Chem-Bio Informatics Journal, Vol. 3, No. 2, pp.78-85(2003)

Consistent Charge Equilibration Method Combined with Universal Force Field: Application to Amino Acid Molecules

Tetsuji Ogawa [1*], Osamu Kitao [2], Noriyuki Kurita [1], Hideo Sekino [1], Shigenori Tanaka [3]

[1] *Department of Knowledge-based Information Engineering, Toyohashi University of Technology, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan*
[2] *Photoreaction Control Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 5, Tsukuba, Ibaraki 305-8565, Japan*
[3] *Corporate R&D Center, Toshiba Corporation, Kawasaki 212-8582, Japan*

**E-mail: ogawa@cochem2.tutkie.tut.ac.jp*

(Received March 11, 2003; accepted April 20, 2003; published online May 31, 2003)

Abstract

In the previous paper (O. Kitao and T. Ogawa, Mol. Phys., 101, 3-17 (2003).), we have proposed the consistent charge equilibration (CQEq) method. The CQEq energy term was combined with the universal force field (UFF) to develop the CQEq with UFF (CUFF). In this article, to confirm the accuracy of the CUFF, geometry optimizations by the CUFF were performed for a series of amino acid molecules. The CUFF can well reproduce the HF/6-31G geometries aside from some flexible dihedral angles. The partial charges obtained by the CQEq deviate somewhat from those by the restrained electrostatic potential fit; this result suggested us a way to improve the CQEq and the CUFF.**

Key Words: Charge Equilibration Method, Atomic Charge, Universal Force Field, CQEq, CUFF, Amino Acid

Area of Interest: Molecular Computing

1. Introduction

The reliability of molecular simulation usually depends on the quality of the force field employed. One of the important terms of the force field is the electrostatic (ES) term. ES force persists even over a long range and constitutes a major part of intermolecular interaction. The majority of standard molecular simulations use a simple point charge model for the ES term. In these models, fixed partial charges are assumed for each component atom. For example, the AMBER [1] force field assigns fixed partial charges defined by the restricted electrostatic potential (RESP) fit [2].

The atomic charges, however, should vary depending on the environmental field and the

geometry of the molecule. The simulations with fixed charges thus cannot describe polarization and charge-transfer effects. This issue is particularly essential in biopolymeric systems such as protein and nucleic acid molecules. Some models that include polarization or charge transfer effects are proposed [3][4][5][6] to overcome limitations in the fixed charge model.

Some approaches for estimating the atomic partial charges according to the molecular geometry are based on a density functional concept. The charge equilibration (QEq) method proposed by Rappé and Goddard [3] is one of these approaches. In this framework, we have derived the consistent charge equilibration (CQEq) method [7], which uses an identical energy expression for the calculations of both partial charges and the electrostatic energy gradient. The CQEq method completely retains the consistency in molecular geometry optimization or molecular dynamics simulation with ES interactions. The CQEq energy term is then combined with a generic force field, the universal force field (UFF)[8], to develop the consistent charge equilibration with universal force field (CUFF). In this work, the CUFF is applied to the modeling of amino acid molecules, and the reliability and capability of the CUFF are discussed.

2. Theory and Calculation

The total energy in the UFF [8] is given by the following equation,

$$E = E_R + E_\theta + E_\phi + E_\omega + E_{\mathrm{vdw}} + E_{\mathrm{ES}}. \tag{1}$$

E_R, E_θ, E_ϕ and E_ω are valence terms of bond stretching, bond angle bending, dihedral angle torsion, and inversion energies, respectively. E_{vdw} and E_{ES} are nonbonding terms of van der Waals and electrostatic (ES) energies. The ES term in the UFF is expressed as

$$E_{\mathrm{ES}}^{(\mathrm{UFF})} = \frac{1}{2}\sum_i \sum_{j\neq i} \frac{q_i q_j}{R_{ij}}. \tag{2}$$

The partial charges q_i are obtained by using the charge equilibration (QEq) scheme. In the UFF, the van der Waals and ES interactions for atoms that are bonded to each other (1-2 interactions) and bonded to a common atom (1-3 interactions) are excluded.

The energy expression in the QEq developed by Rappé and Goddard [3] is expressed as follows,

$$\begin{aligned} E_{\mathrm{ES}}^{(\mathrm{QEq})} = & \sum_{i\in\mathrm{Hydrogen}} \left[E_i^0 + \chi_i^0 q_i + \frac{1}{2} J_{ii}^0 \left(1 + \frac{q_i}{\zeta_{\mathrm{H}}^0} \right) q_i^2 \right] \\ & + \sum_{i\notin\mathrm{Hydrogen}} \left(E_i^0 + \chi_i^0 q_i + \frac{1}{2} J_{ii}^0 q_i^2 \right) + \frac{1}{2}\sum_i \sum_{j\neq i} J_{ij} q_i q_j \end{aligned}. \tag{3}$$

The first and second terms express one-center components of ES energy for hydrogen atoms or non-hydrogen atoms, respectively. The subscript *i* stands for the number of an atom in the system and the summations of the first and second terms are for hydrogen atoms and non-hydrogen atoms, respectively. The third term stands for the two-center component of the ES energy. The E_i^0, χ_i^0 and J_{ii}^0 terms are the ES energy of the neutral atom (treated as constant value), atomic electronegativity, and idempotential (self-Coulomb)[3] of the *i*-th atomic site. J_{ii}^0 is double the atomic hardness η_i^0. The terms χ_i^0 and J_{ii}^0 are atomic parameters given to each element, and q_i is the partial charge for each atom to be obtained. J_{ij} is a two-center electron repulsion integral

Table 1. Atomic parameters for the CQEq calculations.

	χ^0 (eV)	J_{ii}^0 (eV)	R (Å)
Hydrogen	4.528	13.8904	0.371
Carbon	6.27	10.00	0.757
Nitrogen	7.30	14.46	0.700
Oxygen	7.54	12.16	0.680
Sulfur	6.22	8.28	1.047

$J_{ij} = \frac{|\phi_i|^2 |\phi_j|^2}{R_{ij}}$, using the normalized *s*-type Slater function $\phi_{m,\zeta}(R) = N_m R^{m-1} \exp(-\zeta R)$ located on each atomic site. The exponent factor ζ_i is connected to the atomic radii R_i given for each element by

$$\zeta_i = a \frac{2m+1}{2R_i}, \tag{4}$$

and we adopt $a = 0.5$ [3]. For hydrogen atoms, ζ_{H} is treated as a charge-dependent variable, $\zeta_{\mathrm{H}} = \zeta_{\mathrm{H}}^0 + q_{\mathrm{H}}$, where ζ_{H}^0 is the initial exponential value calculated by equation 4. Table 1 lists the parameter set used in this study. The atomic electronegativity χ^0 and hardness $\eta^0 (= \frac{1}{2} J_{ii}^0)$ are taken from Pearson's work [9], except for hydrogen, whose values are taken from the original QEq [3]. The atomic radii *R* are taken from the UFF [8]; we adopt the longest value of *R* among those given for the specific element.

The set of partial charges giving a stationary point to the ES energy with the constraint for total charge Q_{total} is obtained by the following procedure. We consider the functional *L* with the Lagrange multiplier λ,

$$L = E_{\mathrm{ES}} + \lambda \left(Q_{\mathrm{total}} - \sum_i q_i \right). \tag{5}$$

In the consistent QEq (CQEq) scheme [7], the derivative of *L* with respect to q_i is

$$\begin{aligned} \frac{\partial L}{\partial q_i} &= \frac{\partial E_{\mathrm{ES}}}{\partial q_i} - \lambda \\ &= \begin{cases} \chi_i^0 + J_{ii}^0 \left(1 + \frac{3q_i}{2\zeta_{\mathrm{H}}^0} \right) q_i + \sum_{j \neq i} \left(J_{ij} + q_i \frac{\partial J_{ij}}{\partial q_i} \right) q_j - \lambda & (i \in \text{Hydrogen}) \\ \chi_i^0 + J_{ii}^0 q_i + \sum_{j \neq i} J_{ij} q_j - \lambda & (i \notin \text{Hydrogen}) \end{cases} \end{aligned} \tag{6}$$

In the original QEq, the charge-dependency of ζ_{H} is neglected when calculating $\frac{\partial L}{\partial q_i}$. This makes charge calculations less time-consuming, but also makes it hard to obtain correct energy gradients. In the CUFF, the charge-dependency of ζ_{H} is taken into account. The energy gradient with respect

to the Cartesian coordinate $R_k^a (a = x, y, z)$ in the CUFF is, in turn, given as follows,

$$\frac{\partial E_{\mathrm{ES}}}{\partial R_k^a} = \sum_{j \neq k} q_k q_j \frac{\partial J_{kj}}{\partial R_a^k} . \tag{7}$$

This gradient expression is simple in comparison with the QEq when the exact formula for $\frac{\partial L}{\partial q_i}$ is obtained. We substitute the CQEq energy term (equation 3) for the electrostatic term in the UFF (equation 2) and evaluate the gradient of the ES term by using equation 7. We denote this procedure as CUFF (consistent QEq with UFF).

We performed test calculations on 20 standard amino acid molecules. The N-acetyl, N'-methyl amino acid amide (CH_3-CO-NH-C_αHR-CO-NH-CH_3) was used. For reference, *ab initio* molecular orbital calculations were performed, in which geometry was optimized by the Hartree-Fock (HF) method with the 6-31G** basis set using the Gaussian 98 program [10]. Lysine and arginine were optimized in cationic form and glutamic acid and aspartic acid were in anionic form. Histidine was optimized in three forms; cationic and with a hydrogen atom bound to the N_δ atom and to the N_ε atom. Initial geometries of the CUFF optimization were the HF/6-31G** optimized one. The obtained charge distribution by the CUFF was compared with that by HF/6-31G** to assess the accuracy of the CUFF.

3. Results and Discussion

We constructed the CUFF by adopting the CQEq electrostatic energy term in the UFF. Fixed-point charges are assumed in the UFF, whereas variable and spatially distributed charges in the form of the *s*-type Slater function are assumed in the CUFF. This improvement of the CUFF enables us to obtain a more accurate charge distribution and electrostatic field than those obtainable by the UFF.

Another difference between the CUFF and the UFF is that the ES term of the UFF excludes the 1-2 and 1-3 interactions, which are effectively included in valence terms, whereas in the CUFF all atomic pairs are summed up, including the 1-2 and 1-3 interactions. The substitution of the ES term in the CUFF, however, may result in double counting of the 1-2 and 1-3 electrostatic energies through the UFF valence term and the CQEq terms. To evaluate the effect of this double counting, we performed geometry optimizations by both the UFF and CUFF. In the UFF calculation, partial charges were determined by the CQEq scheme for the CUFF-optimized geometry, and the fixed partial charges were used during the geometry optimization. On the other hand, partial charges were determined at each optimization step in the CUFF calculation. The 22 structures of 20 standard amino acid molecules were optimized. The bond lengths, bond angles, and dihedral angles in the Z-matrix were compared. The root-mean-square (RMS) errors between the UFF and the CUFF optimized structures were 0.0060 Å (bond length), 0.64° (bond angle), and 4.0° (dihedral angle). This revealed that the double counting in the CUFF had little effect on the geometry optimization.

To confirm the accuracy of the CUFF, geometry optimizations in the level of HF/6-31G** were also performed for the 22 structures of 20 standard amino acid molecules. The optimized geometries obtained by the CUFF were compared with those by HF/6-31G**. Figure 1 shows the histograms of the differences of bond lengths, bond angles, and dihedral angles between the HF and CUFF geometries. The RMS errors between the HF and CUFF geometries were 0.026 Å (bond length), 2.2° (bond angle), and 10.9° (dihedral angle). It is seen that the bond lengths of the CUFF geometry tend to be about 0.02 Å longer than those of the HF. The differences of bond angles fall in

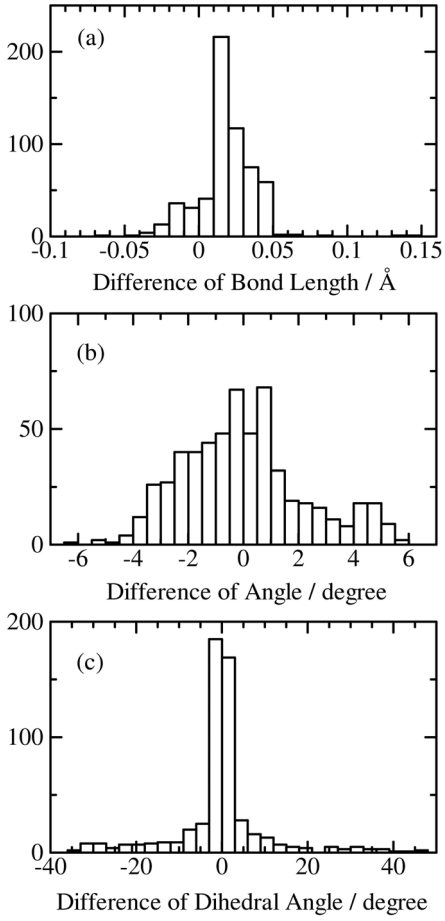

Figure 1. The differences in the optimized structures by the CUFF and HF/6-31G** for the 22 amino acid geometries, for bond lengths (a), band angles (b), and dihedral angles (c).

the range of ±5°. In regard to dihedral angles, about 80 % of the angles are in agreement within the error of ±10°, although some angles have a discrepancy larger than 30° between the HF and CUFF geometries. These angles are concerned with the flexible parts of the amino acid molecules; e.g., the backbone dihedral angles concerning the C_α atom (ϕ, ψ) and the terminal -CH_3 group. We thus concluded that the CUFF can well reproduce the HF geometry aside from some flexible dihedral angles.

To examine the accuracy of the partial charges obtained by the CUFF, they were compared with the charges derived from an *ab initio* molecular orbital method. Unfortunately, there is no unique way to determine atomic partial charge in a quantum mechanical framework, and many charge assignment schemes have been proposed. Among them, we adopted the restrained electrostatic potential (RESP) fit [2]. In the RESP, a set of point charges is derived so as to reproduce the electrostatic potential (ESP). The RESP scheme is employed for determining partial charges in the AMBER force field [1], which have been successfully used in many biopolymer simulations. We calculated the RESP charges for the standard amino acids, whose geometries were optimized by HF/6-31G**. The CQEq and the RESP charges obtained for the 22 amino acid molecules are compared in Figure 2(a)-(d) for hydrogen, carbon, nitrogen, and oxygen atoms, respectively. The diagonal line in the figure indicates that the CQEq and the RESP charges are equal. Deviation from

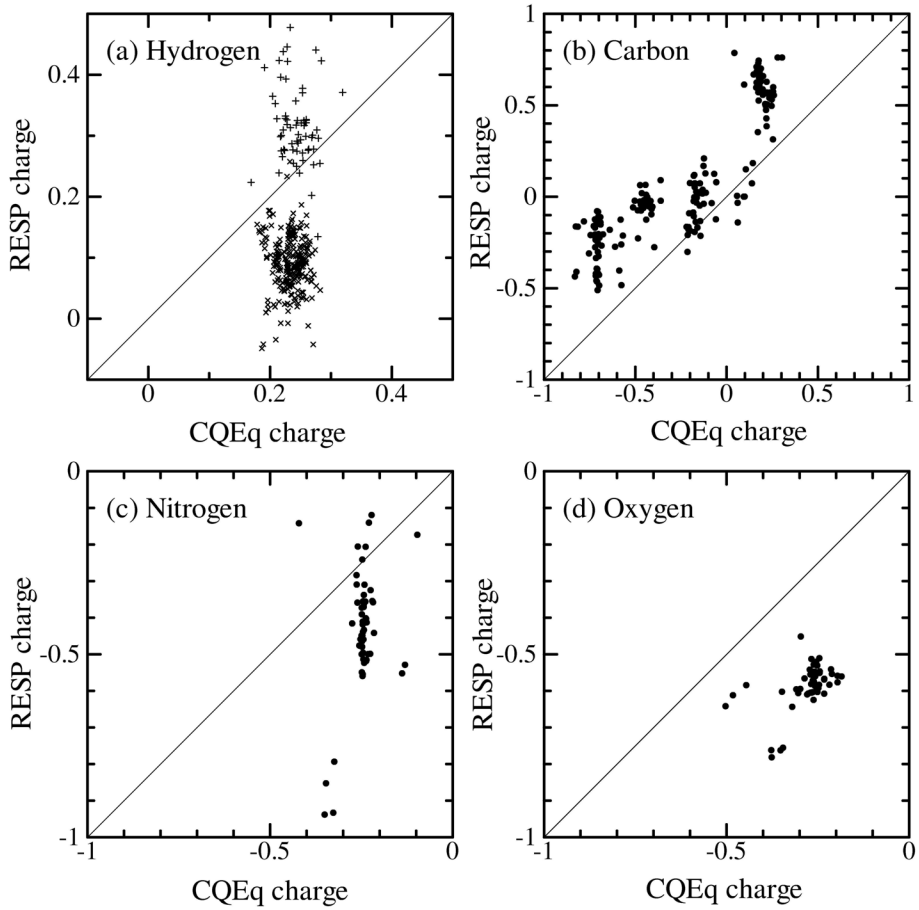

Figure 2. The partial charges obtained by the CQEq and RESP based on the HF/6-31G** result for the 22 amino acid geometries. The partial charges on hydrogen (a), carbon (b), nitrogen (c), and oxygen (d) are shown. The partial charges on hydrogen connected to carbon are plotted by a cross (×) and those connected to nitrogen, oxygen, or sulfur are plotted by a hatched cross (+).

the diagonal line indicates the discrepancy between the CQEq and the RESP charges. The points in the upper left indicate that the CQEq charges are underestimated with respect to the RESP charges, whereas those in the lower right indicate the opposite.

In Figure 2(a), the partial charges on hydrogen atoms connected to carbon atoms (H_C) were plotted by crosses (×) and those connected to nitrogen, oxygen, or sulfur atoms ($H_{O,N}$) were plotted by hatched crosses (+). It is clearly seen in Figure 2(a) that the group of × symbols and that of + symbols are separately clustered, distinguishing the two types of hydrogen atoms. The partial charges on H_C atoms are estimated to be higher than the RESP charges and those of $H_{O,N}$ are estimated to be lower. One of the reasons for the overestimation of the partial charge on H_C is that the electronegativity of hydrogen atoms (χ_H^0) is too small or that of carbon (χ_C^0) is too large. If the difference between χ_H^0 and χ_C^0 is made smaller, the electron flow from the hydrogen atom to the carbon atom will decrease and the overestimation of partial charge on the hydrogen atom will be improved. On the other hand, the underestimation of partial charges on the hydrogen atoms

connected to nitrogen or oxygen will be improved if the difference between χ^0_{H} and χ^0_{O} or χ^0_{N} is made larger. The requirements for the electronegativity would differ according to the types of hydrogen. If the same set of parameters is given to the two types of hydrogen atoms, it would be difficult to obtain adequate values for partial charges. To obtain accurate partial charges, it would be necessary for the two types of hydrogen atoms to be considered and for different parameters to be assigned to each atom type. Figure 2(b) shows that the partial charges on carbon atoms tend to be underestimated, but the correlation between the CQEq and RESP charges is relatively good. Partial charges on oxygen atoms are underestimated in absolute value, as seen from Figure 2(d), but there is some correlation between the CQEq and RESP charges. The CQEq charges on nitrogen atoms have less variety than do the RESP charges and are almost all in the range from -0.3 to -0.2, as seen from Figure 2(c). These results suggest that the CQEq with the present parameter set may not so accurately describe the charge distributions of amino acid molecules.

One of the ways to improve the CUFF is to refine the CQEq parameters χ^0 and J^0_{ii} to reproduce the electrostatic potential. In fact, there are two types of hydrogen atoms in biomolecules. One is connected to carbon atoms and the other is connected to nitrogen or oxygen atoms. The former tends to be assigned a larger charge and the latter tends to have a smaller charge, when the same electronegativity and hardness were used for these two types of hydrogen atoms. This indicates that different parameters should be used for these two types of hydrogen atoms. We have performed the parameter refinement for DNA base pairs by using the two types of hydrogen atoms [11], and the values of the refined parameters differ clearly between the two types. The improvement of the CUFF in this strategy is now in progress to obtain more accurate charge distribution.

4. Conclusion

We have proposed the consistent charge equilibration (CQEq) method, which employs an identical energy expression for the calculations of both partial charges and electrostatic energy gradient. The CQEq energy term was then combined with a generic force field, universal force field (UFF), to develop the consistent charge equilibration with universal force field (CUFF). To confirm the accuracy of the CUFF, geometry optimizations by the CUFF were performed for a series of amino acid molecules.

Geometry optimizations in the level of the CUFF and HF/6-31G** were performed for the 22 structures of 20 standard amino acid molecules. The optimized geometries by the CUFF were compared with those by HF/6-31G**, showing that the CUFF can well reproduce the HF geometry aside from some flexible dihedral angles.

The partial charges obtained by the CUFF were also compared with the charges derived from the RESP based on the HF/6-31G** electrostatic potential. The partial charges on carbon or oxygen atoms are underestimated in absolute value with the relatively good correlation. The partial charges on nitrogen atoms have less variety than do the RESP charges. The partial charges on hydrogen atoms connected to carbon atoms are overestimated and those connected to nitrogen or oxygen are underestimated. This fact suggests that two types of hydrogen atoms should be considered and the parameters of the CQEq allow for improvement. The refinement of the CQEq parameters is now underway.

This work is supported by Research and Development for Applying Advanced Computational Science and Technology of Japan Science and Technology Corporation (ACT-JST).

References

[1] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179-5197 (1995).
[2] P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *J. Comp. Chem.*, **16**, 1357-1377 (1995).
[3] A. K. Rappé and W. A. Goddard III, *J. Am. Chem. Soc.*, **95**, 3358-3363 (1991).
[4] S. W. Rick, S. J. Stuart and B. J. Berne, *J. Chem. Phys.*, **101**, 6141-6156 (1994).
[5] T. A. Halgren and W. Damm, *Curr. Opin. Struc. Biol.*, **11**, 236-242 (2001).
[6] A. Laio, J. VandeVondele and U. Rothlisberger, *J. Phys. Chem. B*, **106**, 7300-7307 (2002).
[7] O. Kitao and T. Ogawa, *Mol. Phys.*, **101**, 3-17 (2003).
[8] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III and W. M. Skiff, *J. Am. Chem. Soc.*, **114**, 10024-10035 (1992).
[9] R. G. Pearson, *Inorg. Chem.*, **27**, 734-740 (1988).
[10] Gaussian 98 (Revision A.9), M. J. Frisch et al., Gaussian Inc., Pittsburgh PA (1998).
[11] T. Ogawa, O. Kitao, N. Kurita, H. Sekino and S. Tanaka, *Chem. Phys. Lett.*, in press.