

Structure-Property Relationships in Crystal Structures of Molecules with Non-Centrosymmetric Polymorphs

G. J. Tizzard, M. B. Hursthouse

School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK.

The current interest in the mechanics of crystal structure assembly stems from extensive efforts towards the capability to design or predict crystal structures. This project focuses on the particular feature of non-centrosymmetry in some crystal structures and specifically the assembly of non-centrosymmetric structures by both chiral and achiral molecular systems. To give added focus, a collection of polymorphic 'clusters' has been assembled from the Cambridge Structural Database (CSD) each of which comprise members of which at least one is a non-centrosymmetric polymorph.

The study employs a variety of approaches over and above topology and thus molecular properties such as electrostatic charge distributions and organisation of molecular dipoles have been calculated from *ab initio* computer modelling techniques.

In order to facilitate this kind of study, methods are being investigated in which database interrogation and ensuing geometry calculations can be achieved in an automated manner.

INTRODUCTION

There are a number of different attractive forces which determine the packing in molecular crystals and they can be approximately classified as follows: dispersion or London forces, multipolar forces, hydrogen bonding and charge transfer forces. It is the complex interplay of these forces along with repulsion energy which can lead to many local minima in the lattice energy of a crystal which can thus result in polymorphism (i.e. the existence of more than one crystalline form in a substance).

In the study of polymorphism, hydrogen bonds are the highest energy interactions in molecular crystals and thus appear to be the most important attractive force. There is clear evidence that multifunctional molecules (e.g. pharmaceuticals) with multiple H-bonding sites promote polymorphism and that the polymorphism exhibited by these molecules can be ascribed to the different H-bonding topologies. However, polymorphism also occurs in systems without strong hydrogen bonds ($N-H \cdots X$, $O-H \cdots X$, $S-H \cdots X$; $X = N, O, S, F, Cl, Br, I$). In these cases, although H-bonding may still be present in the form of weaker interactions such as $C-H \cdots X$ and $C-H \cdots \Pi$, the overarching importance of hydrogen bonding in defining polymorphism is greatly reduced.

This project is concerned with making a detailed study of the latter type of the above systems. In these systems electrostatic interactions are expected to exert a greater influence on the crystal structure adopted. A particular point of interest is the occurrence of polymorphs, in what

are essentially achiral molecules, that have both centrosymmetric and non-centrosymmetric crystal structures. Interest in the second of these is very important for the development of useful materials with nonlinear optical properties.

SELECTION OF DATASET USING THE CSD

An investigation of the CSD^[1] has been conducted and a collection of 258 polymorphic 'clusters', each with 2 – 4 members with at least one that is non-centrosymmetric has been identified. Unfortunately, there is no simple method to execute this type of investigation with ConQuest, the CSD search engine, and so firstly a text search against "polymorph" (excluding organometallics) was carried out to create an initial dataset of 4703 'hits'. Further refinement of this initial data set, including identification of the polymorphic 'clusters' with non-centrosymmetric members and the discarding of duplicate 'hits' and those with incomplete heavy atom coordinates was carried out manually and resulted in the final dataset described above with 574 'hits'.

The purpose of this search was to construct a dataset of members each with a complete set of crystallographic data that could be investigated using other applications. Therefore, one of the criteria for inclusion in the final dataset was a complete set of atomic coordinates; however, it was also necessary to define as large a range as possible for each polymorphic 'cluster' and thus some members were accepted with incomplete hydrogen atom coordinates. This problem affected 80 members of the final dataset.

USE OF *AB INITIO* COMPUTER MODELLING TECHNIQUES

The molecular modelling package, Spartan '02 for Windows^[2] was used to generate the missing hydrogen atomic coordinates for each of the 80 members of the final dataset. This was accomplished by exporting the crystallographic data from the CSD as a Tripos mol2 file, a copy of which was then imported into Spartan '02. All necessary hydrogens were added and the molecular minimizer function was then used (with all original atoms restrained) to find the minimum energy conformation. This result was then exported as a new Tripos mol2 file.

Unfortunately, Spartan '02 does not save the crystallographic information present in the original mol2 file and so this information has to be manually cut-and-pasted from the original input file. The modelling package Cerius2^[3] was used to translate the data from the mol2 file format to the Shelx ins file format and the now complete set of atomic coordinates was then manually cut-and-pasted back to an original Crystallographic Information File (CIF). The biggest potential problems of this whole procedure are the amount of human intervention required leading to possible errors and due to rounding errors between applications accuracy is only maintained to the third decimal place. This will only be a serious problem for atoms on special positions within the unit cell (i.e. positions in a unit cell that lie on point symmetry elements); where accuracy errors might move these atoms off of special positions.

CALCULATION OF PARTIAL CHARGE DISTRIBUTION AND ELECTROSTATIC POTENTIAL MAPS

The single point energy for each molecule has been calculated using the Hartree-Fock quantum mechanical model with a 6-31G^(*) basis set and from this the charge distribution and electron potential map has been derived. A single point energy calculation was performed using molecular coordinates directly imported from the CSD, i.e. using the molecular conformation adopted in the crystal structure. This is much cheaper computationally than calculating the equilibrium geometry, which may have little relation to that adopted in the crystal.

The Hartree-Fock model is an approximation based ultimately on the Schrödinger equation (as are all quantum mechanical models) and combined with the 6-31G^(*) basis set is

computationally one of the cheapest of these models (a major consideration in its selection) that is reasonably successful at reproducing experimental results.

However, several points should be noted here. Firstly, despite charge distributions on molecules being part of the everyday language of chemists, there is no way to measure them experimentally. The only way to judge the various models is to compare the calculated dipole moments; these reflect the overall charge distribution of the molecule and can be measured experimentally. Secondly, the dipole moments from Hartree-Fock models tend to be 20 to 30% larger than the experimental values^[4]. This is because Hartree-Fock models overestimate the electron concentration and thus charge separation of molecules. Although there are other models that account for electron correlation error e.g. Møller-Plesset models, they are computationally (up to 100 times) more expensive.

RESULTS FROM COMBINATION OF CRYSTAL PACKING DIAGRAMS AND ELECTROSTATIC POTENTIAL MAPS

The electrostatic potential data obtained from Spartan'02 has been used in conjunction with crystal packing diagrams generated with Mercury to generate electrostatic packing diagrams. This has been done by generating the packing diagram for the crystal structure of interest from Mercury and then using the colour-coding from the electrostatic potential map generated from Spartan'02 to overlay the packing diagram. Two examples are shown in the figures overleaf; 5-nitrouracil (CSD code NIMFOE) (figs. 1, 2 and 3) and N,N-dimethyl-8-nitro-naphthaleneamine (CSD code DIWWEL) (figs. 4, 5 and 6). 5-nitrouracil exhibits strong hydrogen bonding whereas N,N-dimethyl-8-nitro-naphthaleneamine does not.

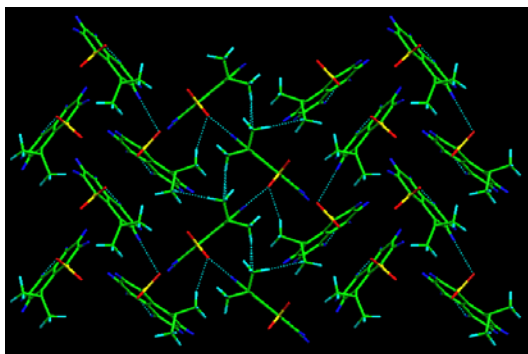


Figure 6. The same view as above (fig. 4) but with the atoms colour-coded as to their electrostatic potential (fig. 5). The network of short-contact distances (sum of VdW radii) is shown; although no measurements are shown as they are identical to those above (fig. 4). As can be seen, electrostatic attraction can account for the interactions between the nitro O and methyl groups and naphthalene 'edge'. However one of the shortest contacts (2.101Å) is between the intermolecular methyl groups and this cannot be accounted for by electrostatic interactions in this model.

As can be seen from the figs. 3 and 6 above, this method of modelling appears useful at highlighting areas of a crystal structure where electrostatic interactions are important (including H-bonding) and those where it is not, e.g. the intermolecular amino methyl interactions of N,N-dimethyl-8-nitro-naphthaleneamine, where perhaps steric considerations may dominate.

Additional points worth noting are as follows. This is a qualitative analysis insofar as the colour-coding for the packing diagrams has been chosen to be close to that of the electrostatic potential map. A more quantitative method would be to base the colours on the point charges of the atoms, however this tends to lead to more extreme results with molecules coloured simply red and blue (and thus a loss of information from the electrostatic potential map). Also, this method is not automated so the colour-coding of the packing diagram has to be done manually which is laborious and slow and obviously not suited to working with large numbers of structures.

FURTHER WORK

From the above method the electrostatic interactions around a given molecule are immediately apparent, however, topological information to allow comparison of different polymorphs is absent. A variety of methods to gain this information are being investigated

which include topological analysis using PLATON^[7], which provides connectivity data about molecular aggregates present based on H-bonding and short contact distances. The application SYNTHON^[8] allows comparison of two polymorphs for the presence of common topological features e.g. dimers, chains, and sheets. The application 'Voronoi'^[9] uses CIF data to generate a unit cell and from the atomic coordinates of the atoms in the unit cell a Voronoi tessellation^[10] is generated and some of the properties of this tessellation are stored and can be compared and analyzed.

The major 'bottlenecks' throughout this project have been workflow related through the transfer of data from one application to another and also from 'driving' the applications to obtain the data. Methods of automation are being investigated including the use of Perl to write data-transfer scripts and spreadsheets to automate calculations. This is with the ultimate aim of providing a complete analysis of the electrostatic interactions of a molecule in the context of its crystal packing as a single callable process.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the EPSRC e-Science programme (GR/R67729, Combechem).

REFERENCES

- [1] F. H. Allen, O. Kennard, *Chem. Des. Autom. News* **1993**, 8, 31.
- [2] 1.0.2 ed., Wavefunction Inc., Irvine, CA, USA, **2002**.
- [3] Accelrys Ltd., 334 Cambridge Scientific Park, Cambridge.
- [4] W. J. Hehre, J. Yu, P. E. Klunzinger, L. Lou, *A Brief Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc., Irvine, CA, **1998**.
- [5] R. S. Gopalan, G. U. Kulkarni, C. N. R. Rao, *ChemPhysChem* **2000**, 1, 127.
- [6] M. Egli, J. D. Wallis, J. D. Dunitz, *Helv. Chim. Acta.* **1986**, 69, 255.
- [7] A. L. Spek, Utrecht University, Netherlands, **2000**.
- [8] T. Gelbrich, University of Southampton, UK, **2002**.
- [9] S. W. Christensen, University of Southampton, UK., **2002**.
- [10] G. F. Voronoi, J. Reine, *Angew. Math.* **1908**, 134, 198.