STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LIII, 2, 2008

PROGRESS IN MODELLING ELECTROSTATICS AND POLARIZATION THROUGH EFFECTIVE MULTIPOLES AND INDUCED CHARGES

C. J. R. ILLIGWORTH[a], P. J. WINN[a], G. G. FERENCZY[b], C. A. REYNOLDS*

ABSTRACT. Here we present a method for modeling polarization in hybrid QM/MM calculations. The method, which expresses the induced dipoles as a set of 'induced' charges, is based on the induced dipole approach and methodology for calculating potential-derived point charges from distributed multipole series. Here we assess the importance of explicit polarization in the classical part of a QM/MM system with regard to improving the classical description and the consequent effects on the quantum description. The main advantages of the induced charge approach are that the method is readily interfaced with quantum mechanical methods and that induced charges are more readily interpreted than induced dipoles. The ease of interpretation is illustrated by analysis of the charges involved in dimeric and trimeric hydrogen bonded systems. The method has been validated using two energy decomposition approaches, which show that MM polarization makes a significant and reliable contribution to the QM – MM interaction energy in a hybrid system. The method has been modified to assess the likely effect of QM and MM polarization on docking. Since the lack of polarization is only one of a number of deficiencies in current docking approaches, we have also used connectivity to assess alternative docking poses.

BACKGROUND

Molecular modelling has evolved into a highly versatile research tool in chemistry, biology and materials science. At its heart lies the atomic charge, which contributes to the energetics of the system through Coulomb's law, and while the concept of the atomic charge is relatively simple, implementation is a little more difficult because the atomic charge is not clearly defined in quantum mechanics. Consequently, the early approaches to atomic charges through population analysis of the quantum mechanical wavefunction [1] were not only arbitrary but also failed to yield accurate quantitative results when compared against the more rigorous methods devised (including our own [2]). The electrostatic potential is, however, precisely defined [3], and so the determination of potential derived charges by Kollman [4] for use in force

[a] *Department of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, CO1 1NW*
[b] *Department of Inorganic Chemistry and Department of Chemical Information Technology, Budapest University of Technology and Economics, Szent Gellért tér 4, H-1111 Budapest, Hungary*

fields [5], following earlier work by Momany [6], was a significant milestone, enabling applications, such as analysis of the electrostatic potential along the grooves of DNA [7], to proceed from a more rigorous foundation [7]. In parallel with these developments, Stone has presented a far more rigorous approach to electrostatics and polarization [8]. While many of these developments provide a benchmark against which other methods can be tested, they have not found widespread use in biological applications. Nevertheless, Stone's distributed multipole analysis (DMA) [9] is extremely useful. Here we provide a perspective on developments at Essex relevant to modelling biological systems that have been heavily influenced by Stones' DMA. The references cited herein provide a more balanced reference to the many other authors who have inspired developments in this field.

POTENTIAL DERIVED CHARGES

During the 1980s however, ab initio calculations on large molecules were difficult. For this reason, the first program to determine semi-empirical potential derived charges was developed [10], later released through Oxford Molecular. Prior to this, the only option for using semi empirical charges was to use Coulson charges [11], which were essentially similar to Mulliken charges [1]. The essential step in the determination of potential derived charges is usually the constrained minimization of the difference between the quantum mechanical and classical molecular electrostatic potential over a numerical grid at and or beyond the surface of the molecule. This method of determining atomic charges formed the basis of much of our work on calculating molecular properties by free energy difference simulations [12-19].

The use of the numerical grid can lead to asymmetries in the atomic charges of symmetrically equivalent atoms and to poorly defined charges on buried atoms that are well removed from the surface of the molecule [20]. By expanding the wavefunction as a distributed multipole analysis (DMA) and exploiting the properties of spherical harmonics, Ferenczy was able to replace the numerical summation over the grid by an analytical integral over a spherical shell surrounding the centre of each multipole series [21]. This method forms the basis of much of our work on modelling polarization through induced charges [20, 22-27] and does not suffer from the known problems of standard potential derived charges [20].

The availability of readily determined semi-empirical potential derived charges enabled a number of molecular mechanics drug-enzyme studies to be completed, most notably the binding of bioreductive DNA minor groove binding ligands [28].

Free energy simulations, however, provided a greater challenge. For example, while determination of the relative partition coefficients of methanol and ethanol could be carried out relatively easily [15], the corresponding

ethanol-propanol system could only be studied [29] through the use of potential derived charges that were valid across a range of conformations [30,31] (this was achieved by weighting the elements of the potential by the Boltzmann population of the conformations from which they were derived). These multiple conformation charges later formed the basis of widely used AMBER force fields [32]. The early potential derived charges were determined using Hartree-Fock wave functions. The fact that the Hartree-Fock wavefunction renders a molecule slightly too polar is fortuitous as it has the effect of including polarization implicitly. The resultant error in the hydration energy arising from a molecule being too polar can be estimated by comparison against charge distributions derived from correlated wavefunctions [22], and this is indeed of the same order (~10-15%) as the polarization energy [22], implying that biological force fields should be based on correlated wavefunctions if they include polarization explicitly [22].

One advantage of the multipole-derived charges, that has not been extensively exploited to date, is the possibility of determining effective multipoles, i.e. charges plus dipoles or charges, dipoles and quadrupoles as these are highly efficient when compared to the underlying DMA in that charges plus dipoles from effective multipoles are roughly equivalent (in determining interaction energies) to octopoles or hexadecapoles from the DMA [23].

POLARIZATION

The classical way to include polarization in a molecular mechanics force field is to use induced dipoles.

$$\boldsymbol{\mu} = \alpha \boldsymbol{E}$$

where $\boldsymbol{\mu}$ is the induced dipole at a given atom, α is the isotropic polarizability of the atom and $\boldsymbol{E}$ is the electric field at the atom. The electric field, $\boldsymbol{E}$, depends on all the other atomic charges and induced dipoles and so the equations are usually solved iteratively. The method is complicated and computationally expensive as the induced dipole is a vector and hence the energy evaluations are inherently more complex. The method is also intellectually inconsistent as the smaller polarization component is handled at the higher dipole (rank 1) level whereas the more important electrostatic component is handled at the lower atomic charge (rank 0) level.

However, the software used in the determination of the multipole-derived charges can be exploited to model polarization since the ability to convert a multipole series into a distributed charge distribution can be used to convert an induced dipole (which is a very simple multipole series) into a set of induced charges on the given atom hosting the induced dipole and those atoms connected to it. There are four distinct advantages of this approach. Firstly,

the induced charges are easier to interpret than induced dipoles and so have more value in aiding chemical intuition [20, 24, 26]. Secondly, induced charges are easier to implement in a variety of other programs and so they can be use to implement for example MM polarization in hybrid QM/MM methods [20, 26]. Thirdly, induced charges are not as computationally expensive as induced dipoles [27] and fourthly, the induced charges are fully consistent with the derivation of most force fields that use quantum mechanical potential derived charges since the induced charges are essentially potential derived charges. For this reason, benchmarking against quantum mechanical calculations has played a key role in developing this approach [20, 24].

The initial implementation of the method focused on a comparison of the induced charges in a number of small molecule complexes (dimers and trimers) with those determined quantum mechanically; considerable quantitative agreement was obtained [24], though this agreement cannot be taken too far as the induced dipole at a point is replaced by a set of charges at the point and its bonded neighbours [20]. Nevertheless, this agreement included reproduction of the slightly enhanced or slightly suppressed induced charges that may occur in cooperative and anti-cooperative hydrogen bonding systems, as illustrated in Figure 1 for water dimers and trimers; these effects have also been reproduced in hybrid QM/MM systems [20].

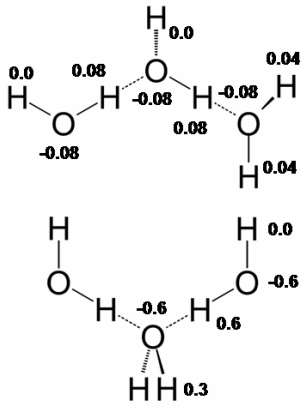

Figure 1. A comparison of the induced charges in a cooperative water trimer (top) and an anti-cooperative water trimer (bottom). The magnitude of the (idealized) induced charges is higher in the cooperative system.

Our first QM/MM implementation of induced charges was used in an investigation into the non-bonded iodine-oxygen interaction in dimethyl-2-iodobenzoylphosphonate, since the oxygen and iodine atoms both approach very close to each other, despite being on the same side of the periodic table

and hence both formally electronegative; it was concluded that the high polarizability of the iodine greatly softened the iodine-oxygen electrostatic repulsion [26]. The potential of the method can be illustrated by two further applications of the QM/MM(IC) method, i.e. the inclusion of induced charges in the molecular mechanics part of a hybrid QM/MM calculation. Within QM/MM(IC) methods, the determination of the field at the MM atoms is more efficient if the wavefunction is expanded as a DMA since the field can then be calculated using Orient [33].

The first involved a study of the contribution of enzyme polarization to transition state stabilization within the enzyme chorismate mutase. Using the previously published structures, we were able to show that the extent of polarization within the enzyme transition state was very similar to that in the reactant geometry [34], and hence that in these particular geometries polarization did not contribute to transition state polarization.

The second study assessed the potential of including MM polarization in docking. Docking programs are design to be computationally efficient rather than accurate and so generally use empirical charges such as those proposed by Gasteiger [35]. Cho et al. have shown that there are distinct advantages for increasing accuracy in using potential derived charges, particularly when the ligand is polarized by the enzyme [36]. At a simple level, this approach requires knowledge of the final answer so that the ligand can be correctly polarized by the enzyme, but the authors have developed a survival of the fittest approach that does not require knowledge of the answer a priori. Against, this background, we have extended this approach to include polarization of the enzyme by the ligand. The ligand was treated quantum mechanically to avoid parameterization issues [37], but in principle, the way to extend this to a fully classical system has already been determined [24]. The method was tested on 10 systems that had previously presented difficulties in Autodock docking experiments. In some cases clear improvements were observed. For example, with the protein - ligand complex from PDB structure 9AAT, the RMS for docking 4'-deoxy-4'-aminopyridoxal-5'-phosphate to mitochondrial aspartate aminotransferase in autodock using the default Gasteiger electrostatics was 7 Å, but with polarization fully included the RMS dropped to 0.9 Å. Not all of the systems showed such clear improvements but clear criteria for assessing when an improvement in the methodology gave rise to an improvement in the results could be determined by analysis of the clustering results.

The previous two applications involved single point determinations of the induced charges since the multipole derived charge methodology [21] cannot readily be extended to include derivatives of the charges. Consequently, a variation on the method has been developed that is suitable for implementation in molecular dynamics simulations [27]. Using only one adjustable parameter

and by modifying the TIP3P water model to include induced charges, we were able to reproduce the oxygen-oxygen radial distribution function for water remarkably well. Moreover, the method is actually based on determination of the potential rather than the field. This article highlighted a problem with induced charges. When the field is perpendicular to the plane, then without the addition of off-atom sites (which can be readily added) the model is not sufficiently flexible to fully reproduce the polarization. However, in a further application we have shown that the extent of polarization energy determined using the induced charge model is generally in line with that determined using a quantum mechanical energy decomposition analysis with medium sized basis sets. In other words, the small amount of polarization energy lost by using induced charges (without off-atom sites) rather than induced dipoles fortuitously corresponds to the amount of polarization energy lost when using a medium sized basis set rather than a large basis set with multiple diffuse functions). The polarization energy determined using the induced dipole model is generally only comparable to that determined using a quantum mechanical energy decomposition analysis when very large basis sets are used that include diffuse functions. This suggests that induced charges are more appropriate for inclusion in QM/MM calculations on biological systems where very large basis sets are generally too expensive [38].

Throughout the development of these methods the possibility of comparison against more rigorous methods such as quantum mechanical calculations, QM/MM calculations or DMA-derived energies has played a key role.

REFERENCES

1. R.S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833.
2. C.A. Reynolds, G.G. Ferenczy, W.G. Richards, *Theochem-J. Mol. Struct.* **1992**, *88*, 249.
3. R. Bonaccorsi, E. Scrocco, J. Tomasi, *J. Chem. Phys.* **1970**, *52*, 5270.
4. U.C. Singh, P.A. Kollman, *J. Comput. Chem* **1984**, *5*, 129.
5. S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P.A. Weiner, *J. Amer. Chem. Soc.* **1984**, *106*, 765.
6. F.A. Momany, *J. Phys. Chem.* **1978**, *82*, 592.
7. J.M. Burridge, P. Quarendon, C.A. Reynolds, P.J. Goodford, *J. Mol. Graphics.* **1987**, *5*, 165.
8. A.J. Stone, The theory of intermolecular forces. Oxford University Press, 1996.
9. A.J. Stone, *Chem. Phys. Lett.* **1981**, *83*, 233.
10. G.G. Ferenczy, C.A. Reynolds, W.G. Richards, *J. Comput. Chem.* **1990**, *11*, 159.

11. B.H. Chirguin, C.A. Coulson, *Proc. R. Soc. London. Ser. A.* **1950**, *201*, 196.
12. S.G. Lister, C.A. Reynolds, W.G. Richards, *Int. J. Quantum Chem.* **1992**, *41*, 293.
13. C.A. Reynolds, P.M. King, W.G. Richards, *Chem. Comm.* **1988**, 1434.
14. C.A. Reynolds, P.M. King, W.G. Richards, *Nature* **1988**, *334*, 80.
15. J.W. Essex, C.A. Reynolds, W.G. Richards, *Chem. Comm.* **1989**, 1152.
16. M.C. Menziani, C.A. Reynolds, W.G. Richards, *Chem. Comm.* **1989**, 853.
17. R.G. Compton, P.M. King, C.A. Reynolds, W.G. Richards, A.M. Waller, *J. Electroanal. Chem.* **1989**, *258*, 79.
18. P.M. King, C.A. Reynolds, W.G. Richards, *J. Mol. Struct. (THEOCHEM)* **1990**, *67*, 205.
19. P.M. King, C.A. Renolds, J.W. Essex, G.A. Worth, W.G. Richards, *Mol. Sim.* **1990**, *5*, 265.
20. C.J. Illingworth, S.R. Gooding, P.J. Winn, G.A. Jones, G.G. Ferenczy, C.A. Reynolds, *J. Phys. Chem. A* **2006**, *110*, 6487.
21. G.G. Ferenczy, *J. Comput. Chem.* **1991**, *12*, 913.
22. P.J. Winn, G.G. Ferenczy, C.A. Reynolds, *J. Phys. Chem. A* **1997**, *101*, 5437.
23. G.G. Ferenczy, P.J. Winn, C.A. Reynolds, *J. Phys. Chem. A* **1997**, *101*, 5446.
24. P.J. Winn, G.G. Ferenczy, C.A. Reynolds, *J. Comput. Chem.* **1999**, *20*, 704.
25. J.H. Wu, P.J. Winn, G.G. Ferenczy, C.A. Reynolds, *Int. J. Quant. Chem.* **1999**, *73*, 229.
26. S.R. Gooding, P.J. Winn, R.I. Maurer, G.G. Ferenczy, J.R. Miller, J.E. Harris, D.V. Griffiths, C.A. Reynolds, *J. Comput. Chem.* **2000**, *21*, 478.
27. G.G. Ferenczy, C.A. Reynolds, *J. Phys. Chem. A* **2001**, *105*, 11470.
28. I.S. Haworth, C. Burt, F. Gago, C.A. Reynolds, W.G.A. Richards, *Anti-Cancer Drug Des.* **1991**, *6*, 59.
29. J.W. Essex, C.A. Reynolds, W.G. Richards, *J. Amer. Chem. Soc.* **1992**, *114*, 3634.
30. C.A. Reynolds, J.W. Essex, W.G. Richards, *J. Amer. Chem. Soc.* **1992**, *114*, 9075.
31. C.A. Reynolds, J.W. Essex, W.G. Richards, *Chem. Phys. Lett.* **1992**, *199*, 257.
32. 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, P.A. Kollman, *J. Amer. Chem. Soc.* **1995**, *117*, 5179.
33. A.J. Stone, A. Dullweber, O. Engkvist, E. Fraschini, M.P. Hodges, A.W. Meredith, D.R. Nutt, P.L.A. Popelier, D.J. Wales, (2002) Orient: a program for studying interactions between molecules, version 4.5', University of Cambridge, Enquiries to AJ Stone, ajs1@cam.ac.uk. 2006.
34. C.J.R. Illingworth, K.E. Parkes, C.R. Snell, S. Marti, V. Moliner, C.A. Reynolds, *J. Phys. Chem. A., in press,* **2008**.
35. J. Gasteiger, M. Marsili, *Tet. Lett.* **1978**, 3181.
36. A.E. Cho, V. Guallar, B.J. Berne, R. Friesner, *J Comput. Chem.* **2005**, *26*, 915.
37. C.J.R. Illingworth, G.M. Morris, K.E.B. Parkes, C.R. Snell, C.A. Reynolds, *J. Phys. Chem. A., in press*.
38. C.J.R. Illingworth, K.E.B. Parkes, C.R. Snell, G.G. Ferenczy, C.A. Reynolds, *J. Phys. Chem. A., in press*.