Comparison of methods for deriving atomic charges from the electrostatic potential and moments

Sigfridsson, Emma; Ryde, Ulf

Published in:
Journal of Computational Chemistry

DOI:
10.1002/(SICI)1096-987X(199803)19:4<377::AID-JCC1>3.0.CO;2-P

1998

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Sigfridsson, E., & Ryde, U. (1998). Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *Journal of Computational Chemistry*, *19*(4), 377-395. https://doi.org/10.1002/(SICI)1096-987X(199803)19:4<377::AID-JCC1>3.0.CO;2-P

Total number of authors:
2

Creative Commons License:
Unspecified

April 1, 2017

A comparison of methods for deriving atomic charges from the electrostatic potential and moments

Emma Sigfridsson and Ulf Ryde

Department of Theoretical Chemistry,

Lund University

Chemical Centre, P.O.B 124

S-221 00 Lund, Sweden

Tel: 46-46-2224502

Fax: 46-46-2224543

Email: teoulf@garm.teokem.lu.se

Abstract

Four methods for deriving partial atomic charges from the quantum chemical electrostatic potential (CHELP, CHELPG, Merz-Kollman, and RESP) have been compared and critically evaluated. It is shown the charges strongly depend on how and where the potential points are selected. Two alternative methods are suggested to avoid the arbitrariness in the point-selection schemes and Van der Waals exclusion radii: CHELP-BOW, which also estimate the charges from the electrostatic potential, but with potential points that are Boltzmann-weighted after their occurrence in actual simulations using the energy function of the program in which the charges will be used, and CHELMO, which estimates the charges directly from the electrostatic multipole moments. Different criteria for the quality of the charges are discussed. The CHELMO method gives the best multipole moments for small and medium-sized polar systems, whereas the CHELP-BOW charges reproduce best the total interaction energy in actual simulations. Among the standard methods, the Merz-Kollman charges give the best moments and potentials, but they show an appreciable dependence on the orientation of the molecule.

We have also examined the recent warning that charges derived by a least-squares fit to the electrostatic potential normally are not statistically valid. It is shown that no rank-deficiency problems are encountered for molecules with up to 84 atoms if the least-squares fit is performed using pseudoinverses calculated by singular value decomposition and if constraints are treated by elimination.

Key words:

atomic charges, copper complexes, electrostatic potential charges, molecular simulation, rank-deficiency problems.

Introduction

Molecular simulation methods have become an important technique in many areas of chemistry through the recent advent of effective and wide-spread software for molecular mechanics, molecular dynamics, and Monte Carlo simulations, and procedures for the estimation of free energy differences. In all such methods, a proper description of the electrostatic interactions within the simulated system is of key importance, since the electrostatics dominates the non-bonding energy between polar molecules. Electrostatic interactions can be treated at various levels of sophistication in classical simulation. However, in most implementations, especially in the multi-purpose biochemical simulation packages, the simplest possible solution is used, an atom-centred point-charge model where each atom (or groups of atoms) is assigned a (partial) electric charge and only the charge–charge Coulombic interaction is considered. Thus, most classical simulation methods need a point-charge parameterisation of the molecules of interest.

Unfortunately, atomic charges are not observables, i.e. they cannot unambiguously be determined by experiments or quantum chemical calculations. Therefore, a large number of methods have been suggested for the estimation of point charges [1]. Several groups have tried to derive the charges directly from experimental quantities, e.g. dipole moments, electrostatic potentials, or free energy differences [2-4]. Yet, relevant data is usually missing or too scarce to allow a determination of all charges in interesting molecules. Instead, most techniques derive atomic charges from quantum chemical calculations.

The simplest way to determine quantum chemical charges is the Mulliken population analysis. In this method, the charge is distributed according to the atomic orbital occupation and the overlap population between pairs of atoms is evenly divided between the two atoms without taking any account of differences in atom type, coefficients, electronegativity, etc. [5]. Although, Mulliken charges are known to strongly depend on the basis set [6] and to reproduce electrostatic moments poorly [1], they are still widely used due to their simplicity. Several other orbital-based methods have been devised to cure the problems of the Mulliken charges [1], e.g. Löwdin population analysis [7] and natural population analysis (NPA) [8].

Another quantum chemical approach is to divide the molecule into volumes occupied by each atom, and assign all electron density within each volume to the corresponding atom. In the atoms in molecules (AIM) method, the volumes are defined by the zero-flux surfaces of the electron density [9]. This is a very attractive approach, since it removes all arbitrariness in the division of the charge among the atoms. However, the procedure is computationally tedious and the resulting atomic charges reproduce the electrostatic potential rather poorly when they are used in a monopole approximation (as they are in most simulation packages) [10].

The most widely used methods to estimate atomic charges derive them from a least-squares fit to the electrostatic potential (ESP; or to the electrostatic field [11,12]) calculated in a large number of points around the molecule of interest. Examples of such potential-based methods are CHELP, CHELPG, and the Merz-Kollman scheme. They differ mainly in the choice of the points were the electrostatic potential is calculated. In the CHELP (charges from electrostatic potential) method [13], points are selected symmetrically on spherical shells around each atom (fourteen points on each shell). The Merz-Kollman scheme [14,15] uses instead points on nested Connolly surfaces, with a density of 1 point/Å, whereas in the CHELPG [16] method, points are selected on a regularly spaced cubic grid with over ten times higher point density than for the other two methods (0.3 Å between the grids). Recently, it has been suggested that the points should be sampled randomly around the molecule [17-19] to avoid any dependence on the choice of coordinate axes.

The electrostatic potential methods do not include in the fit potential points within the Van der Waals radii of the atoms, or points that too far from the molecule. Unfortunately, the exclusion limits, as well as the Van der Waals radii themselves, vary appreciably between the different methods. The Merz-Kollman scheme [14,15] sample points at 1.4, 1.6, 1.8, and 2.0 times the Van der Waals radius of the atoms. The CHELP method [13] sample points at 2.5, 3.5, 4.5, 5.5, and 6.5 Å from each atom, and excludes points within the Van der Waals radius of any atom. The CHELPG [16] method, includes points between 0 and 2.8 Å plus the Van der Waals radius. Several other schemes have been suggested [12,17,19], but normally no points within the Van der Waals radii or more than 3-7 Å from all atoms are used in the fit. Naturally, this arbitrariness in the choice of the potential points will affect the resulting charges.

In the electrostatic potential methods, charges are determined by a least-squares fit to the potential, usually under the constraint that the total charge of the molecule should be correct. Recently, it has been noted that such a fit does not always behave properly [18,19]. Even if the number of potential points is much greater than the number of atoms, it is not certain that the system is over-determined. On the contrary, linear dependence of the potential in nearby points may make the system under-determined, leading to nonsense charges. The method of choice to detect such problem is singular value decomposition (SVD) [18,19]. By this method, the rank of a matrix can be determined and it can be unambiguously concluded if a system is over-determined or not. Francl et al. have used this method to examine the least-squares matrix in a modification of the CHELP method and they observe for 16 typical medium-sized molecules that the matrix in all cases is rank-deficient [18]. Hinsen and Roux [19] have suggested an improved least-squares procedure, which uses pseudoinverses calculated by singular value decomposition. In this way, the problem with ill-conditioned matrices becomes less severe.

Another problem with the least-squares fit is that some charges might be less well-determined than others. This is especially pronounced for the charges of buried atoms, e.g. sp^3 carbons. It has been suggested that this may explain why potential-derived charges are less transferable than are Mulliken charges (similar chemical groups have widely differing charges in related molecules), why they depend so strongly on the conformation of the molecule, and why they often seem to be too large, giving rise to exaggerated intermolecular interactions. Bayly et al. have suggested that these problems may be prevented if all charges are restrained to zero using a hyperbolic penalty function. This method is called the restrained electrostatic potential model, RESP [20].

In this paper, we make a critical analysis of the four most popular potential-based point-charge methods, Merz-Kollman, CHELP, CHELPG, and RESP. It is shown that these methods may give widely different results and possible explanations for this are discussed. Two alternative methods for deriving atomic charges are suggested, which avoid the arbitrariness in the selection of potential points, and their performance is judged using a number of different quality criteria.

Methods

Computational details

If not otherwise stated, all calculations were performed with the 6-31G* basis set [21], except for copper and iron, for which the double-ζ basis sets (62111111/33111/311) of Schäfer et al. [22] were used, enhanced with diffuse *d*, *p* and *f* functions with exponents 0.174, 0.132, and 0.39 for copper, and 0.134915, 0.041843 (two *p* functions), 0.1244, and 1.339 for iron (called DZpdf). The electrostatic moments, the electrostatic potentials, CHELP, CHELPG, Merz-Kollman, Mulliken, NPA, and AIM charges were calculated with the quantum chemistry software Gaussian 94 [23] using the hybrid density functional method B3LYP [24], which has been shown to give charges of the same quality as correlated ab initio methods [25]. The default values for all methods were used, including point densities and atomic radii (2.0 Å was used for copper). The RESP fitting was performed with the Amber 4.1 software [26], using the standard two-stage method [20]. The potential points in the RESP method were sampled with the Merz-Kollman scheme in Gaussian 94 and the potentials were printed out using the undocumented IOp(6/33=2) option. In some cases, calculations were performed also with the ab initio Hartree-Fock (HF) method and with Møller-Plesset second-order perturbation theory (MP2). These calculations employed also the 3-21G [21] basis sets.

For the comparison of different charge models, the charges of 20 small molecules were calculated (see Tables VI and VII). The geometries of these molecules were optimised with the B3LYP method, except propane, for which an extended, model-built, conformation was used. In addition, a small model of the copper coordination sphere in the blue copper proteins was used: $Cu(NH_3)_2(SH)(SH_2)$ with either Cu(I) or Cu(II). These models are called the Cu(I) and Cu(II) complex, respectively. The complexes were optimised with the B3LYP method, as described by Ryde et al. [27].

In order to study the rank of least-squares matrix, 14 additional molecules were studied (Table XI). These were optimised at the SCF level with the 3-21G(*) basis set, as described by Francl et al. [5]. Moreover, the haem group of myoglobin was studied, including the axial histidine and dioxygen ligands (84 atoms). The coordinates of this system were taken from the PDB file 1MBO

[28], and the moments and potentials were calculated by the B3LYP method using the DZpdf / 6-31G* basis sets. All calculations were run on IBM RISC RS6000 workstations.

Charge fitting procedure

The fitting procedure described here is a slight modification of the one suggested by Hinsen and Roux [19]. They use this procedure to fit charges to electrostatic potential points, but since it is general it may be used for fitting the charges directly to electrostatic moments as well.

General linear constraints and restraints can be expressed as

$$w_k \sum_i B_{ki} q_i = w_k c_k \; ; \quad k = 1,2,\ldots,n_c \tag{1a}$$

or in matrix notation

$$\mathbf{wBq} = \mathbf{wc} \tag{1b}$$

where q_i are the partial charges, c_k the constraints or restraints and w_k are weight factors. All restraints or constraints that are linear in the charges can be expressed in this way, e.g. the total charge, electrostatic moments, electrostatic potential, or equality of certain charges due to symmetry. For example, for an electrostatic potential, c_k is the calculated potential and B_{ki} is the distance between the point where the potential is calculated and atom i, whereas for the z-component of the dipole moment, c_k is the dipole moment and B_{ki} is the z-coordinate of atom i.

Equation (1) is a linear equation for the partial charges q_i. It has a solution if and only if

$$\mathbf{B}'\mathbf{B}'^{+}\mathbf{cw} = \mathbf{cw} \tag{2}$$

where $\mathbf{B}' = \mathbf{wB}$ and $\mathbf{B}'^{+}$ is the pseudoinverse of the $\mathbf{B}'$ matrix. If this is fulfilled, the general solution is given by

$$\mathbf{q} = \mathbf{B}'^{+}\mathbf{c}\mathbf{w} + \left(\mathbf{1} - \mathbf{B}'^{+}\mathbf{B}'\right)\mathbf{q}' \tag{3}$$

where $\mathbf{q}'$ is an arbitrary vector.

In the following we use **B** and **c** for constraints and **A** and **f** for restraints. For the constraints, it is meaningless to use different weights and therefore **w** is omitted in these formulae (so $\mathbf{B}' = \mathbf{B}$). To find the optimal point charges for the restraints, we minimise the quantity

$$c^{2} = \left|\mathbf{wAq} - \mathbf{wf}\right|^{2}$$

Constraints are imposed in this equation by inserting equation (3):

$$c^{2} = \left|\mathbf{wAB}^{+}\mathbf{c} + \mathbf{wA}\left(\mathbf{1} - \mathbf{B}^{+}\mathbf{B}\right)\mathbf{q}' - \mathbf{wf}\right|^{2} \tag{5}$$

The general least-squares solution of this equation is

$$\mathbf{q}' = \mathbf{A}'^{+}\left(\mathbf{wf} - \mathbf{wAB}^{+}\mathbf{c}\right) + \left(\mathbf{1} - \mathbf{A}'^{+}\mathbf{A}'\right) \cdot \mathbf{y} \tag{6}$$

where

$$\mathbf{A}' = \mathbf{wA}\left(\mathbf{1} - \mathbf{B}^{+}\mathbf{B}\right) \tag{7}$$

and **y** is an arbitrary vector. To make sure that the smallest charges consistent with the fit are chosen if the **A** matrix is a rank deficient, **y** is set to be zero. This choice turns equation (6) into

$$\mathbf{q}' = \mathbf{A}'^{+}\left(\mathbf{wf} - \mathbf{wAB}^{+}\mathbf{c}\right) \tag{8}$$

and the final charges can be found by inserting $\mathbf{q}'$ in equation (3).

Pseudoinverses were calculated using the numerically stable singular value decomposition (SVD) method. This method gives also the singular values of the matrix, and the quotient of the smallest and largest singular values is the condition number that reflects the rank of the matrix. Following Francl et al. [18], we considered the least-squares matrix rank-deficient if the condition number is less than 10^{-5}. We used the singular value decomposition procedure DGESVD from the Lapack library.

The CHELMO method

To avoid the large number of adjustable parameters in the electrostatic potential methods, we have developed a new method for estimating partial charges, CHELMO (CHarges from ELectrostatic MOments). In this method, charges are fitted directly to the electrostatic multipole moments and no potentials are needed. The choice of the form of the multipole moments is important. It is mandatory that the moments are in the traceless form, but linearly dependent moments do not have to be removed (they are automatically removed by the fitting procedure). We used the real spherical harmonic form of the moments, since then the resulting charges are independent of the orientation of the molecule.

All Cartesian electrostatic moments are linear in the charges; for example, the components of the second Cartesian moment are given by $Q_{xx} = \sum_i q_i x_i x_i$, $Q_{xy} = \sum_i q_i x_i y_i$, and so on. Furthermore, the transformation of the Cartesian moment to the multipole moments in the traceless form or to the spherical harmonics form is a linear transformation (a matrix multiplication), so the multipole moments are still linear in the charges. Thus, the multipole moments can be used as constraints or restraints in the charge fitting procedure described above, and if the number of independent moments is greater than the number of independent atoms, no further information is needed to estimate the point charges. Thus, no potential points are needed and the charges will not depend on Van der Waals radii or on the way the potentials are sampled.

Only two choices have to be made in this fitting procedure: the division of the moments to be used as constraints or restraints, and the weight factor for the different restraints. It is customary to constrain the charges so that the total charge of the molecule is exactly reproduced when calculating atomic charges, and we will stick to this rule. Furthermore, we will argue that it is also important to constrain the charges to reproduce also the dipole moment, if possible, since the dipole moment determines the leading electrostatic energy term of uncharged, polar molecules. In addition, it is necessary, at least in molecular dynamics simulations, that equivalent atoms have the same charge (i.e. atoms that exchange rapidly on a molecular dynamics time scale, e.g. hydrogen atoms bound to the same heavy atom). In most electrostatic potential methods this is accomplished by simply averaging the charges *after* the fit; however, as was noted by Bayly et al. [20], this gives an appreciably worse result than if the charges are constrained to be equal within the fit. Therefore, we always constrain charges of equivalent atoms to be equal in the fit.

The higher moments (including the dipole moment if the number of independent atoms minus the number of non-vanishing elements of the dipole moment is less than two) are used as restraints. Moments up to hexadecapole moments are used since they are obtained in the output of ordinary quantum mechanical programs such as Gaussian. From a multipole expansion it is clear that it is more important to reproduce the lower moments than the higher moments. Moreover, the unit of the different moments, Cm^n, varies with the order of the moment n:. Therefore, the restraints must be weighted with a factor with the unit m^{-n}, and we suggest a factor of the form

$$w = \frac{1}{(r+x)^n} \tag{9}$$

where r is the radius of the molecule and x is a parameter. We used the distance from the centre of nuclear charge to the most distant atom in the molecule as an estimate of r, and 3 Å was found to be an appropriate choice for x (see below).

The CHELP-BOW method

As will be discussed below, the only reasonable way (in terms of the accuracy of the total electrostatic interaction energy) to select potential points for the fit of atomic charges, which are intended to be used in molecular dynamics simulations, is by weighting them after their occurrence in actual simulations. In order to illustrate the performance of such a technique, we here propose a simple implementation. In a forthcoming publication we will refine the method and optimise its performance.

Coordinates at which the electrostatic potential was calculated were selected randomly within 8 Å from any atom in the molecule. A point density of 2500 points/atom was used, since the charges of all atoms have been shown to be converged at this density [17, 29]. Points with a weight of less than 10^{-6} were directly discarded (see below).

In the fitting procedure, the total charge, the dipole moment (if the number of independent atoms minus the number of non-vanishing components of the dipole moment is more than one), the quadrupole moment (if the number of independent atoms minus the number of non-vanishing components of the dipole and quadrupole moments is more than two), and equal charges on equivalent atoms were used as constraints. The charges were restrained to fit the electrostatic potential, but the potentials were weighted by a Boltzmann factor calculated from the distance between the potential point and the atoms in the molecule:

$$w = e^{\frac{-E}{RT}} \tag{10}$$

where the E is the Lennard-Jones potential between the atoms and a probe molecule at the potential point:

$$E = \sum_{\text{atom } i} \begin{cases} A_i r_i^{-6} + B_i r_i^{-12}, & \text{if } r_i \le r_{i\min} \\ E_{i\min}, & \text{if } r_i \ge r_{i\min} \end{cases} \tag{11}$$

Here, r_i is the distance between the electrostatic potential point and atom i, and A_i and B_i are the Lennard-Jones constants obtained from the force-field of the simulation package to be used in the simulations. $E_{i\min}$ is the minimum energy of the Lennard-Jones potential, and it is the energy assumed at the distance $r_i = r_{i\min}$. Thus, only the repulsive part of the Lennard-Jones potential is considered.

In the present implementation of CHELP-BOW, we used the Amber 4.1 force field [30] and the TIP3P water parameters for the probe at the potential point. Thus,

$$A_i = -2e_{iw}\left(r_{iw}^{*}\right)^6 \tag{12}$$

$$B_i = \varepsilon_{iw}\left(r_{iw}^{*}\right)^{12} \tag{13}$$

$$\varepsilon_{iw} = \sqrt{\varepsilon_i \varepsilon_w} \tag{14}$$

$$r_{iw}^{*} = r_i^{*} + r_w^{*} - r_c \tag{15}$$

where ε_i and ε_w are the potential well depth of the atoms and of the water probe, respectively, r_i^{*} and r_w^{*} are the Van der Waals radii of the atoms and the probe, and r_c is a factor that corrects for the neglect of electrostatic interactions. For simplicity, the same ε_i and r_i^{*} were used for all atoms of the same element (except for polar and non-polar hydrogen); the values used are collected in Table I. Moreover, we used $r_c = 0.4$ Å, since this gives a correct O–O and O–H distance in water. With these parameters, we simply have $r_{i\mathrm{min}} = r_{iw}^{*}$ and $E_{i\mathrm{min}} = \varepsilon_{iw}$, .

Result and discussion

The problem with the potential-derived charges

The starting point of the present investigation was to obtain atomic charges of copper complexes for the use in classical simulations of blue copper proteins. Therefore, we started to test the three electrostatic potential methods available in the Gaussian software and the RESP method on the Cu(II) complex. As can be seen in Table II, the charges obtained with the four methods differ appreciably. The variation is most pronounced for the charges on the copper and nitrogen atoms, which vary between 0.11 and 0.38 *e* and between –0.13 and –0.56 *e* for the various methods, respectively. In classical simulations such a variation is unacceptable. The rank of the least-squares matrix indicated no deficiency, so this is clearly not the problem.

There is no default Van der Waals radius for copper in these methods; for the calculations in Table II we used a value of 2.0 Å, which is the default value in the Mulliken [31] software. Table III shows how the CHELPG charges vary when the copper radius is varied from 0.5 to 3.0 Å. For small radii (0.5–1 Å), the magnitude of all charges is high. Between 1.5 and 2.0 Å, the charges are rather constant, but at even larger radii, the charges begin to increase again. This show that the electrostatic potential charges strongly depend on the Van der Waals radii and it also gives a clue to why the charges in Table II vary so much.

The copper complex is approximately spherical with the copper ion at the centre. For a small copper radius, many electrostatic potential points are selected near the copper ion. There, the electrostatic potential is large and varies rapidly in space in a way that is hard to describe by a point-charge model. Since the electrostatic potential is high in magnitude, these points are very important for the fit, and in principle the charges will be determined by a small number of potentials close to the copper ion. Thus, the charges will strongly depend on the exact position of these potential points. When the copper radius is increased, these points are excluded and the fit becomes more stable. The increase in magnitude of the charges seen at an even larger radius is probably due to that the charges become less well determined, since the electrostatic potentials at large distances from the complex are determined mainly by the lowest one or two multipole moments.

The point selection scheme also differs among the three methods and this may explain some of the variation in the charges. Table IV shows the charges calculated with the CHELP, CHELPG, and Merz-Kollman methods using the same the Van der Waals exclusion radii (those of the CHELPG method and 2.0 Å for copper). A comparison of these charges with those in Table II directly shows that the charges do not change much when the Van der Waals radii are altered; the Merz-Kollman copper and nitrogen charges change 0.06-0.09 *e*, whereas all the other charges change by less than 0.03 *e*. Moreover, the charges hardly converge. This might indicate that the difference between the three methods lies mainly in the way the points are selected. However, the CHELPG method includes in the fit all points directly outside the Van der Waals envelop, whereas in the Merz-Kollman scheme, points are sampled between 1.4 and 2.0 times the Van der Waals radii. Thus, in principle the Merz-Kollman method uses radii that are 1.4 times larger than those of the other methods. In order to estimate the effect of this difference, we computed a third set of Merz-Kollman charges using the CHELPG radii divided by a factor of 1.4. With these radii, the charges almost coincide with those obtained by the CHELPG method (within 0.02 *e*).Thus, we can conclude that the main difference between the Merz-Kollman and the CHELPG method is the 1.4 times larger effective radii of the former method, whereas the sampling schemes are almost equivalent.

The reason for the difference between the CHELP and CHELPG is similar. In the CHELPG method, points are sampled on a dense cubic grid, which means that they in essence are sampled directly outside the Van der Waals radii. The CHELP points, on the other hand, are sampled on spherical shells, starting at 2.5 Å distance from each atom. This means that most points are sampled at this and larger distances, while only a few occasional points are sampled near the Van der Waals radii (points from the shells of nearby atoms). Thus, the effective Van der Waals radius of all atom in the CHELP method is 2.5 Å, and we again see that the charges are determined mainly by the location of potential points near to the molecule.

Criteria for the quality of atomic charges

A main problem in comparing different point-charge models is that there is no clear criterion for the quality of the charges. This is probably the reason why so many charge models have been suggested. Furthermore, different applications put different demands on the charges. For example, in molecular dynamics, the molecules move, so the charges must be able to describe the electrostatics properly in all accessible points in the phase space and they should also be invariant to changes in the internal coordinates of the molecule. On the other hand in some calculations of redox potentials or free energies, the molecule and the surroundings is fixed and it is then clear where other molecules actually are encountered; thus, it may suffice to describe the electrostatics well in these points.

Several different quality criteria for atomic charges have been used. A natural demand is that the charges are independent of the basis sets used and the orientation of the molecule during in the calculation [10,12,16]. Furthermore, they should be reasonably independent of the internal geometry of the molecule [20]. Other important issues are the electrostatic properties (the electrostatic potential, electrostatic field, dipole moment, and higher moments calculated from the atomic charges should reproduce those calculated from the wavefunction) [11-13,32,33] and chemical trends (introduction of substituents should affect the charge of adjacent atoms according to the electronegativity of the introduced group) [10,33]. Sometimes, charges have also been judged after how well they reproduce experimental quantities, e.g. dipole moments, solvation free energies, or conformational energies [34,35]. Unfortunately, no charge model is best in all respects. Topological and orbital-based methods give charges that are independent of the orientation of the molecule and usually give the best chemical properties. On the other hand, the potential-based methods give the best description of the electrostatic properties of the molecule [1,10]. Since our aim is to obtain charges to be used in molecular simulations, we have concentrated on the latter methods.

All potential-based charge models involve the implicit assumption that the key quality criterion is a minimal quadratic sum of the difference between the electrostatic potential calculated from the charges and from the wavefunction at the points used in the fit. Such a criterion is quite natural, since it minimises the error in the Coulomb interaction energy between the molecule and charges at the electrostatic potential points (the Coulomb energy is simply the product of the charge of a particle and the electrostatic potential at the position of the particle). However, it is not clear that a *square sum* of the deviations is the best quantity (apart from its nice properties in the fitting procedure). On the contrary, the square sum gives a high weight to large electrostatic potentials and large deviations (since the deviations are squared), which is not physically motivated. Instead, the total or average *absolute* deviation might be more just.

Furthermore, it is far from clear which electrostatic potential points should be included in the fit, a point that is illustrated by the fact that the three electrostatic potential methods tested in this investigation use different point sampling schemes. All potential-based methods use points only in the vicinity of the molecule. The motivation for this is that the electrostatic potentials are largest there and these values may therefore seem most important. Moreover, it has been shown that these points are important to get well-determined charges [17, 29]. However, the electrostatic potential of a charged or polar molecule decrease with the first or second power of the distance from the molecule, respectively, while the number of interactions increases with the *third* power of the distance. Therefore, for the total interaction energy it is more important to describe the potential well far from the molecule than close to it.

Another argument against the choice of only nearby points in the fit is that this gives unfairly great weights for these points. Logically, the only valid criterion for the choice of electrostatic potential points (in terms of the accuracy of the total electrostatic interaction energy) would be their occurrence in the actual simulations. Seen in this way, a Boltzmann average (based on the potential function of the program to be used) would be most appropriate for charges intended for a molecular dynamics simulation. Thus, points close to the molecule should have a *low* weight since they occur relatively seldom in actual simulations due to the molecule–molecule repulsion.

A third, argument against using nearby points is that the electron density is significant at these points, thereby perturbing the potential. In fact, Colonna et al. have shown that such electron-cloud penetration effects are significant up to 3 Å from heavy atoms and to 2 Å from hydrogen atoms [37]. For example, they find that at 1.5 Å (the CHELPG radius) from the carbon atom in formamide, the ab initio potential is less than 50% of the potential without penetration effects. The effect can be avoided by special methods to calculate the potential, e.g. the overlap multipole expansion technique [37].

In the following, we have chosen two different criteria for the quality of estimated charges. The first is based on the absolute deviation of the electrostatic potential calculated from the wavefunction and the charges. We have decided to use the merged four sets of electrostatic potentials used in the Merz-Kollman, CHELP, CHELPG, and CHELP-BOW (see below) methods. This was done to avoid any bias from the sampling techniques and Van der Waals radii. However, since an unweighted fit gives unfairly high weights to the potential points close to the molecule and ignores distant interactions, we have compared both the unweighted and Boltzmann-weighted potentials. The second criterion is the average absolute deviations of the multipole moments (in the real spherical harmonics form) calculated from the charges and the wavefunction. The absolute deviation is judged to be a better criterion than the relative or maximum deviation, since the total energy is linearly related to the moments as well as to the electrostatic potential.

The CHELMO method

In order to better illustrate our discussion of the quality criteria of the charges, we have developed two new methods to calculate atomic charges from quantum chemical calculations, CHELMO (charges from electrostatic moments) and CHELP-BOW (charges from Boltzmann-weighted electrostatic potentials). Both are devised to minimise the number of arbitrary choices in the fitting procedure and to give the optimal atomic charges in terms of the total electrostatic interaction energy in actual simulations.

The CHELMO method does not use the electrostatic potential; instead the charges are fitted directly to the electrostatic moments. Therefore, all choices related to the selection of potential points are avoided as are effects of electron-cloud penetration. In fact, CHELMO involves only two choices: the weight factors for the different restraints and the moments to be used as constraints.

As was discussed in the Methods section, we have used a weight factor for the restraints to the multipole moment of the form $(r+x)^{-n}$ (equation 9), where r is the radius of the molecule, n is the order of the moment, and x is a parameter. We have tested different values of x are for twelve different molecules and compared the mean absolute deviation of the electrostatic potential calculated from the fitted charges and from the wavefunction (results available as supplementary material). The x value that gives the minimum deviation varies for the different molecules. We have chosen $x = 3$ Å, since this was the average of the optimal x values for the various molecules. It is notable that for most of the molecules, the minimum average deviation of the electrostatic potential for the CHELMO method is smaller than the deviation obtained for a direct fit to the electrostatic potential. This is because the fitting procedure uses least-squares fitting to the electrostatic potential while we compare the absolute deviation. Pentane and imidazole were included in the test set to ensure that the weight in equation (9) works well also for elongated and flat molecules. Encouragingly, the results obtained with these two molecules do not differ from those of the other molecules.

We have used as constraints in CHELMO the total charge, the dipole moment (if possible), and equal charges on equivalent atoms. Whether the dipole moment should be used as a constraint or restraint depends on the number of independent atoms in the molecule and the number of non-vanishing components of the dipole moment. Naturally, the number of independent atoms has to be larger than the number of non-vanishing dipole components if it should be possible to constrain the dipole moment. It turned out that already when the number of independent atoms is one more than the number of non-vanishing dipole components, the dipole moment can safely be used as a constraint without deteriorating the higher moments (data not shown). Of course, it would be possible to constrain also the quadrupole moment if the number of independent atoms is high enough. However, we have not seen any advantage with such a procedure.

The CHELP-BOW method

Above, we suggest that electrostatic potential points should be selected according to their occurrence in actual simulations and that points far from the molecule are more important for the charges than points close to the molecule. In order to illustrate the effect of these suggestions, the CHELP-BOW method was developed. In the present implementation, the CHELP-BOW method involves many approximations, but this is acceptable for our illustrating purposes. However, the preliminary results are so promising that in a forthcoming publication, we will thoroughly test and calibrate the method for general use.

In the CHELP-BOW method, points are sampled within a large box enclosing the molecule. In principle, no selection of points is necessary before the fit, but we have decided to remove points with a very small weight in order to minimise the number of potential points. The points are sampled at random, since this has been recommended lately to avoid any orientational dependence [17-19]. Random sampling also gives points that are evenly distributed in the box, i.e. points with a correct frequency at different distances from the molecule if no interactions were present. Furthermore, the point density is very high (2500 points per atom) in order to ensure that the charges are well-determined and do not depend on the orientation of the molecule [16,17,29].

In the least-squares fit, the potential points are weighted according to their distance to the atoms in the molecule. More specifically, the points are weighted by the Boltzmann factor of the interaction energy of a probe atom at the potential point and all the atoms in the molecule. The atoms in the molecule are assigned parameters directly from the molecular simulation package in which the atomic charges will be used. By this choice, the Van der Waals parameters are no longer arbitrary. On the contrary, they are completely determined by the force field in which the charges will be used. This clearly shows the connection between the estimated charges and the force field to be used.

Unfortunately, there are some problems with such an approach. First, the probe molecule is not defined; in the simulation it might be any molecule. However, without further information, a water molecule would be the most natural choice. We have selected a TIP3P water molecule, since this is the default water model in the Amber software, which we intend to use. Such a probe is appropriate also because its Van der Waals interaction is determined only by the oxygen atom so we do not have to speculate about the position of the hydrogen atoms. Second, the interaction energy of the molecule with a probe depends on the charge of the atoms in the molecule, i.e. on the charges we want to determine. Moreover, it also depends on the charge and position of the atoms in the water probe. There are methods to solve these problems, but for simplicity we decided to ignore the charges and only use the Van der Waals interaction energy. Yet, this gives a too long distance between the probe and the molecule (e.g. 3.43 and 2.37 Å between O–O and O–H, respectively). We have corrected this artefact by decreasing the Van der Waals radius of the water probe by 0.4 Å, which gives O–O and O–H equilibrium distances close to the experimentally observed distances.

Normally, the Van der Waals interaction energy is strongly repulsive at short distances, shows an attractive minimum at the sum of the Van der Waals radii of the two interacting atoms, and slowly increases to zero at long distances. Since the Van der Waals interaction in practice is significant only for molecules in direct contact, while the energy of more distant molecules is determined by the electrostatics and the location of all other molecules, we only use the repulsive part of the Van der Waals interaction energy. Thus, if the distance between the probe and an atom in the molecule is larger than the sum of their Van der Waals radii, the interaction energy is set to the minimum energy. In this way, all positions outside the Van der Waals envelope of the molecule will have the same weight.

The wish to describe the electrostatics properly at large distances from the molecule poses another problem. Naturally, all points up to infinity cannot be included; at some finite distance the point selection has to stop. A solution to the problem would be to make a more sparse sampling of points at larger distances from the molecule, and compensate this by a higher weight in the fit. However, we have chosen a simpler solution. We include points up to 8 Å from any atom in the molecule, and in order to get the correct asymptotic behaviour of the electrostatic potential, we constrain the total charge and the dipole moment to reproduce the actual charge and the dipole moment calculated from the wavefunction. Thus, the two lowest moments, with an interaction energy that falls off slower than the increase in the number of interactions, are constrained to be correct. If the number of independent atoms is more than about 12 we recommend that also the quadrupole moment is used as a constraint, since its interaction energy falls of with the same pace as the increase in interactions. However, this is not the case for any molecule in the present investigation.

If the Boltzmann-weight behaves properly, it would compensate for most of the differences in the point-selection schemes in the traditional potential-based methods. In fact, this is the cases as is illustrated for methane in Table V. With normal, unweighted, fits the CHELP, CHELPG, and Merz-Kollman charges differ considerably (the charge on the carbon atom varies between 0 and –0.5 *e*). However, if the electrostatic potential is Boltzmann weighted as in the CHELP-BOW method, the charge on the carbon atom converges towards –0.52 to –0.57 *e* . Interestingly, this is close to the charge obtained with the full CHELP-BOW method (–0.55 *e*) as well as with the CHELMO method

(–0.57 *e*). Thus, the Boltzmann weight works properly, eliminating most of the differences due to different sampling schemes. Moreover, CHELP-BOW gives results that are similar to those of the CHELMO method, although the latter method estimates the charges directly from the electrostatic moments instead of the electrostatic potentials. Thus, we have constructed two methods that are based on different techniques, but nevertheless give very similar results, which is quite satisfactory.

Comparison of the different methods

In order to test the performance of the different methods, partial charges were calculated for 22 small and medium sized-molecules (Table VI) using the CHELP, CHELPG, Merz-Kollman, RESP, CHELMO, and CHELP-BOW methods. For comparison, Mulliken, NPA, and AIM charges were also included in the investigation. The quality of the charges was estimated from the mean absolute deviation of the multipole moments and the electrostatic potential (both Boltzmann-weighted and unweighted) calculated form the wavefunction and from the charges (Table VII). In order to get comparable results, the charges of equivalent atoms were constrained to be equal for all the electrostatic potential methods, and our numerically stable singular value decomposition procedure was used for the fits.

As can be seen from Table VI, the charges obtained with the various methods vary appreciably (by up to 1.24 *e*). In general, the charge on carbon atoms varies most, but the charges on other buried atom such as copper and nitrogen also vary appreciably. Interestingly, this variation is most pronounced in the five molecules involving sulphur. As has been noted before [1], the AIM charges and also the NBA charges differ appreciably from the other charges, often by being larger in magnitude. It is also notable that among the other charges, the RESP method give charges with the lowest magnitude, whereas the CHELMO, CHELP-BOW, and Merz-Kollman methods give charges with the largest magnitudes. This is probably an effect of that fewer (or no) electrostatic potentials near the molecule are used in the fit of the latter three methods.

In general, the CHELMO and CHELP-BOW charges are similar to at least some of the electrostatic potential charges. Thus, these new methods give sensible charges. Moreover, the two new methods give rather similar charges; the average difference is only 0.031 *e* and the largest difference is 0.20 *e*. This is quite satisfactory, since the methods are constructed with the same philosophy although technically they are quite different. The CHELP-BOW charges are also fairly similar to the Merz-Kollman charges, with an average deviation of 0.037 *e*. In fact, the CHELP-BOW charges usually lie between the CHELMO and Merz-Kollman charges. The difference to the other potential-based charges is larger, especially to the CHELP charges (average deviation 0.12 *e*).

If the multipole moments are used as a quality criterion of the charges, the CHELMO method is superior. It gives the best dipole, octupole, and hexadecapole moments, and the third best quadrupole moments. However, the CHELP-BOW charges are not much worse, giving the best quadrupole moment, the second best dipole and octupole moments, and the third best hexadecapole moment, and the Merz-Kollman charges also give good moments. AIM, NPA, Mulliken, and CHELP charges (in this order) give the worst moments, often with twice as large deviations as the CHELMO or CHELP-BOW charges.

If instead the absolute deviation between the electrostatic potentials calculated from the charges and from the wavefunction is used as a quality criterion, the Merz-Kollman charges are the best. However, all the potential-based methods (except CHELP) and CHELMO give similar results within 10 %. Again, AIM, NPA, and Mulliken charges (in this order) give bad result, with deviations up to 9 times larger than for the Merz-Kollman method.

It should be noted that the electrostatic potential points were the merged set of points used for the Merz-Kollman, CHELP, CHELPG, and CHELP-BOW fits. As was discussed above, these potential sets (especially the former three) give an unfair weight to points near the molecule and they ignore points far from the molecule where the most important interactions appear (in terms of total energy) and where the CHELMO and CHELP-BOW methods can be predicted to be superior. This is illustrated by the data in Table VIII, where the mean absolute deviation of the electrostatic potentials is listed, grouped after the distance of the electrostatic potential point to the closest atom in the molecule. We have also included a large number of electrostatic potential points far from the molecule (up to 25.5 Å). It can be seen that CHELP-BOW and CHELMO actually gives the smallest deviation of the electrostatic potentials for all distances except those closest (1.5-4.9 Å) to the molecule. Yet, it is only the latter points that are used in the fit of the Merz-Kollman and CHELPG methods. This clearly illustrates the difference between CHELMO and CHELP-BOW on one hand and the other electrostatic potential methods on the other hand. In the potential-based methods, the charges are mainly determined by the points close to the molecule, whereas the CHELMO and CHELP-BOW fit is constructed to preferably reproduce the electrostatic potential at larger distances from the molecule.

In order to compensate for the bias for potential points close to the molecules, we also made a weighted comparison of the potentials, using the same weights as in the CHELP-BOW method (Table VII). With weighted potentials, the CHELP-BOW method performs best, whereas the Merz-Kollman and CHELMO charges perform only slightly worse. Thus in conclusion, the CHELP-BOW method seems to give the most general charges, which reproduce the exact moments as well as the potential very well. In essence, it seems to provide a good compromise between the excellent moments of the CHELMO method and the good potential of the Merz-Kollman scheme.

It is notable the Merz-Kollman scheme gives better charges than the CHELP and CHELPG methods. The reason for this is most likely that this method selects electrostatic potential points more distant from the molecules than the other methods, because points are excluded within 1.4 times the Van der Waals radii. Judging from our criteria, it is disadvantageous to include points too near the molecule. It is also interesting that the CHELP charges perform appreciably worse than the other potential-based methods and that the RESP charges give significantly worse moments and potentials (up to 40%) than the Merz-Kollman charges, although they are based on the same potentials. Finally, we can conclude that the Mulliken, NPA, and AIM charges are unsuited to use when electrostatic properties are of interest. In particular, it is eye-catching that the NPA method constitutes a significantly impairment the Mulliken charges in this respect.

We have also examined to what extent the charges depend on the quantum chemical method and the basis sets. In Table IX, Merz-Kollman, CHELP-BOW, and CHELMO charges for the Cu(I) complex are calculated with different methods and basis sets. It can be seen that all three charge sets show a moderate dependence on the basis set. This is not a disadvantage of these methods, however, but rather a necessary result of the variation of the electrostatic moments and the electrostatic potential with basis sets; the charges cannot be expected to converge more rapidly than the moments with the size of the basis set. The Merz-Kollman charges show the smallest variation, slightly smaller than for the CHELP-BOW charges, whereas the variation is significantly larger for the CHELMO charges. Apparently, the electrostatic moments vary more than the electrostatic potential with the level of theory and the basis sets. The charges calculated with the MP2 and B3LYP methods are closely similar, while the Hartree-Fock charges are quite different. This agrees with the conclusions of De Proft et al. [25] that partial charges calculated with the B3LYP method are in good agreement with high-level ab initio results.

The electrostatic potential methods (especially CHELP) have been criticised for being dependent on the choice of the coordinate system [16]. We have therefore examined how the potential-based charges vary when the Cu(I) complex is rotated. As can be seen in Table X, all methods show a rotational dependence, although it is minimal for the CHELPG and CHELP-BOW charges (0.001 *e* in average), i.e. the methods with a large number of electrostatic potential points. For the CHELP-BOW method, the variation of the charges with the rotation reflects rather the uncertainty in the charges than the rotational dependence, since the points are selected randomly. The CHELP and Merz-Kollman charges show an appreciable dependence of the orientation of the molecule, in average 0.04-0.05 *e*. If the real spherical harmonics form of the multipole moments are used in the CHELMO method, it is rotational invariant, as are the Mulliken, NBO, and AIM methods. Naturally, this is an advantage of these methods.

The result in Table X represents a typical rotation. A closer investigation of different rotations shows that similar variations are obtained with other rotations, independently of the rotation axis and the magnitude of the rotation (Fig. 1). Thus, even for very small rotations (a few degrees) a 0.05 *e* variation of the Merz-Kollman charges can be expected. This can partly explain the large variations in electrostatic potential charges encountered during quantum molecular dynamics simulations [36].

The conditioning of the least-squares matrix

Recently, it has been pointed out that the least-squares matrix for the electrostatic potential methods in general is rank deficient, meaning that the corresponding charges are not statistically valid. This is a very serious problem, and therefore we have used numerically stable algorithms based on singular value decomposition for our fitting. The singular values of a matrix directly show whether it is rank-deficient or not. We used the same criteria for rank-deficiency as Francl et al [18], that the condition number (i.e. the quotient between the largest and the smallest singular values) is less than 10^{-5}, on both the constraint (**B**) and the restraint (**A**) matrices. Quite unexpectedly, we did not encounter any rank-deficiencies for any of the molecules in Table VI with any method.

In order to deduce why our results differ so strongly from those of Francl et al. [18], we decided to study exactly the same system as in their investigation. The molecules have between 6 and 36 atoms and they are studied with the CHELP method. In addition, a more than twice as large molecule was also included, the haem group and its axial ligands in oxygen-loaded myoglobin (84 atoms). The results are collected in Table XI, showing the condition numbers of the restraint matrix (in these calculations, the constraint matrix **A** is one-dimensional and therefore always have condition number of 1). It can be seen that the condition number decreases with the number of atoms in the molecule, but even in the 84-atom myoglobin model, it is still 0.001, i.e. 100 times larger than the limit for a rank-deficiency. On the other hand, Francl et al. [18] encountered deficiencies in the rank of the least-squares matrix for all molecules (between 1 and 12).

This discrepancy is partly due to the method of the least-squares fit. In our fitting procedure, the equation $\mathbf{Aq} = \mathbf{f}$ is solved using pseudoinverses, whereas Francl et al. [18] solves the equation $\mathbf{A}^{\mathbf{T}}\mathbf{Aq} = \mathbf{A}^{\mathbf{T}}\mathbf{f}$ by forming the inverse of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$. As was discussed by Hinsen and Roux [17], this latter multiplication with $\mathbf{A}^{\mathbf{T}}$effectively squares the condition number of the **A** matrix, thereby strongly increasing the risk of yielding an ill-conditioned problem. Yet, this does not explain all the discrepancy. Even if our condition numbers are squared, only the condition number of the myoglobin model will become less than 10^{-5}. All the charges of the other molecules will still be well-determined. Probably, the rest of the problem lies in the treatment of the constraints. Francl et al. use Lagrangian multipliers which increase the dimension of the **A** matrix. We, on the other hand, eliminate the constraints using the **B** matrix, which effectively reduces the size of the **A** matrix.

In conclusion, the results in Table XI illustrate the advantage of our fitting scheme involving pseudoinverses calculated by singular value decomposition and elimination of constraints. Clearly, it is possible to estimate well-determined atomic charges for molecules with at least 100 atoms if a proper method for the least-squares fit is used.

Concluding remarks

Potential-based charges strongly depend on the way the potential points are selected. We have presented and tested a new method that avoids such a dependence, CHELMO. It fits the charges directly to the electrostatic moments and it performs as well, or better, as the best electrostatic potential method judging from the calculated electrostatic potential and moments. In fact, it describes the electrostatic potential especially well far from the molecule, which is appropriate since those interactions are most important for the total interaction energy. The method is easy and fast to use, since the electrostatic moments are contained in the output of normal quantum chemical programs and no electrostatic potential needs to be calculated. Moreover, it avoids the problem of electron-cloud penetration. The major disadvantage of the method is that it cannot be used for large molecules, since there is a limited number of moments. No more than 25 independent charges in a molecule can be determined if all moments up to hexadecapole moments are used. In practice it should not be used for molecules with more than about 20 atoms since otherwise the higher moments may be poorly reproduced and the statistics of the fit may become poor. This could be remedied if moments higher than hexadecapole moments are calculated but such moments are not included in the output of normal quantum chemical packages.

In order to get a method that can be used for larger system, we constructed the CHELP-BOW method, which estimates the charges from a Boltzmann-weighted fit to the electrostatic potential. It has the advantage over the other electrostatic potential methods that it gives the best possible electrostatic interaction energy in actual simulations by weighting the electrostatic potential points according to their occurrence in the simulations. Thus, it does not give an unfairly high weight to electrostatic potential points close to the molecule as in the other methods. Furthermore, it does not contain any arbitrary parameters for the potential point selection; instead it employs the potential function of the program for which the charges will be used. The present implementation of the CHELP-BOW method is very simplified, but in a future publication we will refine and thoroughly test the method. However, the results are conclusive enough to show that the method is very promising and has the desired behaviour. In fact, it gives both excellent electrostatic moments and electrostatic potentials, thereby combining the attractive characteristics of both the CHELMO and Merz-Kollman methods. Clearly, this is the method we recommend for general use.

An important advantage of the methods developed in this paper is that they are general, i.e. they can be used with any quantum chemical program and they can use experimental data (e.g. multipole moments) as well. The only thing that has to be changed is the input section of the programs. Thus any quantum chemical basis sets can be used, and any level of theory that gives a wavefunction may be employed. Furthermore, the methods can easily be adapted to use the same hyperbolic restraints as in the RESP method. However, this means that a non-linear fitting method has to be used and it is not clear how such a method behaves with linearly dependent electrostatic potential points (i.e. if the method is sensitive to the least-squares rank problem). Furthermore, our results show that RESP clearly impede the electrostatic behaviour of the fitted charges (by up to 40%). On the other hand, with a non-linear method, it would also be possible to fit the charges to the absolute (instead of the squared) deviation of the electrostatic potentials or moments, which we believe is an advantage.

Finally, a comment on the traditional electrostatic potential methods. As was discussed above, the Merz-Kollman method gives the best charges according to our electrostatic criteria, but they show an appreciable dependence on the orientation of the molecule compared to the CHELPG method. Therefore, if any of the standard electrostatic potential method should be used, we recommend the use of the Merz-Kollman method, but with a much higher point density than default (at least 2000 points per atom). From our results, it is also clear that the standard CHELP method is inferior in all respects (bad moments, electrostatic potentials, and rotational dependence) to the other potential-based methods. This is probably due to the highly uneven distribution of the electrostatic potential points. Thus, we cannot see any justification to use CHELP charges except for backward comparisons.

Acknowledgements

This investigation has been supported by grants from the Swedish Natural Science Research Council (NFR). We thank Gunnar Karlström and Per-Åke Malmqvist for fruitful discussions and suggestions.

References

1. S. M. Bachrach, Rev. Comp. Chem., 5, 171 (1994).
2. D. P. Pearlman and S. H. Kim, Biopolymers, 24, 327 (1985).
3. A. K. Rappé and W. A. Goddard, J. Phys. Chem. 95, 3358 (1991).
4. P. Coppens, Annu. Rev. Phys. Chem 43, 663 (1992).
5. R. S. Mulliken, J. Chem. Phys, 36, 3428 (1962).
6. R. S. Mulliken and P. Politzer, J. Chem. Phys, 55, 5135 (1971).
7. P.-O. Löwdin, Adv. Quantum Chem. 5, 185 (1970).
8. A. E. Reed, R. B. Weinstock, and F. A. Weinhold, J. Chem. Phys., 83, 735 (1985).
9. R. F. W. Bader, Atoms in Molecules. A quantum theory, Clarendon Press, Oxford, UK, 1990.
10. K. B. Wiberg and P. R. Rablen, J. Comput. Chem. 14, 1504 (1993).
11. Chipot, C., Maigret, B., Rivail, J.- L. and Scheraga, H. A., J. Phys. Chem. 96, 10276 (1992).
12. Ángyán, J. G. and Chipot, C., Chem. Phys. Lett., 241, 51 (1995).
13. L. E. Chirlain and M. M. Francl, J. Comp. Chem., 6, 894 (1987).
14. U. C. Sing and P. A. Kollman, J. Comp. Chem., 5, 129 (1984).
15. B. H. Besler, K. M. Merz and P. A. Kollman, J. Comput. Chem. 11, 431 (1990).
16. C. M. Breneman and K. B. Wiberg, J. Comp. Chem., 11, 361 (1990).
17. R. J. Woods, M. Khalil, W.Pell, S. H.Moffat and V. H. Smith, J. Comput. Chem. 11, 297 (1990).
18. M. M. Francl, C. Carey, L. E. Chirlain and D. M. Gange, J. Comp. Chem., 17, 367 (1996).
19. K. Hinsen and B. Roux, J. Comp. Chem., 18, 368 (1997).
20. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, J. Phys. Chem, 97, 10269 (1993).
21. W. J. Here, L. Radom, P. V. R. Scheyer and J. A. Pople, Ab Initio Molecular orbital theory, Wiley-Interscience, New York (1986).
22. A. Schäfer, H. Horn and R. Ahlrichs, J. Chem. Phys. 97, 2571 (1992).
23. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A.

Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian 94, Revision D.1, Gaussian, Inc., Pittsburgh PA, 1995.

24. V. Barone, Chem. Phys. Letters, 226, 392 (1994).

25. F. De Proft, J. M. L. Martin and P. Geerlings, Chem. Phys. Letters, 250, 393 (1996).

26. D. A. Pearlman, D. A. Case, J. W. Caldwell. W. S. Ross, T. E. Cheatham III, D. M. Ferguson, G. L. Seibel. U. C. Sing, P. K. Weiner and P. A. Kollman, AMBER 4.1, University of California, San Francisco, 1995.

27. U. Ryde, M. H. M. Olsson, K. Pierloot and B. O. Roos, J. Mol. Biol., 261, 586 (1996).

28. S. E. V. Philips and B. P. Schoenborn, Nature 292, 81 (1981).

29. B.Beck, T. Clark and R. C. Glen, J. Comput. Chem. 18, 744 (1997).

30. W. D.Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, J. Am. Chem. Soc. 117, 5179 (1995).

31. J.E. Rice, H. Horn, B. H. Lengsfiels, A. D. McLean, J. T. Carter, E. S. Replogle, L. A. Barnes, S. A. Maluendes, G. C. Lie, M. Gutwski, W. E. Rude, S. P. A. Sauer, R. Lindh, K. Andersson, T. S. Chevalier, P. -O. Widmark, D. Bouzida, G. Pacansky, K. Singh, C.J. Gillan, P. Carnevali, W. C. Swope and B. Liu, Mulliken^{TM} Version 2.25b, internal release, IBM Corporation, Almaden, USA, 1995.

32. S. Tsuzuki, T. Uchimaru, K. Tanabe and A. Ylinniemela, J. Mol. Structure, 365, 81 (1996)

33. I. Rozas, Int. J. Quant. Chem. 62, 477 (1997).

34. W. D.Cornell, P. Cieplak, C. I. Bayly and P. A. Kollman, J. Am. Chem. Soc. 115, 9620 (1993).

35. J. W.Storer, D. J. Giesen, C. J. Cramer and D. G. Truhlar, J. Comput. Aided Mol. Design. 9, 87 (1995).

36. K.M. Merz, in Molecular modeling and dynamics of biological molecules containing metal ions, NATO advanced research workshop, Pisa, p. 39, 1997.

37. F. Colonna and E. Evleth, J. Comput. Chem. 13, 1234 (1992).

38. L. J. Bartolotti, L. G. Pedersen and P. S. Charifson, J. Comput. Chem. 12, 1125 (1991).

39. Abrahamsson and I. Pascher, Acta Cryst. 21, 79 (1966).

Table I. The Van der Waals parameters used in the CHELP-BOW method [30, 38].

Element	r_i^* (Å)	ε_i (kJ/mole)
H polar	0.6	0.0657
H non-polar	1.487	0.0657
C	1.908	0.3598
N	1.824	0.7113
O	1.6686	0.6360
P	2.100	0.8368
S	2.000	0.8368
Cl	2.47	0.4184
Cu	1.17	4.7698
Water probe	1.7683	0.6360

Table II. Charges of the Cu(II) complex calculated with five different methods.

	Mulliken	CHELP	CHELPG	RESP	Merz-Kollman
Cu	0.40	0.11	0.24	0.26	0.38
S_{SH2}	–0.26	–0.20	–0.29	–0.28	–0.34
H_{SH2}	0.17	0.18	0.21	0.19	0.22
S_{SH}	–0.07	–0.16	–0.20	–0.24	–0.23
H_{SH}	0.13	0.13	0.16	0.19	0.15
N	–0.94	–0.13	–0.42	–0.33	–0.56
H_N	0.38 to 0.39	0.16 to 0.18	0.25 to 0.26	0.23	0.28 to 0.30

Table III. The dependence of the electrostatic potential charges on the copper radius[a].

	$r = 0.5$ Å	$r = 1.0$ Å	$r = 1.5$ Å	$r = 2.0$ Å	$r = 2.5$ Å	$r = 3.0$ Å
Cu	2.20	0.70	0.25	0.24	0.31	0.40
S_{SH2}	–0.57	–0.36	–0.30	–0.29	–0.30	–0.31
H_{SH2}	0.27	0.22	0.21	0.21	0.20	0.20
S_{SH}	–0.54	–0.29	–0.20	–0.20	–0.22	–0.25
H_{SH}	0.11	0.14	0.15	0.16	0.16	0.16
N	–3.41	–0.94	–0.41	–0.42	–0.46	–0.52
H_N	0.87 to 1.16	0.35 to 0.41	0.25 to 0.26	0.25 to 0.26	0.25 to 0.27	0.26 to 0.29

[a] The charges of the Cu(II) complex were determined by the CHELPG method using different radii for copper and the default radii for the other atoms.

Table IV. The dependence of the charges on the point-selection scheme[a].

Method	CHELP	CHELPG	Merz-Kollman	Merz-Kollman[b]
Cu	0.11	0.24	0.32	0.22
S_{SH2}	–0.20	–0.29	–0.35	–0.28
H_{SH2}	0.18	0.21	0.22	0.20
S_{SH}	–0.15	–0.20	–0.23	–0.18
H_{SH}	0.11	0.16	0.16	0.15
N	–0.14	–0.42	–0.47	–0.43
H_N	0.16 to 0.18	0.25 to 0.26	0.26 to 0.27	0.25 to 0.26

[a] The charges of the Cu(II) complex were calculated with the three different methods, all using the same Van der Waals radii, viz. those of the CHELPG method (and 2.0 Å for copper).

[b] Using CHELPG radii divided by 1.4.

Table V. Charges for methane calculated with and without Boltzmann weights.

Method	C	H
Merz-Kollman	–0.50	0.12
CHELP	–0.00	0.00
CHELPG	–0.37	0.09
weighted Merz-Kollman	–0.52	0.13
weighted CHELP	–0.54	0.13
weighted CHELPG	–0.57	0.14
CHELMO	–0.57	0.14
CHELP-BOW	–0.55	0.14

[a] The electrostatic potential points were sampled with the CHELP, CHELPG, and Merz-Kollman methods. Charges were fitted to the electrostatic potential points with our without a Boltzmann weight as described for the CHELP-BOW method. The charges on the four hydrogen atoms were constrained to be equal.

Table VIII. The quality of the various point charge models for the electrostatic potential calculated at different distances from the Cu(II) complex[a].

Distances (Å)	1.5-4.9	4.9-8.3	8.3-11.8	11.8-15.2	15.2-18.6	18.6-22.1	22.1-25.5	Total
# points	3715	4868	6816	6959	7921	4307	4411	38997
Mulliken	4005	1276	603	400	284	171	158	812
CHELP	1747	362	142	88	59	38	32	272
CHELPG	1280	184	72	45	33	20	18	176
RESP	1287	155	62	39	28	17	15	169
Merz-Kollman	1171	138	55	36	26	15	14	153
CHELP-BOW	1191	85	19	8	4	2	1	130
CHELMO	1454	101	27	12	6	4	2	160

[a] The absolute mean deviation of the electrostatic potential calculated from the fitted charges and from the wavefunction at different distances from the closest atom is listed for the two copper complexes. The points were selected in shells around the molecule as in the Merz-Kollman scheme, but between 1.0 and 8.5 times the Van der Waals radii. Unit: 10^{-6} a.u.

Table IX. Merz-Kollman and CHELMO charges calculated for the Cu(I) complex with different basis set and quantum chemical methods.

Merz-Kollman charges

	Hartree-Fock		B3LYP		MP2	
Basis set:	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf
Cu	0.48	0.54	0.17	0.28	0.19	0.28
S_{SH2}	–0.49	–0.41	–0.47	–0.38	–0.45	–0.39
H_{SH2}	0.24	0.20	0.21	0.20	0.22	0.21
S_{SH}	–0.92	–0.88	–0.77	–0.78	–0.81	–0.80
H_{SH}	0.15	0.12	0.16	0.13	0.15	0.14
N	–0.83	–0.66	–0.60	–0.52	–0.59	–0.55
H_N	0.33	0.26	0.28	0.23	0.28	0.24

CHELP-BOW charges

	Hartree-Fock		B3LYP		MP2	
Basis set:	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf
Cu	0.45	0.51	0.10	0.25	0.11	0.25
S_{SH2}	–0.49	–0.44	–0.45	–0.41	–0.45	–0.42
H_{SH2}	0.24	0.22	0.21	0.21	0.23	0.22
S_{SH}	–0.91	–0.88	–0.76	–0.78	–0.80	–0.80
H_{SH}	0.15	0.12	0.17	0.14	0.15	0.14
N	–0.86	–0.63	–0.61	–0.48	–0.60	–0.51
H_N	0.34	0.25	0.29	0.22	0.29	0.23

CHELMO charges

	Hartree-Fock		B3LYP		MP2	
Basis set:	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf	3-21G	6-31G* / DZpdf
Cu	0.48	0.65	0.07	0.38	0.14	0.39
S_{SH2}	–0.52	–0.50	–0.47	–0.46	–0.47	–0.48
H_{SH2}	0.25	0.22	0.22	0.21	0.23	0.22
S_{SH}	–0.94	–0.92	–0.77	–0.82	–0.82	–0.84
H_{SH}	0.16	0.13	0.18	0.14	0.17	0.15
N	–0.82	–0.57	–0.55	–0.42	–0.56	–0.45
H_N	0.33	0.23	0.27	0.20	0.27	0.21

Table X. The dependence of the various charges on the coordinate system[a].

Method	CHELP	CHELPG	Merz-Kollman	CHELP-BOW
Cu	0.106	0.007	0.044	0.002
S_{SH2}	0.062	0.001	0.003	0.000
H_{SH2}	0.017 to 0.079	0.000 to 0.001	0.000 to 0.004	0.000
S_{SH}	0.006	0.000	0.007	0.004
H_{SH}	0.010	0.001	0.009	0.004
N	0.170 to 0.065	0.000	0.046	0.000
H_N	0.008 to 0.077	0.000 to 0.001	0.001 to 0.118	0.000
Average	0.046	0.001	0.037	0.001

[a] The change in the charges caused by a rotation of the Cu(I) complex 110° around the *x*-axis is listed.

Table XI. The rank estimates of CHELP least-squares fit[a].

Molecule	# atoms	Condition no.	CHELP-SVD[b]
Formamide	6	0.079	1
Ethane	8	0.037	1
Acetamide	9	0.031	2
Methyl acetate	11	0.026	2
Dimethylphosphate	13	0.038	3
Phenol	13	0.023	3
L-cysteine	14	0.033	3
t-Butoxide	14	0.025	3
2,4-Pentanedione	15	0.020	3
Adenine	15	0.027	4
Neopentane	17	0.027	4
Alanine dipeptide (tr)	22	0.011	5
Glucose	24	0.014	7
GPC[c]	36	0.010	12
Myoglobin model	84	0.001	–

[a] The molecules were optimised at the HF/3-21G(*) level, except GPC and the myoglobin model, for which the coordinates were taken from literature [28,39]. The electrostatic potential points were sampled with the CHELP method. The fit was performed as described in the Methods, with the total charge as the only constraint. Therefore, the condition number of the constraint matrix is always 1. The condition number of the restraint matrix is defined as the quotient of the smallest and the largest singular values.

[b] Rank deficiencies encountered by the CHELP-SVD method [18].

[c] L-α-glycerylphosphorylcholin

Figure caption

Figure 1. The variation of the carbon charge in CH_3SH on the orientation of the molecule. The charges are calculated with the standard Merz-Kollman method. The molecule is rotated around the *x*-axis starting from the Gaussian 94 standard orientation.