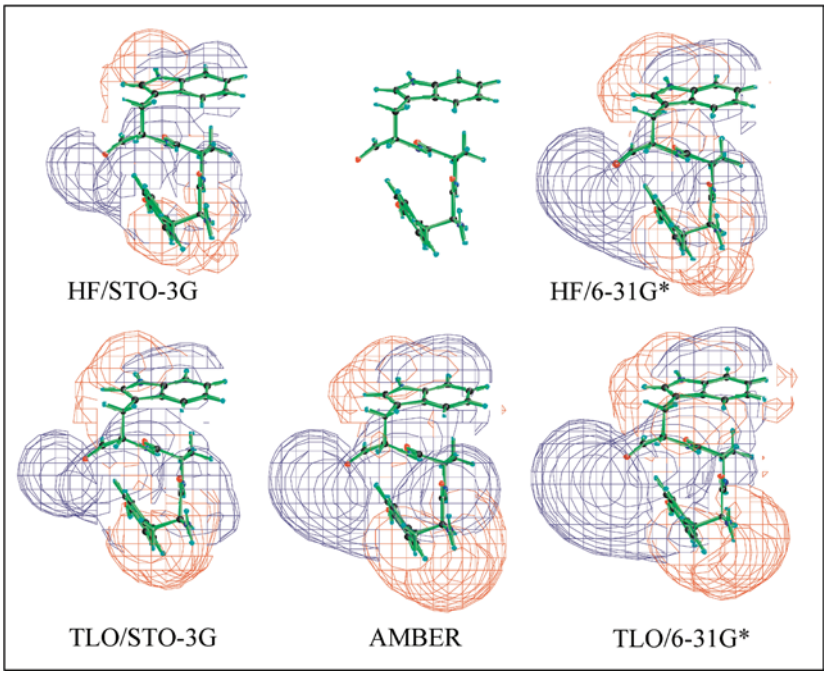

Fig. 6. MEP of TRP-ALA-TYR produced by the Merz-Kollman partial charges (*blue* and *red* surfaces = −10 and +10 kcal/mol, respectively) derived from the best fit to the TLO/STO-3G, HF/STO-3G, TLO/6–31G* and HF/6–31G* MEP and by the AMBER partial charges

Table 3. Percentage of similarity obtained overlapping TLO and HF molecular electrostatic potential maps for the molecules listed

Molecules	Similarity (%)	
	STO–3G	6–31G*
Methanol	94.4	94.4
Ethanol	92.9	93.3
1-Propanol(*gauche*)	92.6	91.1
1-Propanol(*anti*)	92.3	91.5
2-Propanol	91.1	91.3
Phenol	90.2	93.0
1,2-Ethanediol	91.4	92.4
Water	97.2	97.2
Methylformate	88.1	89.2
Methylacetate	83.1	83.6
β-Butyrolactone	87.7	85.2
γ-Butyrolactone	86.6	86.3
Formaldehyde	89.2	92.9
Ethanal(acetaldehyde)	90.2	91.8
Propanone(acetone)	89.6	91.1
2-Butanone	91.1	90.7
Cyclopentanone	92.0	91.6
Ethanoicacid (acetic acid)	85.1	83.1
Acetaceticacid	84.7	85.6
Malonicacid	84.1	83.0
Dimethylether	94.5	93.3
Anisole(methoxybenzene)	92.2	91.7
1,3-Dioxane	88.5	89.1
1,4-Dioxane	91.3	91.3
Oxethane	93.6	97.6
Tetrahydrofuran	86.1	92.2
Methylamine	94.5	97.9
Ethylamine	97.5	96.0
Dimethylamine	97.8	95.2
Ammonia	98.5	97.4
Piperidine	97.0	94.3
Ethanethiol	91.9	90.1
Methanethiol	94.1	93.7
Thiophene	99.5	99.6
Benzene	92.0	91.6

Test molecules of limited size In Table 3 the percentages of similarity between the TLO and HF MEP values are reported both at the STO-3G and 6–31G* levels for a series of test molecules belonging to several chemical families. The lowest percentage found is about 83%, though the average values are well above 90% both at the STO-3G and 6–31G* levels.

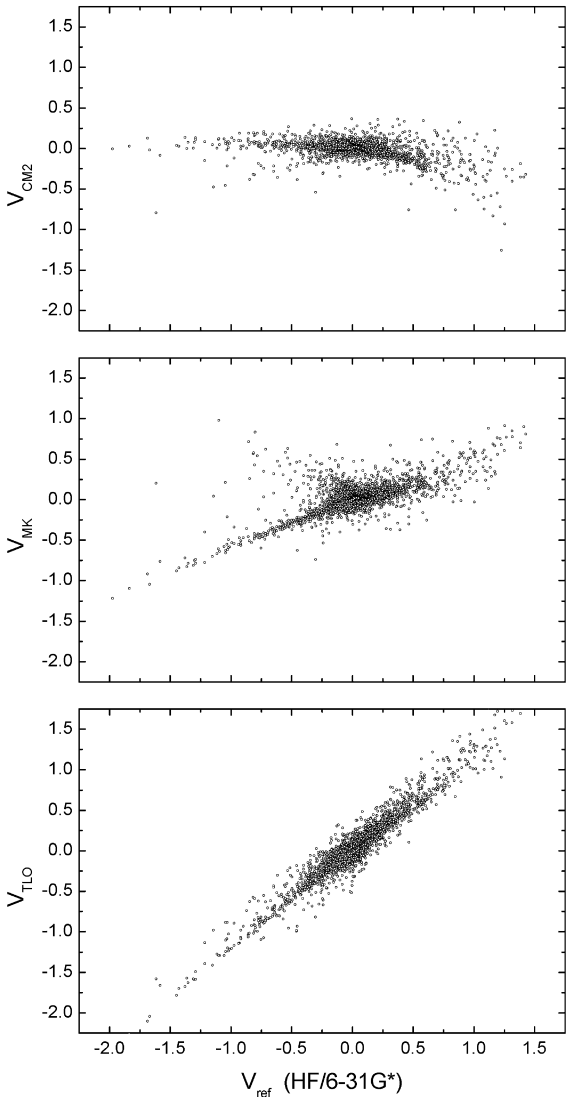

Fig. 7. Correlation between the HF/6–31G* MEP values (V_{ref}) of *n*-pentane along a grid of points (see text) and approximate MEP values produced by: (*lower part*) TLO/6–31G*, (*middle part*) Merz-Kollman (*MK*) and (*upper part*) CM2 partial charges. Values are in kcal/mol

TLO versus HF densities

Actually, a comparison carried out using partial charge models is reliable as long as the point charges are able to reproduce the system features. Every time an *ab initio* QM calculation is affordable, MK charges represent an excellent electrostatic model, with the exception only of small apolar compounds where there can be large errors, because of the difficulty in reproducing a feeble potential with a limited set of charges. A more adequate and proper comparison between MEP is obtainable using the TLO and HF densities, provided that a way to quantify the degree of similarity without uncertainty is available.

In order to validate the procedure used for this purpose, the TLO performance along the series of *n*-alkanes up to hexane was compared to the HF description. In Fig. 7 (lower part), the linear correlation between MEP values computed at the HF/6–31G* and TLO/6–31G* levels on a regular 3D grid of points around pentane (33,298 points with the exclusion of the closest ones, i.e. falling below 1.9 Å from H and 2.9 Å from C, radii used by Williams [23]) is displayed. As expected, the TLO MEP values slightly overestimate the HF ones (the slope is 1.14, but the regression coefficient is satisfactory ($r = 0.957$)). In Fig. 7, the linear correlation between the MEP values produced by the MK (middle part) or CM2 (upper part) models with respect to the HF/6–31G* ones, for the same system and on the same points, are also shown. It is apparent that atom-centered point

charge models fail to correctly reproduce the MEP values, although to different extents (MK: slope = 0.396, $r = 0.693$; CM2: slope = −0.138, $r = -0.434$).

In order to evaluate the reliability of the MEP values computed without actually resorting to their graphical representation, the rms between the HF/6–31G* MEP values (V_{ref}) and the approximate ones (V_{appr}) using either the TLO density or the partial charges (MK or CM2) have been determined on the grid points (excluding those whose separation is below the given radii).

In Fig. 8 (lower part), the correlation lines for the first six *n*-alkanes between the afore-mentioned indexes are shown, together with the corresponding rrms (upper part). The TLO MEP indexes (solid circles) turn out to be fairly well correlated ($r = 0.9986$) to the HF/6–31G* ones, with a sufficiently low rrms. In contrast, apart from methane, which is described very well by MK charges due to its spherical symmetry (rrms < 0.2), the atom-centered models turn out to be dismaying, as already noted [23]. The CM2 charges, though internally consistent ($r = 0.9722$), produce an rrms∼0.8 for methane and > 1 for ethane through hexane). Both ESP and RESP MK charges for ethane are extremely low and fail to reproduce its MEP (rrms > 1), giving MEP values very close to zero everywhere; furthermore, this failure sharply deteriorates the regression coefficient. The situation is slightly better for propane to hexane (rrms∼0.7).

When a number of small molecules with a variety of functional groups, listed in Table 3, is considered, the results decidedly favor the MK charges, as expected.

The correlation lines between the indexes, reported in Fig. 9, show an almost perfect trend for the MK description ($r = 0.9997$, slope = 1.02), even in the presence of non-polar molecules such as cyclohexane and 3,5-dimethylhexane, while *n*-alkanes have been left aside on purpose. Also, the CM2 charges behave very accurately ($r = 0.9809$, slope = 0.94), whereas the TLO description is slightly less satisfactory because of the limited size of the systems considered. Nonetheless, the MEP maps obtained are satisfactory for most applications, as can be derived from a perusal of Table 3.

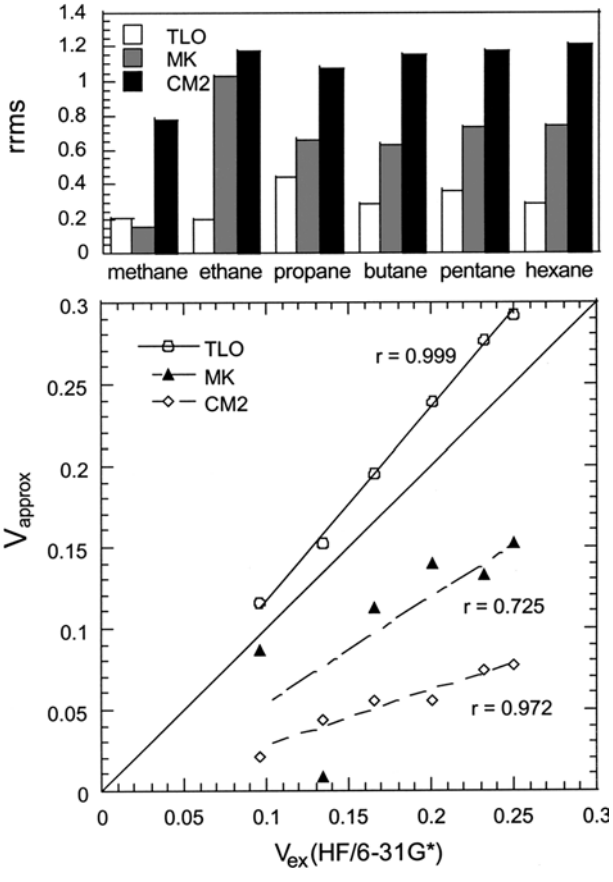

Fig. 8. In the *lower part* is shown the correlation between indexes (see text, values in kcal/mol) V_{ex}, related to the HF/6–31G* MEP values of *n*-alkanes along a grid of points, and analogous indexes V_{approx}, related to approximate MEP values produced by: TLO/6–31G*, MK and CM2 partial charges. The regression coefficients, *r*, are also shown. Relative root mean square deviations of the approximate MEP values produced by: TLO/6–31G*, Merz-Kollman and CM2 partial charges (see legend) with respect to V_{ex} are shown in the *upper part*

Energy comparisons

As a test concerning the applicability of this inexpensive method to get a useful first-order description of interaction energies and conformational curves, some comparisons of interaction energies, stabilization energies and conformational profiles of small molecules are reported. The systems chosen can be treated at a very high level of theory. A large number of theoretical studies have been reported in the literature thus far, which can be useful for comparison. Of course, critical systems are presented, because when dealing with large complex

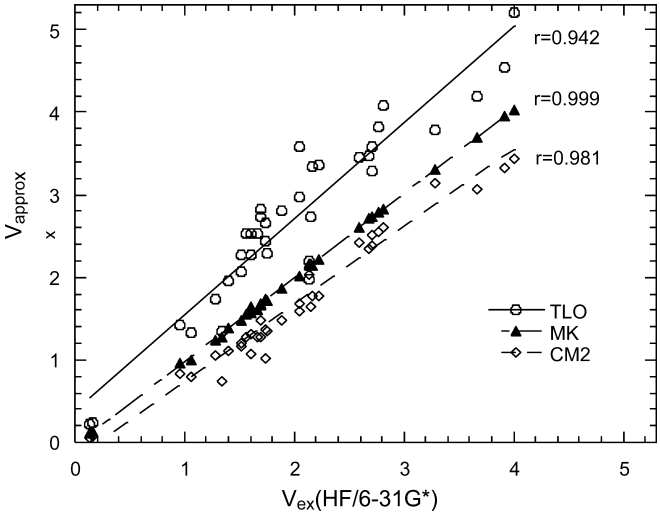

Fig. 9. Correlation between indexes (see text, values in kcal/mol) V_{ex}, related to the HF/6–31G* MEP values of the molecules listed in Table 3 along a grid of points, and analogous indexes V_{approx}, related to approximate MEP values produced by: TLO/6–31G*, MK and CM2 partial charges. The regression coefficients, *r*, are also shown

systems, apart from the difficulty of obtaining reliable reference values, a better performance of the method can be obtained, due to error compensations.

Interaction energies The calculations were carried out on three adducts belonging to different families: a classic hydrogen-bonded system such as the water dimer, a weakly hydrogen-bonded system such as the methane-water adduct and a van der Waals system such as the methane dimer.

The TLO interaction energies for variable separations at fixed mutual orientations of the two partners were compared to the *ab initio* results obtained at the SCF and MP2 levels with STO-3G and 6–31G* basis sets on internal geometries optimized in the isolated partners, using the Tripos force field and the Gasteiger-Hückel charges. Intermolecular interactions in hydrogen-bonded systems are an admixture of van der Waals and electrostatic effects. For the methane-water complex the starting conformation was chosen with methane oriented with one of its hydrogens pointing toward the oxygen of water. A linear configuration of the water dimer was selected as the starting conformation. An energy scan was performed on the methane dimer as a function of the C-C distance for a face to face relative orientation of the two molecules, which is considered to be the most stable one [87]. By the term face we mean the face of the regular tetrahedron with the C nucleus at the centre of the tetrahedron and an H atom at each vertex. Trends of the interaction energies and adduct starting conformations are displayed in Fig. 10. The counterpoise (CP) correction to the basis set superposition error (BSSE) was also introduced for the 6–31G* basis set, both at the HF and MP2 levels. What is apparent at first sight are the analogous behaviors of the TLO, CP-corrected MP2 and HF curves for the water dimer, even though the TLO/STO-3G interaction energies represent an upper boundary and the TLO/6–31G* a lower boundary. In the water-methane adduct, the TLO descriptions produce interaction energies close to the CP-corrected HF ones that are slightly less favorable than the CP-corrected MP2 interaction energies, in turn less favorable than the uncorrected ones. The HF/STO-3G description favors shorter equilibrium distances than the other methods, both for the water dimer (in analogy to AMBER) and the water-methane adduct, while the TLO equilibrium separations are closer to CP-corrected HF and MP2 values.

In 1992, Feller reported an exhaustive theoretical study of the water dimer [88]. Experimental estimates of the interaction energy place the global minimum of the water dimer at -5.4 ± 0.7 kcal/mol with an O-O distance of $\sim$2.98 Å. The complete basis set limit for the SCF interaction energy was estimated to be -3.55 kcal/mol, which became -5.1 kcal/mol at the optimal O-O distance of $\sim$2.91 Å, adding the correlation contributions. Our CP-corrected MP2 calculations predicting an interaction energy of -5.1 kcal/mol at an equilibrium distance $R_{OO} \sim 3.0$ Å are in good agreement with those theoretical results, as well as the TLO/6–31G* description that however produces a binding energy of $\sim$6.6 kcal/mol at $R_{OO} = 3.0$ Å. In contrast, the TLO/STO-3G description is less satisfactory.

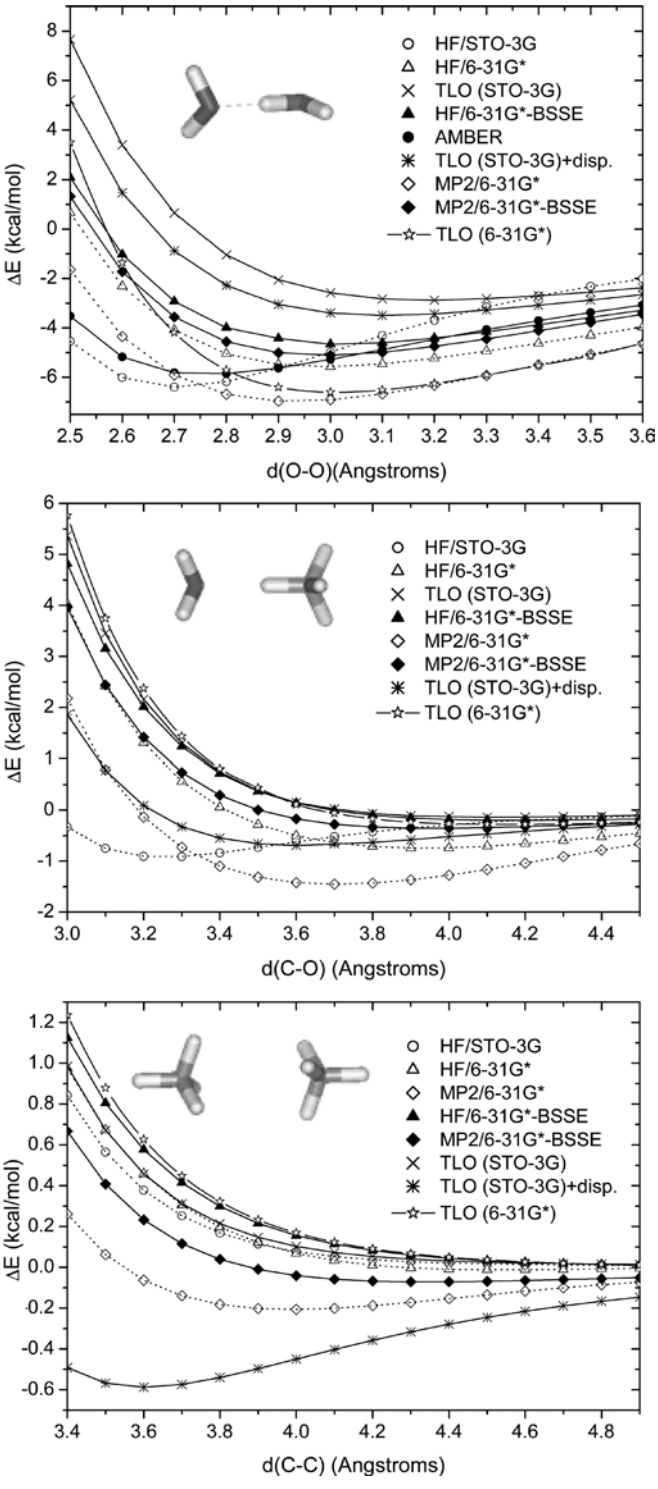

Fig. 10. TLO/STO-3G and TLO/6–31G* interaction energies along the approaching path for the water dimer (*top*), the methane-water adduct (*middle*) and the methane dimer (*bottom*) as compared to the HF or MP2/6–31G* ones with and without the inclusion of counterpoise corrections to BSSE and to the HF/STO-3G ones. TLO/STO-3G values with the inclusion of dispersion corrections are also shown

There have been a few studies of the methane-water adduct. Scheiner et al. [89] reported BSSE-corrected MP2/aug-cc-pVDZ and MP2/6–311+G** binding energies of 0.43 kcal/mol and 0.35 kcal/mol; Novoa et al. [90] reported a BSSE-corrected MP2/aug-cc-pVQZ(-g,-f) binding energy of 0.42 kcal/mol; the binding energy obtained here at the CP-corrected MP2/6–31G* level is about 0.36 kcal/mol, whereas the TLO/6–31G* value is ~0.28 kcal/mol (~0.14 kcal/mol for TLO/STO-3G), and intermediate with respect to the CP-corrected HF/6–31G* value (~0.20 kcal/mol).

The methane dimer is bound at the MP2 level of theory, but not at the HF and TLO levels. The weakest intermolecular interaction is found at the MP2/6–31G* CP-corrected level with a potential well of ~0.07 kcal/mol and an equilibrium distance R_{CC}~4.3 Å. As CH_4 does not possess a permanent dipole moment, the intermolecular attraction is dominated by the dispersion energy term. Whilst the repulsion can be calculated reliably even at the HF level, dispersion forces require post-HF treatments of electron correlation. The dispersion interaction has its origin in molecular polarization and electron correlation [91, 92]. Since the calculated molecular polarizability depends crucially on the basis set used, to improve the quality of the computed dispersion energies, very large basis sets containing a large number of polarization and diffuse functions were developed. However post-HF calculations such as MP2 with very large basis sets on molecules made of a great number of atoms are computationally very demanding and, for the time being, unfeasible.

The van der Waals (vdW) energy is modeled empirically in molecular mechanics force fields such as AMBER. In these methods the vdW energy is determined solely by the positions of nuclei and calculated using a Lennard-Jones potential with empirical parameters. Thus, an additional attraction energy term for each atom pair, separated by a distance R, $E_{vdW} = -C6/R^6$, can be added the interaction energies. The C6 coefficients were taken from the AMBER force field. The effect of this correction is shown for the TLO/STO-3G interaction energies. Although the C6 coefficients could be questionable to explore the potential energy surface of the CH_4 dimer, the calculated interaction energy shows a minimum for R_{CC}~3.6 Å with a potential well of ~0.59 kcal/mol, which is closer than the other values to the experimental estimates of 0.33–0.46 kcal/mol based on an isotropic averaged potential [93, 94, 95]. Of course, the agreement with the experimental estimates is even better correcting the TLO/6–31G* results. For the methane-water adduct a similar behavior can be observed: the equilibrium distance, R_{OC}, is shortened by ~0.5 Å and the interaction energy minimum becomes deeper (~0.70 kcal/mol).

Adding the empirical dispersion term to the TLO/STO-3G interaction energy in the case of the water dimer has a favorable effect as well: equilibrium distance and potential well come closer to the CP-corrected MP2 values.

Another interesting feature of the water dimer is the interaction energy connected to changes in the mutual orientation of the partners. In Fig. 11, the interaction energy profiles are displayed for the HF and TLO descriptions at the STO-3G and 6–31G* levels. While the HF profiles present two minima when the bridging H is collinear with the O lone-pairs, both TLO profiles show a wide minimum shifted towards the deepest minimum in the HF profile. Even though the fine details of HF profiles are not reproduced, the TLO trends are, however, satisfactory in the region of the minimum.

Conformational energies An accurate theoretical prediction of the barriers to internal rotation in polyatomic molecules requires a careful investigation of total energies at the relevant geometries and can be achieved by employing high-level methods and large basis sets. However the accuracy of quantum chemical calculations is severely limited by the computational resources, results with chemical accuracy lower than 1 kcal/mol being still limited to relatively small molecular systems. Therefore approximate methods and basis sets of manageable size are generally used. This is the reason why the TLO performance has also been examined in this field.

In Fig. 12, the barrier height to internal rotations and the conformational energy profiles for ethane (top), methanol (middle) and 1,2-ethanediol (bottom), computed employing TLO and HF methods at the STO-3G and 6–31G* levels, are shown. Bond lengths and bond angles were not optimized. For ethane the relative energies tend to increase with the size of the basis set. Even if the trend is correct, since the experimental value of the barrier is about 2.89 kcal/mol [96], the barrier height is underestimated by 0.4–0.6 kcal/mol, respectively, by TLO/6–31G* and TLO/STO-3G descriptions and overestimated at the HF level. For methanol the

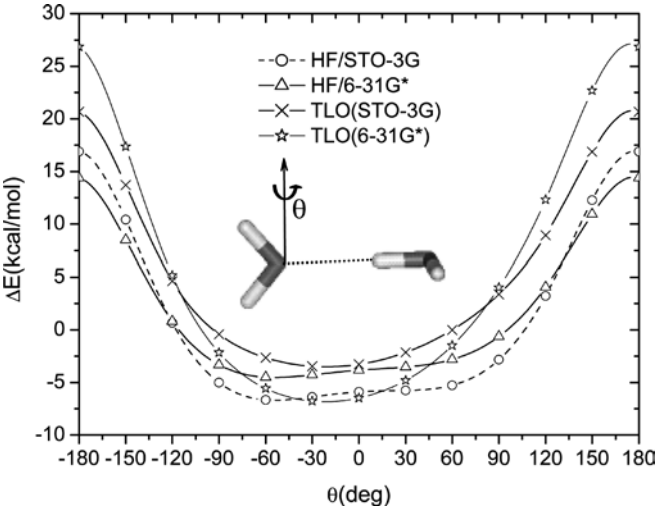

Fig. 11. Potential energy profiles for the rigid rotation about θ of the left hand side water molecule in the water dimer (the $\theta = 0°$ arrangement is displayed) at the HF and TLO levels, as described by STO-3G and 6–31G* basis sets

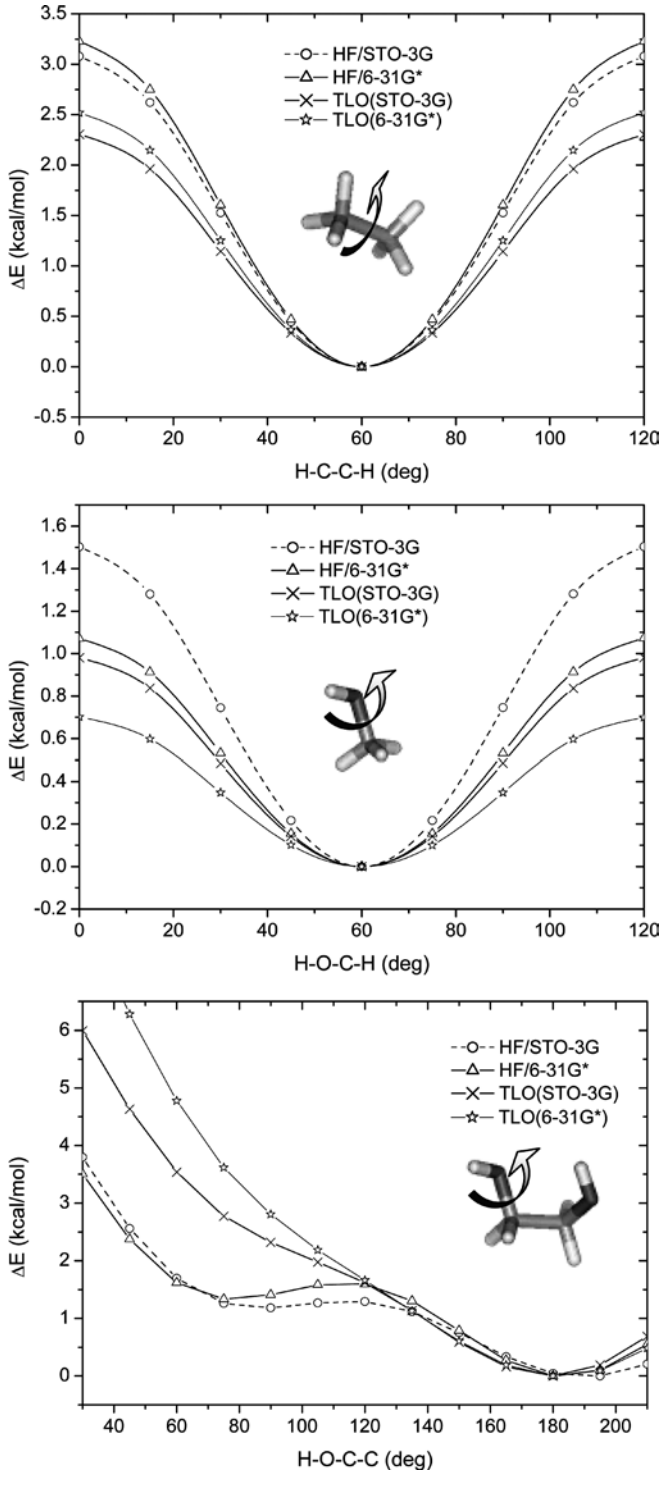

Fig. 12. Potential energy profiles at the HF and TLO levels, as described by STO-3G and 6–31G* basis sets for the rigid rotation of ethane about the C-C bond (*top*); methanol about the C-O bond (*middle*) and 1,2-ethanediol about the C1-O1 bond (*bottom*)

relative energies tend, in contrast, to decrease with the size of the basis set. The TLO energies are again more favorable than those obtained at the HF level (by ~0.4–0.5 kcal/mol), but the trend is correct. The 1,2-ethanediol system is different from the previous ones because some conformations, generated by changing the torsional angle about one of the C-O bonds, while keeping the position of the other hydrogen *gauche* with respect to the C-C bond, possess intramolecular hydrogen bonds. Intramolecular interactions can produce a polarization of some localized orbitals that is not accounted for by the TLO approximation. Thus the TLO curves, instead of having a maximum at about 120°, have an inflexion point and do not adequately represent the minimum about 70°. As a further development of this method, we are considering the possibility of introducing polarization effects perturbatively, as has been recently reported [97].

Conclusions

Since the concept of transferability is well established, we have implemented and extended to any kind of compound the TLO methodology, a technique proposed and tested on small molecules by Scrocco and coworkers many years ago. This approach allows an accurate description of the MEP of a molecule without calculating in advance its molecular wavefunction, thus reducing computational times. Such reduction becomes particularly significant for large molecules. Of course, the use of localized orbitals conflicts with the aromatic character of several biochemically and pharmacologically important compounds. Therefore, fragments extending over a consistently larger subset of atoms, with respect to common fragments made up of one-center or two-center units (i.e. inner shells, lone pairs and bond orbitals) have also been modeled with σ/π orbitals. The fragments, allowing the description of all the chemical groups including C, N, O, and S in various hybridizations, besides H, F and Cl, have been stored in two different libraries (STO-3G and 6–31G*). Two computer codes, one to recognize the fragments contained in the molecule of interest and reorder (and rename) its atoms consistently with the library, the other to assign the proper transferable contributions to each fragment and re-orient them according to the actual position taken in the molecule, have been set up.

Even though MEP can be determined directly from the TLO description, as it is advisable to do for non polar systems (specially for *n*-alkanes), this level of approximation can be successfully applied to determine point charge models for a large variety of systems, not only those of biological interest. MK potential-derived partial charges can be computed from the approximate wavefunctions, though a word of caution is necessary where dipole moments are also sought, because the deletion of tails can affect this observable. Partial charges from this source can be employed in combination with any empirical potential function, sharply expanding the force field capabilities, avoiding having to resort to erratic charges or to rely completely upon those already contained in force fields.

For systems with low non-electrostatic contributions, relative values of stabilization energies and interaction energies are sufficiently well reproduced already by the TLO approximate density or partial charges. The percentage of similarity with the *ab initio* MEP is in general higher for the TLO charges than for the CM2/PM3 ones, while the CM2/AM1 charges are slightly worse.

The quality of the calculation depends on the basis set and on the number and size of building blocks used to define the molecule. The error is larger when the molecule is divided into inadequately small fragments, while satisfactory results can be obtained using apt groups. The results obtained suggest that the TLO approximation is reliable for polypeptides, proteins and other systems containing also atoms of the second row of the periodic table.

Acknowledgements. S.C. is grateful to Menarini Ricerche SpA for a PhD grant.

References

1. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179
2. Wang J, Cieplak P, Kollman PA (2000) J Comput Chem 21:1049
3. Dixon RW, Kollman PA (1997) J Comput Chem 18:1632
4. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657
5. Mayo SL, Olafson BD, Goddard III WA (1990) J Phys Chem 94:8897
6. Bonaccorsi R, Petrongolo C, Scrocco E, Tomasi J (1971) Theor Chim Acta 20:331
7. Alagona G, Cimiraglia R, Scrocco E, Tomasi J (1972) Theor Chim Acta 25:103
8. Alagona G, Pullman A, Scrocco E, Tomasi J. (1973) Int J Peptide Protein Res 5:251
9. Pullman A, Alagona G, Tomasi J (1974) Theor Chim Acta 33:87
10. Bonaccorsi R, Scrocco E, Tomasi J (1977) J Am Chem Soc 99:4546
11. Singh UC, Kollman PA (1984) J Comput Chem 5:129
12. Besler BH, Merz Jr KM, Kollman PA (1973) J Comput Chem 11:431
13. Bonaccorsi R, Scrocco E, Tomasi J (1970) J Chem Phys 52:5270
14. Scrocco E, Tomasi J (1973) Top Curr Chem 43:95
15. Scrocco E, Tomasi J (1978) Adv Quant Chem 11:116
16. Politzer P, Truhlar D (eds) (1981) Chemical applications of atomic and molecular electrostatic potentials. Plenum, New York
17. Tomasi J (1982) Electrostatic molecular potential model and its application to the study of molecular aggregation. In: Ratajczak H, Orville-Thomas WJ (eds) Molecular Interactions. Vol. 3, p. 119. Wiley, New York
18. Náray-Szabó G (1979) Int J Quantum Chem 16:265
19. Hobza P, Kabelac M, Sponer J, Mejzlik P, Vondrasek J (1997) J Comput Chem 18:1136
20. Gundertofte K, Liljefors T, Norrby P-O, Petterson IA (1996) J Comput Chem 17:429
21. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) J Am Chem Soc 115:9620
22. Woods RJ, Chappelle R (2000) J Mol Struct (THEOCHEM) 527:149
23. Williams DE (1994) J Comput Chem 15:719
24. Espinosa E, Lecomte C, Ghermani NE, Devémy J, Rohmer M-M, Bérnard M, Molins E (1996) J Am Chem Soc 118:2501
25. Ghio C, Tomasi J (1973) Theor Chim Acta 30:151
26. Náray-Szabó G, Ferenczy G (1995) Chem Rev 95:829
27. Warshel A, Aquist J (1991) Ann Rev Biophys Chem 29:267
28. Sharp KA, Honig B, Harvey SC (1990) Biochemistry 29:340
29. Kenny PW (1994) J Chem Soc Perkin Trans 2:199
30. Bonaccorsi R, Scrocco E, Tomasi J (1976) J Am Chem Soc 98:4049
31. Ghio C, Scrocco E; Tomasi J (1978) Theor Chim Acta 50:117
32. Agresti A, Bonaccorsi R, Tomasi J (1979) Theor Chim Acta 53:215
33. Ghio C, Scrocco E, Tomasi J (1980) Theor Chim Acta 56:61
34. Ghio C, Scrocco E, Tomasi J (1980) Theor Chim Acta 56:75
35. Bonaccorsi R, Ghio C, Scrocco E, Tomasi J (1980) Israel J Chem 19:109
36. Bonaccorsi R, Ghio C, Tomasi J (1984) Int J Quantum Chem 26:637
37. Alagona G, Bonaccorsi R, Ghio C, Tomasi J (1986) J Mol Struct (THEOCHEM) 135:39
38. Ghio C, Tomasi J, Weill J, Sillion B (1986) J Mol Struct (THEOCHEM) 135:299
39. Alagona G, Bonaccorsi R, Ghio C, Montagnani R, Tomasi J (1988) Pure Appl Chem 60:231
40. Alagona G, Ghio C, Igual J, Tomasi J (1989) J Am Chem Soc 111:3417
41. Christoffersen RE (1972) *Ab initio* calculations on large molecules. In: Advances in quantum chemistry, vol. 6. AP, New York, p 333 and references quoted therein
42. Genson DW, Christoffersen RE (1973) J Am Chem Soc 95:362
43. Shipman LL, Christoffersen RE (1973) J Am Chem Soc 95:4733
44. Maggiora GM, Christoffersen RE (1976) J Am Chem Soc 98:8325
45. Orozco M, Luque FJ (1990) J Comput-Aided Mol Des 4:411
46. Foster JM, Boys SF (1960) Rev Mod Phys 32:300
47. Boys SF (1966) Localized Orbitals and localized adjustment functions. In: Lowdin PO (ed) Quantum theory of atoms, molecules, and the solid state. Academic, New York, p 253
48. Edminston C, Ruedenberg K (1965) J Chem Phys 43:97
49. von Niessen W (1972) Theor Chim Acta 21:9
50. Smits GF, Altona C (1985) Theor Chim Acta 67:461
51. Náray-Szabó G, Surján PR (1983) Chem Phys Lett 96:499
52. Náray-Szabó G (1984) Croat Chem Acta 57:901
53. Ferenczy GG, Rivail JL, Surján PR, Náray-Szabó G (1992) J Comput Chem 13:830
54. Náray-Szabó G, Ferenczy GG (1992) J Mol Struct (THEOCHEM) 261:55
55. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 312:319
56. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Chem Phys Lett 313:701
57. Nakano T, Kaminuma T, Sato T, Akiyama Y, Uebayasi M, Kitaura K (2000) Chem Phys Lett 318:614
58. Kitaura K, Sugiki S, Nakano T, Komeiji Y, Uebayasi M (2001) Chem Phys Lett 336:163
59. Nakano T, Kaminuma T, Sato T, Fukuzawa K, Akiyama Y, Uebayasi M, Kitaura K (2002) Chem Phys Lett 351:475
60. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899
61. Dewar MJS, Zoebish EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902
62. Stewart JJP (1989) J Comput Chem 10:221
63. Tóth G, Náray-Szabó G (1994) J Chem Phys 100:3742
64. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902
65. Stewart JJP (1989) J Comput Chem 10:209
66. Náray-Szabó G, Tóth G, Ferenczy GG, Csonka G (1994) Int J Quantum Chem QBS 21:227
67. Tóth G, Náray-Szabó G, Ferenczy GG, Csonka G (1997) J Mol Struct (THEOCHEM) 398:129
68. Ferenczy GG, Kádas K (1999) J Mol Struct (THEOCHEM) 463:175
69. Bader RFW, Becker P (1988) Chem Phys Lett 148:452

70. Breneman CM, Rhem M (1997) J Comput Chem 18:182
71. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery Jr JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98 (revision A.6). Gaussian, Pittsburgh, Pa.
72. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) J Comput Chem 14:1347
73. Alagona G, Ghio C, Monti S (2001) MBBM. Menarini Ricerche, Pomezia, Italy
74. Campanile S, Monti S, Cacelli I (2002) Jupiter. Menarini Ricerche, Pomezia, Italy
75. SYBYL (2000) Molecular Modeling Software, Version 6.7 (October 2000). TRIPOS Associates, St. Louis, Mo.
76. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219
77. Gasteiger J, Saller H (1985) Angew Chem, Int Edn Engl 24:687
78. Hückel E (1932) Z Phys 76:628
79. Chirlian LE, Francl MM (1987) J Comput Chem 6:894
80. Breneman CM, Wiberg KB (1990) J Comput Chem 11:361
81. Sigfridsson E, Ryde U (1998) J Comput Chem 19:377
82. Kearsley SK, Smith GM. (1993) An alternate method for the alignment for molecular structures. SEAL, QCPE #634, Bloomington, IN, 47405 USA
83. Connolly M. (1981) Molecular Surface program MS, QCPE #429, Bloomington, IN, 47405 USA
84. Connolly ML (1983) Science 221:709
85. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) J Comput-Aided Mol Des 9:87
86. Li J, Zhu T, Cramer CJ, Truhlar DG (1998) J Phys Chem A 102:1820
87. Novoa JJ, Whangbo M–H, Williams JM (1991) J Chem Phys 94:4835
88. Feller D (1992) J Chem Phys 96:6104
89. Gu Y, Kar T, Scheiner S (1999) J Am Chem Soc 121:9411
90. Novoa JJ, Planas M, Rovira MC (1996) Chem Phys Lett 251:33
91. Hobza P, Selzle HL, Schlag EW (1990) J Chem Phys 93:5893
92. Stone AJ (1996) The theory of intermolecular forces. In: International series of monographs on chemistry, vol. 32. Clarendon, Oxford
93. Reid BP, O'Loughlin MJ, Sparks RK (1985) J Chem Phys 83:5656
94. Boughton CV, Miller RE, Watts RO (1985) Mol Phys 56:363
95. Hoinkis J, Ahlrichs R, Böhm HJ (1983) Int J Quantum Chem 23:821
96. Fantoni R, Van Helvroot K, Knippers W, Reuss J (1986) Chem Phys 110:1
97. Luque FJ, Orozco M (1998) J Comput Chem 19:866
98. Hawkins GD, Giesen DJ, Lynch GC, Chambers CC, Rossi I, Storer JW, Li J, Zhu T, Rinaldi D, Liotard DA, Cramer CJ, Truhlar DG (1998) AMSOL—Version 6.5.3. Regents of the University of Minnesota

Note added in proof. In the meantime, C-C and C-N triple bonds have been included in both fragment libraries described in this article.