Library Project – Part 2 (Questions 1 & 2)

Ben Mills

1) Literature search

Crystal structure prediciton
You should aim to cover:

- *the different approaches*
- *the range of structures which can be predicted successfully – in brief*
- *the challenges involved at each step*
- *the success rate for each method*
- *future perspectives*

Introduction

Crystal structure prediction (CSP) is of fundamental importance[1] and has been a goal of the physical sciences for more than fifty years[2], but as recently as 1994, it was considered impossible by Gavezzotti[3] and has since proved to be a very challenging task.[4,5,6] In the late 2007, a significant advance[7,8] was made by Neumann *et al* in *ab initio* CSP, that is to say, the prediction of molecular crystals, knowing only the connectivity of the molecule[9]. Neumann's new CSP method has already been commercialised[10], but perhaps even more significantly, it paves the way to a much deeper understanding of the crystallization process, which at present is poorly understood[11]. Despite this recent success, great challenges lie ahead before reliable and comprehensive computational methods of modelling crystallization from first principles become available as a tool to scientists, as the focus shifts from predicting the most thermodynamically stable crystal structures to those that are kinetically accessible under given experimental conditions. A full treatment of polymorphism requires extensive and vastly complex modelling of the kinetics of crystallization under real conditions, but the Neumann method is proving a helpful step towards this ambitious goal.[12]

The demand for a reliable method of predicting the crystal structures of crystalline materials is great – mainly from the pharmaceutical industry, since the different polymorphs of a drug are separately patentable and to guarantee consistent bioavailability and pharmacokinetics, production methods must be controlled so that only one polymorph is formed.[13] Other industries that benefit greatly from CSP include those that manufacture agrochemicals, pigments, dyes, explosives and fine chemicals, because such compounds derive many of their key properties from their crystal structures but are often not suitable for traditional diffraction methods of structure determination. Pigments, for example, are very difficult to obtain as single crystals, so few structural determinations are available[14]. The task of fully understanding and modelling crystallization may be daunting, but the economic and scientific impetus is there.

The different approaches

Historically, close-packing arguments and analogies with existing structure determinations (data-mining) have been used to predict the likely structure of a new crystalline solid[15]. For ionic crystals, simple criteria such as Pauling's rules have been developed, but these do not apply to molecular crystals and are not obeyed in ionic crystals when structural features such as lone pairs, hydrogen bonding and metal-metal bonding are present[16]. Crystal structures of organic molecules are very hard to predict and rank in stability because the factors that determine them, the intermolecular interactions, are weak and numerous, they act at long range and they are they have little directionality[2].

It is worth mentioning that there exists a spectrum of methods that are not true *ab initio* crystal structure prediction, but involve the computational modelling of crystal structures, including the calculation of the equilibrium structure of a crystal given an approximate structure, which can lead to useful results such the assignment of the correct space group. These methods are of value in the refinement of crystal structures obtained by diffraction, but are not enough to predict crystal structures from scratch[1].

The original and most popular approach follows Kitaigorodskii[17] and assumes intermolecular interactions are weak and poorly directional (isotropic). The crystal structure is thus evaluated on the basis of close packing, with the preferred structure having the most efficient packing. Modifications are required to improve this rather unrealistic model, in particular the effect of hydrogen bonding. The major drawbacks are (1) the fluctuations in intermolecular forces are too many to calculate properly and are often due to very subtle or distant structural features, and (2) finding the global energy minimum is not equivalent to finding the experimental crystal structures, as noted already. A full dynamic treatment of crystallization is the way forward, but such a thing is enormously demanding of computing power and is not yet within reach for realistic systems. An alternative way of guessing a crystal structure is to search crystallographic databases, such as the CSD, for similar molecules, but this does not help in cases where no similar structures have been determined. The advantage of such a data-mining route is that it may help decide if kinetic crystals are likely for the compound in question, for example, catemers rather than more thermodynamically stable dimers.

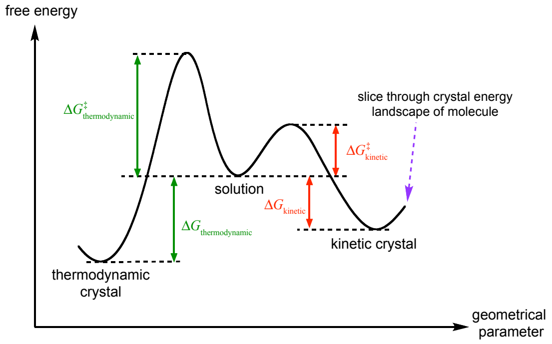

Figure 1: Schematic crystal free energy profile, illustrating the importance of kinetic factors in crystallization, polymorphism and crystal structure prediction. The formation of the kinetic crystal would be faster because there is a lower activation energy barrier to this pathway. The formation of the thermodynamic crystal would form more slowly but would be more stable. Comparing the mimumum energy crystal structure of a molecule obtained by CSP with the observed polymorphs can provide insight into the factors that affect crystallization[12]. Adapted from a diagram in Desiraju[2].

Neumann notes two problems for CSP to address. The physical problem of finding the relative stabilities of *all possible* crystal packings, and the 'mathematical' problem of finding all low-lying minima on the lattice energy landscape – a function of many variables, such as unit cell dimensions, space group, the conformation of molecules in the unit cell and their packing in the lattice – a very demanding problem.

CSP is simplest with small, rigid molecules that have consist only of common atom types (C,H,N,O), lack hydrogen bonding, have no more than one molecule in the asymmetric unit, are not solvates, salts or cocrystals, and crystallize is one of the more common space groups.[18] Unfortunately, these structural features are very common and they add many more degrees of freedom to the global energy optimisation problem at the heart of CSP.[7] It is fairly easy to generate low-energy crystal structures, but there is almost always a choice of many[6], all within a few kJ mol^{-1}. The problem is deciding which one will form under given conditions.[18] High temperatures give more weight to entropic factors, complicating matters. Far more difficult is understanding *how* crystals form[6].

To objectively assess the progress of research towards CSP, a series of four challenges known as the 'blind tests of crystal structure prediction' have been organised by the Cambridge Crystallographic Data Centre (CCDC) and held in 1999, 2001, 2004 and 2007. Participants from different institutions use a range of different programs to attempt to predict the experimental crystal structures of several small (< 30 atoms) organic molecules. The first test[5] revealed both the strengths and weaknesses of CSP at the time, as none of the 11 groups participating managed to consistently predict structures correctly but many of the predictions submitted were 'close' to the experimental structures . The second[19] and third[11] blind tests also gave mixed results. Overall success rates have been between 0% and 35%, but these do not reflect the consistency of any particular protocol. The best outcome so far was achieved in 2007, when three groups predictions' matched three experimental structures each, another three groups had two correct matches each, and the group consisting of Neumann, Leusen and Kendrick correctly predicted the crystal structures all four molecules. There was, however, an element of luck in the success of the Neumann group in that the experimental structures just happened to be the thermodynamically most stable structures.[7,12]

Neumann's procedure is novel because it sandwiches a quantum mechanical DFT step in between two faster classical molecular mechanics (MM) steps. The first MM step ranks the energies of all possible crystal structures, and submits the best to the DFT step, which calculates a customised force field of bond energies. The second MM step takes this force field and uses it to estimate the lattice energy and stability of each candidate structure. The method has broad scope because it treats molecular flexibility explicitly and considers all 230 crystallographic space groups. The method can only predict around 5% of structures useful to the pharmaceutical industry, so the group is now attempting to refine it to reliably perform CSP on more challenging targets: flexible molecules, salts, solvates, co-crystals and structures with more than one molecule in the asymmetric unit.[7] Therocedure has been commercialised by Avant-garde Materials Simulation of Freiburg, who will calculate all the polymorphs of a drug of a molecular weight of around 200-450, in a period of 4 to 22 weeks[10].

Since Neumann's breakthrough, the CSP community has been working on predicting the stability of different polymorphs under different crystallization conditions, such as temperature and pressure. Full dynamic modelling of crystallization process seems necessary to confidently predict which crystal structures are accessible. CSP can currently find energy minima but there are many of them, only some of which are actually found experimentally.[20]

Protein CSP researchers are pleased with the results of the 2007 blind test but cannot yet apply the method because proteins are so large (800-200,000 atoms) that current computing power is not enough to deal with the DFT step[12].

The prediction and rationalisation of the structures of inorganic compounds has a long history of prediction, including such criteria as Pauling's rules, but these are rarely selective enough to be used alone.[21] The crystal structures of inorganic compounds can be predicted computationally, although inorganic CSP has been considered somewhat less of a priority than for organic compounds because polymorphism is less common in inorganics[22]. Comparison with existing structures (data-mining) and quantum chemical methods (such as density functional theory) are the two main approaches. DFT is time consuming, so faster but less accurate methods are employed to create a short list of candidate structures for quantum chemical analysis. Inorganic CSP is valuable, for example, in unravelling the structures of zeolites, which are often not available from diffraction methods. CSP in general is useful to focus experimental effort on the most promising candidate materials, saving time, reducing costs and increasing the rate of innovation[23].

Future developments may include the greater use of evolutionary algorithms, simulations and metadynamics to address the questions surrounding phase transitions and crystallization kinetics. Crystal engineering applications and integration with crystallographic databases are likely to follow[24].

2) Keeping up to date

I have registered with *ISI Web of Knowledge* in order to use the *Citation Alerts* and *Saved Searches* features, and have set up *Citation Alerts* to email me whenever one of the key articles on CSP is cited in another paper. For example, all the Blind Tests, the 2007 breakthrough, and the most relevant review articles (Woodley & Catlow, Desiraju) are on my *Citation Alerts* list. So far, I have received one alert, on 30 January 2009, informing me of a newly-published article by Issa *et al* that cites Neumann (2008).

Issa *et al* (2009) notes the major breakthrough in 2007, citing both Neumann (2008) and an *Acta Cryst. B* article currently in press, which is presumably a more detailed report of the Fourth Blind Test (G.M. Day et al., "Important progress in predicting the crystal structures of small organic molecules – a report of the fourth blind test", *Acta Cryst. B*, in preparation). I could check the *Acta Cryst. B* website regularly to find the article when it is published, but this approach is labour-intensive. I have set up *Citation Alerts* with *Web of Science* for all the Blind Tests, and this as-yet-unpublished paper should cite them, so I expect to receive an email when it is published.

I have set up email alerts for IUCr journals (http://journals.iucr.org/iucr-top/user/ealert.html) to inform me of the contents of each new issue of *Acta Cryst. B* when it becomes available online (http://journals.iucr.org/b/journalhomepage.html), which is sometimes helpful in spotting new developments.

Some useful key words for searching databases for information on crystal structure prediction are "crystal structure prediction", "polymorph prediction" and "crystallization modelling". The more common journals for publishing articles on CSP are *CrystEngComm*, *Acta Crystallographica, Section B*, and *Nature Materials*, plus the usual general chemistry journals for important papers, such as the GDCh's *Angewandte Chemie International Edition*, the ACS's *Journal of the American Chemical Society*, and the RSC's *Chemical Communications*.

Some of the key researchers in the field of CSP are Angelo Gavezzotti at the University of Milan, Marcus Neumann at Avant-garde Materials Simulation in France, Sally Price at UCL, Frank Leusen at the Institute of Pharmaceutical Innovation at the University of Bradford, Gautam Desiraju at the University of Hyderabad in India, Graeme Day at Cambridge, and Sam Motherwell, although he has recently retired from the CCDC.

References

---

[1] S. M. Woodley, R. Catlow, “Crystal structure prediction from first principles”, *Nature Mater.*, 2008, **7**, 937–946.

[found with SciFinder Scholar 2007: after finding Neumann et al (2008) as above, selected Detail of Reference > Get Related… > Citing References]

[2] G. R. Desiraju, “Cryptic crystallography”, *Nature Mater.*, 2002, **1**, 77–79.

[found with Web of Science using the query “crystal structure” AND “prediction”, listing the results in order of number of times cited, the first chemistry-related article was the starting reference (Lommerse et al, 2000). Searched citing articles by clicking on “Times cited”]

[3] A. Gavezzotti, “Are Crystal Structures Predictable?”, *Acc. Chem. Res.*, 1994, **27**, 309–314.

[obtained from reference 5 (Lommerse 2000)]

[4] J. Maddox , “Crystals from first principles”, *Nature*, 1988, **335**, 201.

[obtained from Woodley & Catlow]

[5] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr., Sect. B*, 2000, **56** 697–714.

[found with Web of Science (http://www.isiknowledge.com), searching the “Author” field for “Lommerse J P M” and the “Year Published” field for “2000”]*

[6] J. D. Dunitz, “Are crystal structures predictable?”, *Chem. Commun.*, 2003, 545–548. *[found by same method as reference #2]*

[7] M. A. Neumann, F. J. J. Leusen, J. Kendrick, “A Major Advance in Crystal Structure Prediction”, Angew. Chem., Int. Ed., 2008, **47**, 2427–2430.

[found with SciFinder Scholar 2007: Explore > Research Topic > “crystal structure prediction” > references containing the search query as entered (as opposed to containing the concept)]

[8] G. M. Day, T. G. Cooper, A. J. Cruz Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf and B. Schweizer, , *Acta Crystallogr., Sect. B*, 2009, in preparation.

[found by searching Google for “Dr G M Day” and following links to his publications page on the Cambridge website, http://www-day.ch.cam.ac.uk/publications.html]

[9] R. J. Gdanitz, “*Ab initio* prediction of molecular crystal structures”, *Curr. Opin. Solid State Mater. Sci.*, 1998, **3**, 414–418.

[found with Web of Science search: “crystal structure” AND “predicition, and refining the results of the search to review articles only]

[10] M. Neumann, *Contract research*, Avant-garde Materials Simulation, Freiburg, Germany, 2009. http://www.avmatsim.de/services/contract-research.

[found by Google search for “fourth blind test”]

[11] G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, P. Verwer, “A third blind test of crystal structure prediction”. *Acta Crystallogr., Sect. B*, 2005, **61** 511–527.

[found by searching the IUCr journals website for “blind test”]

[12] S. L. Price, “Computational prediction of organic crystal structures and polymorphism”, *Int. Rev. Phys. Chem.*, 2008, **27**, 541–568.

[found by using Web of Science to set up a Citation Alert for Neumann et al (2008) – received email notification of this paper on 6 February 2009]

[13] S. L. Price, “The computational prediction of pharmaceutical crystal structures and polymorphism”, *Adv. Drug Delivery Rev.*, 2004, **56**, 301–319. *[obtained from reference 7 (Neumann 2008)]*

[14] N. Panina, F. J. J. Leusen, F. F. B. J. Janssen, P. Verwer, H. Meekes, E. Vlieg, G. Deroover, “Crystal structure prediction of organic pigments: quinacridone as an example“, *J. Appl. Cryst.*, 2007, **40**, 105–114.

[found by searching the IUCr journals website (http://journals.iucr.org/services/search.html) for “crystal structure prediction”]

[15] G. R. Desiraju, “Crystal Gazing: Structure Prediction and Polymorphism”, *Science*, 1997, **278**, 404–405.

[obtained from references in reference #2, Desiraju 2002]

---

[16] P. A. Madden, M. Wilson, "'Covalent' Effects in 'Ionic' Systems", *Chem. Soc. Reviews*, 1996, **25**, 339–350.

[found with a Google search for "Pauling's rules", which led to online lecture notes by Dr S. J. Heyes of the University of Oxford (http://www.chem.ox.ac.uk/icl/heyes/structure_of_solids/Lecture3/Lec3.html), containing the reference]

[17] A. I. Kitaigorodskii, *Organic Crystal Chemistry*, Izd. Akad. Nauk SSSR, Moscow, 1955.

[found in reference #2, located in the Worsley Chemical library by searching for "Kitaigorodskii" in the author field of the library catalogue: location H1 KIT]

[18] J. Bernstein, Crystal Structure Prediction and Polymorphism", *ACA Transactions*, 2004, **39**, 14–23

[found by same method as reference #9]

[19] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams, "Crystal structure prediction of small organic molecules: a second blind test", *Acta Crystallogr., Sect. B*, 2002, **58** 647–661. *[found by searching Google for "second blind test"]*

[20] *Major advances made in predicting crystal structures*, University of Cambridge, Cambridge, UK, 2007. http://www.admin.cam.ac.uk/news/dp/2007112604.

[obtained from reference #10]

[21] L. Pauling, "The Principles Determining the Structure of Complex Ionic Crystals", *J. Am. Chem. Soc.*, 1929, **51**, 1010–1026.

[obtained from the Wikipedia article on Pauling's rules (http://en.wikipedia.org/wiki/Pauling's_Rules), itself found by searching Google for "Pauling's rules"]

[22] T. Li, P. W. Ayers, S. Liu, M. J. Swadley, C. Aubrey-Medendorp, "Crystallization Force—A Density Functional Theory Concept for Revealing Intermolecular Interactions and Molecular Packing in Organic Crystals, *Chem. Eur. J.*, 2009, **15**, 361–371.

[found by same method as reference 10 (Price 2008)]

[23] A. Meden, "Inorganic Crystal Structure Prediction – a Dream Coming True?", *Acta Chim. Slov.*, 2006, **53**, 148–152.

[found by same method as reference #9, Gdanitz 1998]

[24] A. Nangia, "Database research in crystal engineering", *CrystEngComm*, 2002, **4**, 93–101

[found by same method as reference #9, Gdanitz 1998]