Developing a Force Field for the Transition State of the Aldol Reaction of Enolborinates: Evaluation of the Use of Fixed Point Charges.

Anna Bernardi,*a Andrea Cassinari,a Angiolina Comotti,a Mark Gardner, b Cesare Gennari,a Jonathan M. Goodman,*b and Ian Patersonb

(a) Dipartimento di Chlmica Organica e Industrlate, Universiti di Mllano, Centm CNR per lo Studio deUe *Sostanze* **Organiche Naturali, via** Venezlan **21.20133 Milaao, Italy. (b) University Chemical Laboratory, Lensfeld Road, Cambridge CB2 IEW England, UK.**

(Received in UK 20 March 1992)

Abstract: *A method for assigning atom-cmtred point charges, within the constraints of the MacroModel implementation of the MM2* force *field, is described. The* method has *been applied to obtain charges with sensible magnitudes for the transition structures of boron enolate aldol reactions. The effect of these new charges on the aldolforcefield model is discussed. and it is concluded that* atom-centred *point charges do not help in the analysis of this transition state, and a* **more** *flexible system for the treatment* of electrostatics **is** *essential.*

The boron aldol reaction is one of the most powerful methods available to the organic chemist for the control of acyclic stereochemistry. This includes asymmetric aldol reactions under either substrate control using chiral ketones, or reagent control using chiral ligands attached to boron **(Scheme 1).**

Scheme 1

As part of a programme to analyse the origins of the stereocontrol, we have developed a force field model for the aldol reaction of ketone derived enolborinates with aldehydes (here referred to as *FFZ,* details of which have already been published^{1,2}). This force field is based on $MM2³$ and uses new parameters derived from *ab initio* calculations on the cyclic transition structures (chairs and boats) and from trial and error optimisation. It successfully reproduces both the relative energies and the geometries of simple unsubstituted and monosubstituted *ab initio* transition structures **(Figure 1,** structures **1** and **2,** with mono-methyl substitution in each of the possible positions, giving sixteen structures in tota14) with quite good agreement. The force field also reproduces the experimental syn : *anti ratios* for the aldol reactions of simple Z and E substituted

enolborinates from ethyl ketones with aldehydes, and the aldehyde si : re selectivity for the syn selective aldol reactions of a range of chiral Z enolborinates.

In these cases, the force field predicts that only chair-like transition structures are important.^{1,2} The force field fails to reproduce quantitatively some experimental results with unsubstituted enolborinates, i.e. \mathbb{R}^2 $=$ H in **Scheme 1.** In our previous paper,¹ this was attributed to the neglect of a third accessible transition structure, boat B (3), during the parametrization process. Such a structure lies only 1.7 kcal mol⁻¹ (RHF/3-21G) above the global minimum, boat A (2), in the unsubstituted case, and therefore may be important in determining the course of some reactions. This structure has also been found by F. Bernardi and co-workers, at the MC-SCF/3-21G level.⁵ Using this higher level of theory, the energy of 3 relative to the lowest energy transition structure 2 is calculated to be 2.4 kcal mol-I.

To assess the importance of boat 8, the geometries of six structures in addition to the original sixteen were calculated (RHF/3-21G), each bearing a methyl substituent on a different position. Boat B was found to be an important transition structure in all cases examined, except for the addition of the Z enol borinate of propionaldehyde to formaldehyde. We then set about parameterizing a force field which would describe these new structures, as well as the original sixteen. Larger barriers had to be introduced to ensure the existence of minima, which made it more difficult to tune the force field to the presence of individual substituents. As a result, a field was obtained (here referred to as FF2) whose fit to the *ab initio* data was an order of magnitude worse than that of the previous field.⁶ Reproduction of the experimental data was also not satisfactory. This could be due to two problems: (i) the trial and error approach to the development of parameters may not be the most effective way of attempting a description of more complex and flexible systems; (ii) the treatment of electrostatics may be inadequate. In the development of a force field, atom centred point charges must be assigned to the atoms of the transition structure. Initially these charges were based on Löwdin analyses of the *ab initio* results, but, as the force field became more sophisticated, this simple method became the weak link in its development. We have investigated better methods of finding atom-centred point charges and report our findings here.

Procedure for Determination of Atom-Centred Point Charges

The model

The MacroModel⁷ implementation of MM2³ was the basis for the model. This treats electrostatics by assigning a fixed partial charge to each atomic centre. The charges on the core atoms of the transition structure are not able to vary with the geometry of the transition structure, *i.e.* the same set of numbers must be used for a chair, a boat A or a boat B. This approximation is widely used, because of its simplicity, but has recently been questioned by several authors.⁸ The charges can vary with substitution on the core, because MacroModel

allows bond dipoles to be defined.9 The MacroModel implementation of MM2 assigns hydrogens partial charges of zero, and this approximation was used in the new force field, in order to give compatibility with the existing force field. The MacroModel default is to use a distance dependent dielectric ($\epsilon_r = r$ in atomic units), which reduces charge-charge interactions with respect to their effects *in vacuo*, and this was used throughout the study.

There are two requirements in selecting suitable charges: (i) it is necessary to determine the best values; (ii) it is important to have quantitative measures of how good these charges are, and of the errors introduced by not allowing change with conformation.

Past work

In our earlier work.l.2.10 methods of summing electron density from ab *initio* calculations and localising it on particular atoms were used. The Mulliken population analysis¹¹ is the simplest of these methods, but it gives very large values for the partial charges, and these are probably not optimal for force fields, although they reproduce the molecular dipoles fairly well. Löwdin analysis¹² gives numbers that seem of a more realistic magnitude, and also generate good values for the total dipole moment of the system. The initial force field, *FF1*, was based on the Löwdin values.

These population analyses assign partial charges to hydrogen atoms, whereas MacroModel usually uses a partial charge of zero for hydrogens. Hence it was decided to assign zero charge to the hydrogen atoms and to add the calculated partial charges of the hydrogen atoms to that of the attached carbon and boron atoms. This gave a model, *FFI,* that fitted well into the existing MM2 force field, but contained unquantifiable errors.

Assigning bond dipoles was also difficult, because ab *initio* calculations on substituted systems give population analyses that differ over the whole molecule, not just on the new substituent and the attached atom. In *FF1*, therefore, the bond dipoles were not set.⁹

A new approach was clearly needed and it was decided to determine partial charges by considering the electrostatic field around the molecule. CHELPG¹³ (Charges from Electrostatic Potential Grid) is a method of obtaining atom-centred partial charges from ab *initio* calculations, which works by a least-squares fit of the electrostatic potential produced by the partial charges to the ab initio-calculated electrostatic potential. With respect to fitting the electrostatic potential over a certain grid of points, it produces the best possible charges. There are two advantages in using CHELPG: (i) the 'goodness' of the charges can be assessed by evaluating the root mean square deviation from the ab *initio* electrostatic potential around the molecule for every set of charges; (ii) the effect of ring substitution in terms of dipoles for the additional bond can be evaluated. This is done simply by constraining all the other charges to a fixed value (the unsubstituted value) and finding which point-centred charges for the substituent and the attached atom gives the best fit to the *ab initio* electrostatic potential. Partial charges are initially assigned for the transition structure with only hydrogen substituents, then the variation of these charges with alkyl substitution is considered, and finally bond dipoles for substituents in each ring position are defined.

The Method

CHELEG was modified to allow atom charges to be constrained. The partial charges of all the hydrogens were constrained to zero, and the partial charges for each of the remaining elements in the allhydrogen transition structures were calculated, using the electrostatic potential calculated by GAUSSIAN8214 at the RHF/3-21G¹⁵ level of theory.^{16,19} The averages of these charges were used as the basis for the methyl substituted cases. All the charges were constrained to the average all-hydrogen transition structure values, except for the carbon of the methyl group and the attached atom. The total charge of the system was constrained to zero. This gave values for the partial charges of these two atoms in each substituted transition structure. The dipole for each substituted bond was then calculated, and parameters for the alteration of charge due to substitution were found, within the limits set by MM2 in its MacroModel implementation.

The Results

Charges

There are three different geometries for the unsubstituted transition structures **(Figure 1)** and so three different sets of charges were calculated (Table **1). The standad** deviation of these numbers is quite large. This suggests that the different transition structures correspond to slightly different points on the reaction coordinate and that the charges may not be transferable. Within MacroModel, only one set of charges can be used and the average of these numbers might lead to large errors. However, we need not be satisfied with a qualitative estimate of these errors, because the root mean square deviation from the electrostatic potential around the molecule provides a quantitative measure of how good the charges are. CHELPG was used to quantify how good the average charges were for each of the transition structures and these results are shown in **Table 2.** Fitting the electrostatic potential with the average charges or with the individually optimized charges gave RMS deviations that were rather similar, and so this approximation seemed sufficiently good to justify continuing this approach. A referee commented that treating the forming carbon-carbon bond as an actual bond means that the electrostatic interaction of these two carbon atoms is neglected. We think that this is a reasonable model, both because this interaction will be implicitly included in the bond stretching and angle bending terms, and also because the small partial charge on C_1 makes this a minor interaction (**Table 1**).

Table 1: Charges (atomic units)

 $(0.5 \text{ Å grid}; ca 1960 points; average RMS = 0.065; all H constrained to 0; charges in atomic units)$

| RMS deviation from | 1 chair | 2 boat A | 3 boat B | Average for three | |
|-------------------------|---------|----------|--------------|-----------------------|--|
| electrostatic potential | | | | transition structures | |
| individual optimisation | 0.0663 | 0.0637 | 0.0648 | 0.065 | |
| average charges | 0.0703 | 0.0665 | 0.0657 | 0.0675 | |

Table 2: RMS deviation from electrostatic potential (atomic units)

The dipoles for the mono-substituted transition structures were calculated next, as described in the method section. The charges found for the substituents are given in **Table 3.** In most cases, the charges are similar for the chair and the two boats, and so it seemed reasonable to use the average value in the force field.

Methyl substitution on boron posed a problem, because systems of synthetic interest have two alkyl groups on the boron and the *ab initio* calculations could not cope with so large a system on the available computers. Fortunately, the effect of replacing one hydrogen with a methyl group hardly alters the charge on the boron **(Table 3). Thus** it is reasonable to assume that replacing a second hydrogen with a methyl group would not have a significant effect on the system, compared to the other approximations in the model. Therefore, the boron-carbon dipole was not set.⁹

Table 3: Calculated charges for the carbon atoms of the substituents $R_1 - R_7$ (atomic units)

| R_4 R_{5} | $Substituent = Me$ | chair | boat A | boat B | average | $\sigma_{\rm n}$ |
|--|--------------------|-----------|--------|-----------|---------|------------------|
| | \mathbf{R}_1 | 0.0555 | 0.0575 | 0.0857 | 0.0662 | 0.013 |
| O R ₂ R_{7} `R, R_3 | R_2 or R_3 | 0.0975 | 0.0475 | 0.0859 | 0.0675 | 0.020 |
| | | 0.0760 | 0.0462 | 0.0517 | | |
| | R_4 or R_5 | -0.0089 | 0.0163 | -0.0057 | 0.0028 | 0.011 |
| | | $-.0028$ | 0.0187 | -0.0006 | | |
| R _s $R_1 - R_7 = H$, | R_6 | 0.0613 | 0.0200 | 0.0384 | | |
| | R ₇ | 0.0669 | 0.0482 | | 0.047 | 0.017 |

unless otherwise stated

The charge assignment for *E* and *Z* enolborinates proved to be more difficult. The force field cannot distinguish between an E methyl group and a Z methyl group, so these must have the same partial charge. However, the *ab initio* calculations and the CHELPG calculations suggest that they should be different, and so it is necessary to use an average. Again, we could quantify the errors in the charges using CI-IELPC and the results are given in **Table 4. The** larger root mean square differences show that the greatest source of error in these new charges is the inability of the program to distinguish E and Z methyl groups.

The charges and dipoles obtained in this way seemed to have reasonable magnitudes and they fitted the *ab initio* electrostatic potentials well. Therefore, we concluded that they might be used to substantially improve our force field for the boron aldol transition structures.

Table 4: RMS deviation from the electrostatic potential using average charges (atomic units)*

a the charges of all core atoms except C₁ were constrained to the unsubstituted average value from **Table 1**

Force Field Results

Trial and error methods were used to develop torsional parameters for these new charges, this time assisted by automatic methods of modifying and testing limited sets of parameters.20 A force field was produced which had the same magnitude of error as *FF2* in fitting the *ab initio* data. However, this field was less good with respect to reproducing the experimental data.

For some substitution patterns, a new boat was found by this force field. This structure was distinct from the two boats that had already been found, and contained a O=C--C=C dihedral angle close to zero. In earlier work,¹ a force field had predicted the presence of a transition structure that was confirmed by *ab initio* calculations. Therefore, an attempt was made to locate this new boat on the RHF/3-21G potential surface, but it could not be found. This suggests, but does not prove, that it does not exist. If this transition structure is important, then *E* enolates would produce some *syn* aldol products. It is found experimentally that the *anti* aldol dominates, so the new boat must be high in energy relative to the other transition structures. Therefore, a reasonable force field should not generate low energy structures corresponding to this boat. This new boat was stabilised by favourable charge-charge interactions, and we could only eliminate it through reducing the charges by 50%, a far more dramatic scaling factor than that suggested by Chirlian and Francl.^{13b} The agreement of this final version (referred to as *FF3) with the ab initio* data is similar to that achieved with *FF2,* while agreement with the experimental data is slightly improved. However, the new field is worse compared to *FFl* with respect to reproducing the experimental data (see **Table 5).** Nevertheless, it is possible to identify and predict highly selective aldol reactions (cf. **Table 5,** entries 4, 5, 9). In these cases, only the chair transition structure is important in determining the reaction stereoselectivity.

Automatic parameter optimisation2oa also gave a better fit to the *ab initio* data when the charges were reduced. A derivative fitting method was also used $20b$ to produce an initial set of torsional parameters for the unsubstituted cases. The charges calculated by electrostatic potential fitting proved to be inferior. However, the situation was improved by reducing the magnitude of the charges, and the best set of parameters was obtained with zero charges.

Table 5: Results from Force Fields

(a) Using "Bu ligands and benzaldehyde in pentane.^{s.m.} (b) Using benzaldehyde in ether.^{s.m.} (c) From **(-)-(lpc)sBOTf (d) using diethylketone.2'c (e) Using "PrCHO. 21c** (9 **See reference2'd (g) From (-)-(lpc)sBCI (h) Experimental anti-syn ratto is 6020, but NMR shows enotate is 60:20 mixture E:Z, so the result from** the *E* enolate is probably 100% anti. ²¹⁶ (j) Using a benzyl ether, in place of the methyl ether. ²¹⁶

There are various possible explanations for these results. The agreement with experiment using the first force field *FFZ may be* fortuitous. This is unlikely, because of the variety of experimental results reproduced, and the chemically intuitive manner in which the force field results can be interpreted.^{1,2} Alternatively, the further efforts to increase the fit to all twenty three *ab initio* calculated structures may have succeeded at the expense of the intervening regions, and so decreased the reliability of the force field, as judged by its ability to reproduce and predict experimental selectivity. In particular, the relative importance of the transition structures derived from the original unsubstituted cases (chair **1,** boat A 2, boat B 3) may not be calculated very well by molecular mechanics. If only the chair transition structure is important, as appears to be the case for Z enolates, then the accuracy of the result will be higher than for cases where two or three transition structures are involved. When the selectivity is low, then the error may be larger than the energy difference between diastereomeric transition structures, and so the wrong sense of asymmetric induction may be predicted by the force field. When the selectivity is high, the same error in the energy will give the right sense of induction.

If the force field predicts high selectivity in a new situation, it is reasonable to expect that the experimental result will also show good selectivity in the same sense. Computer aided design of new chiral ligands for the enolborinate aldol reaction is underway in our laboratories using this hypothesis.

Conclusions

We have investigated three methods for obtaining point charges for the boron aldol transition structures within the constraints of the MacroModel program: Mulliken and Löwdin population analyses, and charges derived from CHELPG. The Mulliken and Löwdin charges are found to be too large for this application and the charges for the hydrogens and their attached atoms have to be added together to give point charges on the first row elements only. The CHELPG method, with constraints, gives exactly the values required: point charges on the first row elements, no point charges on the hydrogens, and dipoles on the methyl substituents. CHELFG charges. therefore, should provide the best procedure for modelling transition structures. Furthermore, they have the additional advantage of quantifying the difference between the point charge model and the *ab initio* electrostatics. The CHELPG calculations quantify the transferability of the partial charges over the different transition structures. This provides the best possible point charges, with respect to the physically significant criterion of the fit to the electrostatic field around the transition structures. Therefore, they should be used, unless there are no torsional parameters that can complement them. For the boron aldol reaction, it appears that there are no torsional parameters that can cope well with these charges.

In developing a force field by trial and error, a subjective assessment of the difficulty of the procedure may prove misleading. In this particular study, however, the trial and error method is evaluated by the derivative fitting method and an automated optimisation process. We conclude that the use of CHELFG charges do not improve the model.

Therefore, we cannot use the most physically-significant charges and also reproduce the *ab initio* calculated geometries and energies of the transition structures. In order to predict and explain experimental results, it is necessary to reproduce these geometries and energies, and so the physical significance of the charges must be abandoned, until more flexible models of electrostatics become available.

For the time being, the force field derived by fitting only fourteen *ab initio* structures and using Uwdin-based charges *(FFI) gives us the* best agreement with experimental data **(Table 5).** This field *can give a* qualitative assessment of whether a reagent is selective, as well as a good quantitative reproduction of the experimental results in the case of highly selective reagents. This force field, therefore, should be of help in designing new chiral ligands for the enolborinate aldol reaction.

Acknowledgements We thank the Commission of the European Community [Grant SC1*.0324.C(JR)], NATO (Grant 0368/88) and The Royal Society for support, A. M. Capelli for helpful discussions, and Clare College, Cambridge for a Research Fellowship (JMG).

References and Notes

- 1. Bemardi, A.; Capelli. A. M.; Gennari, C.; Goodman, J. M.; Paterson, I.; *J. Org. Chem.* 1990,55.3576.
- **2.** Bernardi. A.; Capelli, A. M.; Comotti, A.; Gennari, C.; Gardner, M.; Goodman, J. M.; Paterson, I. *Tetrahedron* **1991,47,3471.**
- **3.** Allinger, N. L. *J.* Am. Chem. Sot. 1977,99,8127.
- **4.** *FFI* was developed from only fourteen of the sixteen possible structures. The missing two structures are the chair and the boat with a methyl group on the less hindered position of boron. It was thought that the geometry of these structures would be very similar to the unsubstituted cases. Recent calculation of the ab *initio* structures shows that they are indeed described well by *FF1*.
- **5.** Bemardi, F.; Robb, M. A.; Suzzi-VaUi, G.; Tagliavini, E.; Trombiii, C.; Umani-Ronchi, A. *J. Org. Gem.* **1991,56,6472.**
- **6.** We used as a measure of the 'goodness' of the field, the sum of the squares of the Boltzmann factors, in order to reduce the importance of the high energy terms.

N is the total number of structures; ΔE_i^{Ψ} is the energy of structure *i* relative to lowest energy *ab initio* structure; ΔE_i^{MM} is the energy of structure *i* relative to lowest energy force field structure. This measure does not take account of differences in the geometry of the structures. However, least squares fits of the force-field minimised structures to the *ab initio* structures showed that the quality of the geometries was similar in all cases. The values of this measure for the three different force fields, averaged over 23 structures *(FF2* and *FF.?)* or 14 structures *(FFl), arc* as follows:

FFI : 0.0015 *FF2* : *0.0350 FF3* : 0.0406

- **7.** Mohamadi. F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comp. Chem.* 1990, 11,440.
- **8.** (a) Bohm, H. J.; Brode, S. *J. Am.* Chem. Sot. 1991,113,7129; (b) Faerman, C. H.; Price, S. L. *J.* Am. Chem. Soc. **1990**, 112, 4915.
- 9. MacroModel uses atom centted point charges to describe the electrostatic properties of a system, but also allows the dipole moment of a bond to be set. If it is set, then this dipole is added to the partial charges already present. The total dipole of the bond is not, therefore, the value to which it is set, unless the two atoms defining the bond have the same charge. If the dipole is not set, then the dipole moment of the bond is not necessarily zero, because the program will use the standard atom centred point charges for the system.
- 10. Goodman, J. M.; Kahn, S. D.; Paterson, I. *J. Org. Chem.* 1990,55,3295.
- 11. Mulliken, R. S. J. Chem. Phys. 1955,23,1833.
- 12. Löwdin, P. O. Phys. Rev. 1955, 97, 1474.
- 13. (a) Singh. U. C.; Kolhnan, P. A. J. *Comp. Chem.* 1984,5, 129; (b) Chirlian, L. E.; Francl, M. M. J. *Comp. Chem.* 1987,8,894; (c) Breneman, C. M.; Wiberg, K. B. J. Camp. Chem. **1990, 11,361.**
- 14. Binkley, J. S.; Frisch, M. J.; DeFrees, D. J.; Raghavacheri, K.; Whiteside, R. A.; Schlegel, R. A.; Fluder, E. M.; Pople, J. A. Carnegie-Mellon University, Pa 15213, U. S. A.
- 15. Binkley, J. S.; Pople, J. A.; Hehre, W. J. J. Am. Chem. Soc. 1980, 102, 939.
- 16. The choice of basis set was largely dictated by the computer power available, and the 3-21G split-valence basis set was chosen. A review by Davidson and Feller¹⁷ concludes that this is adequate for closed shell organic compounds. Chirlian and Francl address this question in detail, $13b$ and show that charges derived from 3-21G correlate well with those derived from larger basis sets, after scaling. In this study, no scaling factor was used. The effect of electron correlation as omitted from this study. A recent study suggests that it should only have a small effect on the charges obtained.¹⁸
- 17. Davidson, E. R.; Feller, D. Chem. Rev. 1986, 86, 681.
- 18. Luque, F. J.; Grozco, M.; Illas, F.; Rubio, J. *J. Am. Chem. Sot.* 1991,113,5203.
- 19. Typical number of points considered: 2000. Atomic radii used: MM2. Scaling factor: 1.0 Grid parameters: spacing = 0.5 Å, maximum distance from a point to the closest atom = 2.8 Å. All the other parameters were the CHELPG default values.
- 20. (a) Gardner, M. *unpublished work;* (b) Goodman, J. M. *unpublished work*
- 21. (a) Evans, D. A.; Nelson, J. V.; Vogel, E.; Taber, T. R. J. Am. Chem. Soc. 1981, 103, 3099; (b) Brown, H. C.; Dhar, R. K.; Bakshi, R. K.; Pandiarjan, P. K.; Singaram, B. J.Am. *Chem. Sot.* 1989,111, 3441; (c) Paterson, I.; Goodman, J. M.; Lister, M. A.; Schumann, R. C.; McClure, C. K.; Norcross, R. D. *Tetrahedron* 1990,46,4663; (d) Duplantier, A. J.; Nantz, M. H.; Roberts, J. C.; Short, R. P.; Somfai, P.; Masamune, S. *Tetrahedron Lett. 1989,30,7357; (e)* Paterson, I.; Goodman, J. M.; Isaka, M. *Tetrahedron* Lett. **1989**, 30, 7121.