Chemical Physics Letters 593 (2014) 165–173

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Octahedral point-charge model and its application to fragment molecular orbital calculations of chemical shifts

Qi Gao [a,b], Satoshi Yokojima [c,d,*], Dmitri G. Fedorov [e], Kazuo Kitaura [f], Minoru Sakurai [b], Shinichiro Nakamura [d]

[a] Mitsubishi Chemical Group Science and Technology Research Center, Inc., 1000 Kamoshida-cho, Aoba-ku, Yokohama 227-8502, Japan
[b] Center for Biological Resources and Informatics, Tokyo Institute of Technology, 4259 Nagatsuda-cho, Midori-ku, Yokohama 226-8501, Japan
[c] School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan
[d] RIKEN Research Cluster for Innovation, Nakamura Laboratory, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[e] NRI, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
[f] Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Hyogo 657-8501, Japan

ARTICLE INFO

Article history:
Received 1 November 2013
In final form 3 January 2014
Available online 8 January 2014

ABSTRACT

To obtain chemical shifts in large molecular systems accurately with a low computational cost, we developed an octahedral point-charge model that mimics the electrostatic potential due to charge distributions. The point-charges in this model are defined to reproduce the multipole moments calculated using the revised distributed multipole analysis. The accuracy of charge representations was tested on formamide. The octahedral point-charge model was used in the fragment molecular orbital method and applied to NMR calculation of a β-sheet polypeptide. The maximum errors relative to conventional *ab initio* NMR calculation were 0.13, 0.73, and 0.09 ppm for ^{13}C, ^{15}N, and ^{1}H, respectively.

1. Introduction

In nuclear magnetic resonance (NMR) spectroscopy, the term chemical shift refers to the difference between the resonance frequency of a nucleus and that of a standard nucleus. The chemical shift of a nucleus in a molecule depends on the conformation and environment of that molecule because chemical shift is sensitive to the group susceptibility, electric field, charge transfer, and solvent effects, all of which depend on the molecular conformation and environment. Numerous experimental studies of these effects on the chemical shifts of amino acids in proteins have been performed for thoroughly investigated protein structures [1–3]. On the other hand, it is important to note that the experimental assignment of chemical shifts in proteins is not an easy task in some cases. For example, it is difficult to assign the peaks of aromatic residues (e.g., tyrosine and phenylalanine) because some of their shifts are very similar in magnitude [4]. Quantum-chemical approaches can be used to predict the values of chemical shifts [5,6].

In practice, however, it is difficult to apply conventional quantum-chemical approaches for the calculation of the chemical shifts of proteins because the size of proteins makes the computational cost very high. Proteins are also observed experimentally in condensed phases, and the effect of the surrounding medium should be considered if the chemical shifts are to be obtained accurately. Including the effect of a protein's environment further increases the computational cost.

Hybrid quantum mechanics/molecular mechanics (QM/MM) methods [7] typically rely on a point charge distribution for electrostatic embedding. The advantage of these methods are that the cost of chemical shift calculations does not increase significantly when a distribution of point charges in the MM region is added to the total Hamiltonian of the system. A lot of QM/MM studies for calculating chemical shifts in biomolecular systems have been conducted [8–11], and various methods involving automatic application of QM/MM in the multiple-center fashion have been proposed [12–14]. Flaig et al. have studied the effect of increasing the QM region size [15].

Some studies, however, indicate that several fundamental problems associated with QM/MM limit the accuracy of the calculations of chemical shifts [16,17]. One major problem stems from the failure to reproduce the embedding electrostatic effects accurately. Chemical shifts are quite sensitive to the charge distribution because it affects the magnetic shielding. The charge distribution is determined by the electrostatic field, which should therefore be reproduced accurately if accurate chemical shift values are to be obtained. For instance, instead of using the predetermined point charges for functional groups, Grant et al. used charge distributions derived from density functional theory and

* Corresponding author at: School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan. Tel.: +81 42 676 1529.
E-mail address: yokojima@toyaku.ac.jp (S. Yokojima).

0009-2614/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cplett.2014.01.001

2nd-order Møller–Plesset perturbation theory to calculate ^{13}C chemical shifts in thiocarbonates [16]. Their method reproduced the electrostatic effects accurately and consequently improved the principle values of the shielding tensor by about 5 ppm [16].

Based on the fragment molecular orbital (FMO) method [18–20], we have proposed a method [21] for calculating NMR chemical shifts that introduces the concept of a merged fragment. Sekino et al. on the other hand, developed a model for FMO calculation of NMR shifts by using the many-body expansion [22], and Frank et al. developed a chemical shift model with the adjustable density matrix assembler approach [23]. The FMO method is a fragment-based method [24] in which fragments and their dimers are calculated self-consistently in the presence of the electrostatic potential (ESP) derived from quantum-chemical calculations.

Two FMO models, differing in the method used to compute the ESP, have been proposed [21]. Model I relies on the atomic point charges (typically, from the Mulliken population analysis), whereas model II is formulated in terms of charge density distributions directly; the point charges and charge distributions are both obtained from FMO calculations. Chemical shifts are sensitive to the ESP, and model II is in general more accurate although more expensive. In addition, only the continuous set of gauge transformations (CSGT) [25] is applicable to the chemical shift calculations in model II, whereas both gauge-including atomic orbitals (GIAO) [26,27] and CSGT are applicable to those in model I [28]. The aim of the work reported in this Letter was to develop a model that yields an adequate description of the ESP with a low computational cost.

The Mulliken population analysis [29] can be inaccurate because the point charges of atoms are determined by dividing the orbital population between the atoms into equal parts [30]. These point charges are not well suited for accurately evaluating the surrounding molecular ESP, and several other point-charge models have been developed [31–33]. Distributed multipole analysis (DMA) [31] reproduces the ESP accurately. The important feature of this method is that it uses molecular multipole expansions at atomic positions, whereas other popular point-charge models such as the generalized atomic polar tensor (GAPT) [32] and molecular electrostatic potential derived (MEP) [33] methods use only atomic charges (i.e., monopoles at atomic positions). It has been shown [31] that DMA with multipole expansion up to quadrupoles reproduces the ESP accurately. A better representation of ESP can be obtained by using higher-order multipoles (including octapoles is usually good enough) [34].

In this Letter we propose a method for condensing DMA multipoles into a set of several charges per atom. We use point charges rather than a multipole because there is so far no program that can calculate NMR chemical shift using multipole moments. By using point charges to mimic multipole moments, we can calculate NMR chemical shifts with better ESP but with the existing program. Our method has three steps. Step I: obtain the charge density distributions of the whole system by quantum-chemical calculations. Step II: for each atom, calculate DMA multipoles up to quadrupoles based on the charge density distributions. Step III: generate charges at the center and vertices of an octahedron around each atom. These seven charges are defined to reproduce the DMA multipoles. The octahedral point-charge model is used in the FMO method and applied to a polypeptide.

2. Methods

2.1. FMO1(merged) method

We now summarize the details of the FMO1(merged) method [21,28,35] applicable to chemical shift calculations in large molecular systems such as proteins. The molecular system is divided into fragments (also called monomers) each consisting of two adjacent amino acid residues. The energy of each monomer is then obtained by using the embedding ESP $V^x_{\mu\nu}$ which is due to the ESP from all the other monomers.

$$V^x_{\mu\nu} = \sum_{K\neq x}\left\{\sum_{A\in K}\left\langle\mu\left|-\frac{1}{4\pi\varepsilon_0}\frac{Z_Ae^2}{|\boldsymbol{r}-\boldsymbol{R}_A|}\right|\nu\right\rangle + \frac{e^2}{4\pi\varepsilon_0}\sum_{\rho\sigma\in K}D^K_{\rho\sigma}(\mu\nu|\rho\sigma)\right\}, \tag{1}$$

where the superscript x represents a monomer ($x = I$) or a dimer ($x = IJ$); μ, ν, ρ, and σ denote atomic orbitals; Z_A and $\boldsymbol{R}_A$ are respectively the charge and position of atom A; and D^k is the density matrix of monomer K. A merged fragment is then constructed by assembling all monomers within a cutoff distance L_{cutoff} from the center of mass $\boldsymbol{R}_a$ of the residue for which chemical shifts are calculated. The distance R_{Ia} between monomer I and residue a is defined as the closest distance between atoms in I and the center of mass of a as follows:

$$R_{Ia} = \min_{i\in I}\{|\boldsymbol{R}_i - \boldsymbol{R}_a|\}, \tag{2}$$

where $\boldsymbol{R}_i$ is the position of atom i in monomer I. In model I the ESP surrounding the merged fragment Q is then calculated using point charges of surrounding monomers:

$$V^{Q(L_{\text{cutoff}})}_{\mu\nu} = \sum_{K\notin Q(L_{\text{cutoff}})}\left\{\sum_{A\in K}\left\langle\mu\left|-\frac{1}{4\pi\varepsilon_0}\frac{Z_Ae^2}{|\boldsymbol{r}-\boldsymbol{R}_A|}\right|\nu\right\rangle + \sum_{A\in K}\left\langle\mu\left|-\frac{1}{4\pi\varepsilon_0}\frac{Z^K_Ae^2}{|\boldsymbol{r}-\boldsymbol{R}_A|}\right|\nu\right\rangle\right\}, \tag{3}$$

where Z^K_A is the atomic population on atom A in monomer K. In model II the surrounding ESP is computed using the density matrices:

$$V^{Q(L_{\text{cutoff}})}_{\mu\nu} = \sum_{K\notin Q(L_{\text{cutoff}})}\left\{\sum_{A\in K}\left\langle\mu\left|-\frac{1}{4\pi\varepsilon_0}\frac{Z_Ae^2}{|\boldsymbol{r}-\boldsymbol{R}_A|}\right|\nu\right\rangle + \frac{e^2}{4\pi\varepsilon_0}\sum_{\rho\sigma\in K}D^K_{\rho\sigma}(\mu\nu|\rho\sigma)\right\}. \tag{4}$$

Finally, chemical shifts of the atoms in the merged fragment are calculated by solving coupled perturbed Hartree–Fock (CPHF) equations in the presence of the ESPs defined in Eqs. (3) or (4).

We previously found that the chemical shifts obtained using model I are less computationally expensive but their accuracy is unsatisfactory (e.g., the maximum error of the ^{15}N isotropic shielding constant is 2.21 ppm) [21]. The results obtained using model I are inaccurate because surrounding ESPs are poorly described by Mulliken charges. These results indicate the elaboration of assignment of charges to atoms is a key point in reproducing ESPs accurately.

2.2. The distributed multipole analysis (DMA) method

When the ESP at position $\boldsymbol{R}$ is calculated from the charge density distribution $\rho(\boldsymbol{r})$ of a molecule, it is given by

$$V(\boldsymbol{R}) = \frac{e}{4\pi\varepsilon_0}\int_{-\infty}^{\infty}\frac{\rho(\boldsymbol{r})}{|\boldsymbol{R}-\boldsymbol{r}|}d\boldsymbol{r} \tag{5}$$

or, for point charges representing $\rho(\boldsymbol{r})$, by

$$V(\boldsymbol{R}) = \sum_a\frac{e_a}{4\pi\varepsilon_0|\boldsymbol{R}-\boldsymbol{r}^a|}, \tag{6}$$

where e_a is the charge at position $\boldsymbol{r}^a$. When all $|\boldsymbol{R}-\boldsymbol{r}^a|$ are large, Eq. (6) can be expanded as a Taylor series. Assuming for simplicity one atomic center, we have

$$V(\boldsymbol{R}) = e\int_{-\infty}^{\infty}\frac{\rho(\boldsymbol{r})}{4\pi\varepsilon_0}\left\{\frac{1}{R} - \sum_\alpha r_\alpha\frac{\partial}{\partial R_\alpha}\left(\frac{1}{R}\right) + \frac{1}{2}\sum_{\alpha,\beta}r_\alpha r_\beta\frac{\partial^2}{\partial R_\alpha\partial R_\beta}\left(\frac{1}{R}\right) - \cdots\right\}d\boldsymbol{r}, \tag{7}$$