ChemComm
www.rsc.org/chemcomm

Are crystal structures predictable?

Jack D. Dunitz
Chemistry Department LOC, ETH-Hönggerberg, CH-8093 Zurich, Switzerland. E-mail: dunitz@org.chem.ethz.ch

The one-word answer to the title question is still "No", although at certain levels of discussion a "Maybe", or even a conditional "Yes", may be entertained as possible responses.

Are crystal structures predictable? Eight years ago, a terse monosyllabic negative reply to this question was given by Gavezzotti.[1] Has anything important happened since then to change the answer? Experts in the field have tried their skills in two blind tests and discussed their methods, together with their successes and failures, at two meetings organized by the Cambridge Crystallographic Data Centre in 1999 and 2001.[2,3] Although some successes were reported, no method gave consistently reliable predictions. How is it that, although physicists believe we are approaching a theory of everything, we still lack a theory of crystallization? Is the formation of a crystal more complicated and difficult to understand than the formation of the universe? Or do we mean different things when we say we understand crystallization or galaxy formation?

There is no problem about understanding *why* crystals form. The Gibbs equation $G = H - TS$ expresses the balance between the tendency towards maximal disorder of the molecules in a system and the opposing tendency for attractive intermolecular forces to bring the system into a state where its potential energy is minimized. As the temperature is lowered, the second tendency gains the upper hand. The attractive forces cause the system to condense into a liquid and on further cooling into the state of maximum order, which is crystalline. The entropy S of a perfectly ordered crystal is zero; the location of every molecule can in principle be described by an address, a set of numbers specifying its position in the three-dimensionally periodic array. In fact, real crystals are not perfectly ordered but this does not seriously affect the argument. Crystals form because the state of minimal potential energy is perfectly ordered.

Nowadays it is no great problem to calculate the potential energy of a crystal built from small to medium-sized organic molecules (with up to, say, 20–30 non-hydrogen atoms) by pairwise addition of atom-atom potentials, such as, *e.g.*, $E = \Sigma_i\Sigma_j A\ \exp(-BR_{ij}) - CR_{ij}^{-6}$ where R_{ij} is the distance between atoms i and j in different molecules. In the summation, i runs over the atoms of a reference molecule and j over atoms of all surrounding molecules. The exponential term represents the repulsion that sets in at short interatomic distance, and the inverse sixth power term represents the attraction in the form of a dispersion (van der Waals) interaction between the given atom pair. The parameters A, B, C need to be calibrated either against experimental data or theoretical calculations, and additional terms $q_iq_jR_{ij}^{-1}$ may be added to represent the electrostatic interaction between atoms with assigned charges at the nuclear positions. Usually the summation converges once the summation goes past the first coordination shell. For a stable crystal structure, E is negative, its first derivatives are zero, and its second derivatives are positive with respect to the various degrees of freedom.

Given a molecular structure, assumed for the moment to be rigid, why not simply calculate E for very many hypothetical crystalline arrangements and find the one with the lowest potential energy? Thousands (millions) of possible structures can be generated by grid search or random search techniques, or by systematic generation of dimers or strings of molecules over crystal space group symmetry elements, and these (or a selection of them) can be used as starting points for energy minimization calculations. The problem is that such calculations typically lead not to a single crystal structure that is much more stable than any of its competitors but rather to many possible crystal structures within a quite small energy range. This kind of result is compatible with the frequent occurrence of polymorphic forms of closely similar energy, but it is obtained also for compounds for which no

Jack Dunitz (b. 1923) studied chemistry at Glasgow University. After a decade of peripatetic research at Oxford University, Caltech, N.I.H. and the Royal Institution in London, in 1957 he was appointed Professor of Chemical Crystallography at the Swiss Federal Institute of Technology (E.T.H.) in Zurich, a position that he held until his retirement in 1990. As a survivor of the heroic age of X-ray crystallography before the advent of computers, automatic diffractometers and direct methods, he has witnessed how the subject has developed over the last half-century. He has written more than 350 scientific papers and is the author of *X-Ray Analysis and the Structure of Organic Molecules* (Cornell University Press, 1979; Verlag HCA, Basel, 1995) and *Reflections on Symmetry in Chemistry … and Elsewhere* (with E. Heilbronner, Verlag HCA, Basel, 1993).

DOI: 10.1039/b211531j

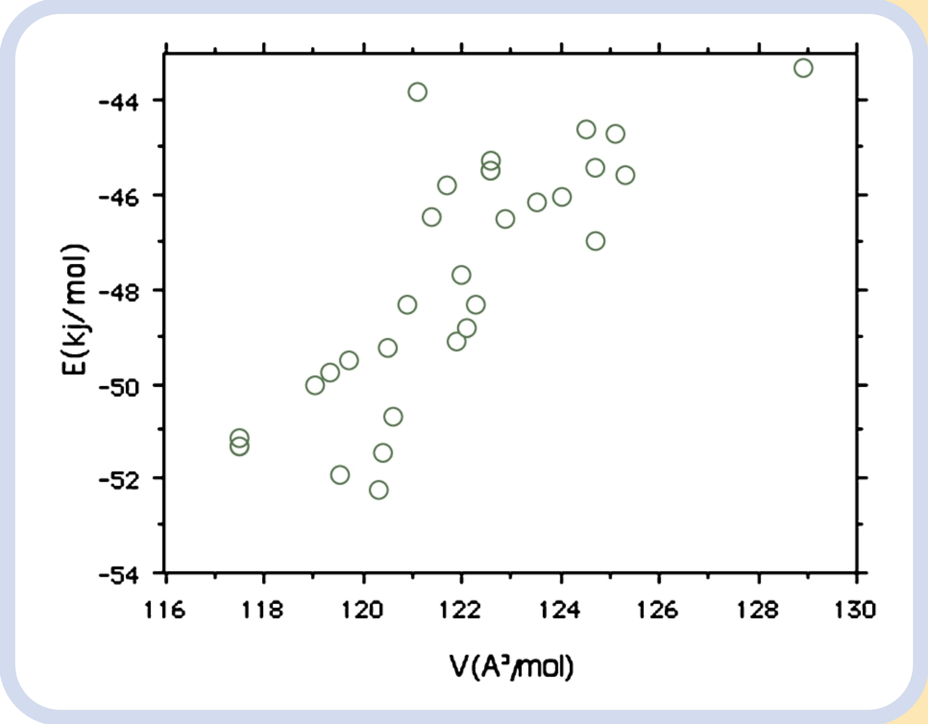


Fig. 1 Scatter plot of calculated lattice energy versus net molecular volume (unit cell volume divided by number of molecules in unit cell) for 30 possible crystal structures of benzene within an energy window of 10 kJ mol–1, based on results from ref. 6. Clearly low energy structures tend to be associated with tighter packing. The structure with the lowest lattice energy corresponds to the normal crystal structure of benzene, the one with the smallest volume to the structure at 25 kbar.

polymorphic forms are known.

For example, benzene. Under normal laboratory conditions benzene crystallizes in the orthorhombic space group *Pbca* and retains this structure down to very low temperatures.[4] At high pressure (circa 25 kbar) an alternative monoclinic structure is obtained.[5] According to a thorough recent study,[6] however, there are at least 30 possible crystal structures with calculated potential energies within a 10 kJ mol^{-1} range (Fig. 1). Indeed, the one with the lowest calculated energy corresponds to the known *Pbca* structure, but it is only a small fraction of a kJ mol^{-1} more stable that its nearest competitors. Also, the one with the highest calculated density corresponds to the known high-pressure structure.

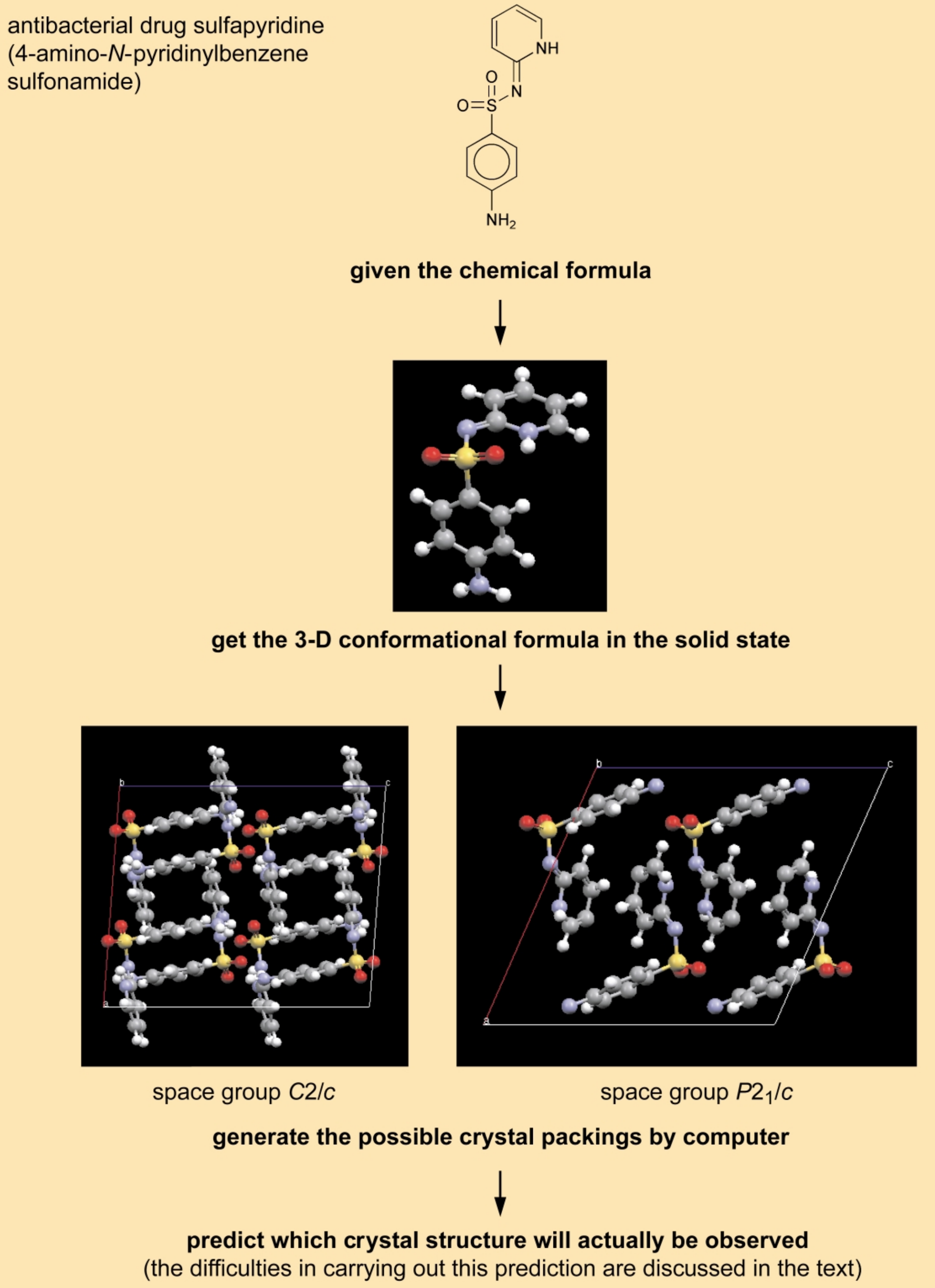
The problem is not so much a matter of generating stable crystal structures but rather one of selecting one or more from many almost equi-energetic possibilities (Fig. 2). Fine details of the calculation and of the parametrization can easily alter the energy ranking. Moreover, the potential energies estimated by such force-field calculations refer to collections of atoms at rest and take no account of the fact that in the real crystal the molecules are vibrating about their equilibrium positions and orientations. It is possible to allow for the vibrational energies and entropies of the calculated structures by making suitable approximations but this does not usually lead to more confident predictions. On the contrary, the less tightly packed the structure, the higher the potential energy tends to be, and also the larger the vibrational entropy.[7] Thus a slightly less favorable potential energy can be compensated at 300 K by a more favorable entropy, and the energy ranking becomes even more uncertain. In any case, any prediction based on the calculated energy ranking would implicitly assume that crystal formation depends on thermodynamic factors rather than kinetic ones.

The question of *how* crystals form is even more problematic, that is to say, very little is known for certain and it is much more difficult to make realistic computer calculations. Crystallization is not an equilibrium process. It requires supercooling or supersaturation. At low degree of supercooling or supersaturation, molecular clusters are presumably being formed and dispersed. As supercooling or supersaturation increases, one type of cluster may tend to grow faster than the others, and such clusters may develop into crystal nuclei. This nucleation process is probably the rate-limiting step in crystal formation, but one may not assume that the fastest forming nuclei necessarily correspond to those of the most stable crystal structure. Indeed, there are many cases known where the crystals first formed from solution or from a melt do not correspond to the thermodynamically stable crystal under normal laboratory conditions.

Thus, computational methods for predicting crystal structures of organic compounds cannot yet be regarded as reliable. From a more qualitative and descriptive viewpoint has come the notion that certain groupings in organic molecules exercise attractive intermolecular interactions and so guide the molecules into distinctive patterns in their crystal structures — structure directing 'synthons'.[8] This indeed has become one of the tenets of crystal engineering. The prime example of a structure directing interaction is, of course, the hydrogen bond, an interaction that has been widened from its original sense ($F–H\ldots F^-$, $O–H\ldots O$, $N–H\ldots O$, etc.) to include gradually weaker interactions such as $C–H\ldots O$, $C–H\ldots F$, $C–H\ldots\pi$, and so on into virtual nothingness. One can *describe* almost any observed crystal structure in terms of such interactions. The trouble with this approach is that its physical basis is shaky and its predictive capacity is poor. Of course, inspired guessing may occasionally lead to a correct prediction with any approach. Clearly, the preoccupation with those particular atoms or groupings on the peripheries of different molecules that come 'into contact' with one another is compatible with the r^{-6} dependence of the main attractive term in the usual atom-atom pair potentials referred to earlier (for any given interaction, a 12% increase in interatomic distance corresponds to a halving of the attraction energy).

Indeed, the atom-atom pair potentials described above place the main emphasis on dispersion energy as the dominating attraction term, with electrostatic energy appearing only as a correction term and then in a vastly oversimplified manner with assigned electric charges at the nuclear positions. In more elaborate models for individual molecules, electrostatic effects are expressed in terms of distributed multipoles to simulate not only atomic charges but also features such as lone pairs and π electrons.[9] However, the main emphasis is still provided by the repulsion–dispersion model, and crystal structure predictions using potentials based on the distributed multipole approach are only marginally more successful than those based on simple transferable atom–atom potentials.[2,3]

Very recently, Gavezzotti has described a new procedure for calculating intermolecular interaction energies by direct numerical integration over electron densities.[10] The method is being extended to the calculation of polarization, exchange



This figure is courtesy of the Cambridge Crystallographic Data Centre (CCDC) which serves the scientific community through the acquisition, evaluation, dissemination and use of the world's output of small molecule crystal structures, especially through the compilation of the Cambridge Structural Database. *For more information on the CCDC see http://www.ccdc.cam.ac.uk/*

Fig. 2 Crystal structure prediction.

repulsion, and dispersion terms.[11] In place of the many adjustable parameters of atom-atom pair potential schemes, the new method includes only three disposable numerical parameters and seems to be applicable to a wide range of molecular systems. Sublimation energies of many organic crystals are well reproduced. Although destabilization and repulsion at very short distances are not correctly reproduced, the new method seems to provide reliable intermolecular energies at distances characteristic of molecular crystals, where 'penetration energy' arising from overlap of molecular electron densities is not necessarily repulsive. In particular, at such distances, electrostatic energies calculated with this procedure bear little relation to those obtained with point-charge or distributed multipole models. They are generally much larger. It is too early to judge whether this new approach will be more successful than others in the challenge of crystal structure prediction. What it does seem to show is that interpretations of intermolecular attractions and repulsions based mainly or exclusively on r^{-6} interactions between atoms in close contact may be misleading. In fact, it suggests that although the widely used atom-atom approximation is moderately successful in calculating packing energies of crystals, it lacks any sound physical basis. Does this matter?

What next? Doubtless we can look forward to improved force fields tailor-made for individual molecules. Better methods of converting 0 K potential energies into free energies would tell us about the temperatures where free-energy curves cross and hence about the range of thermodynamic stability of hypothetical crystal structures. Faster computing should enable extensions of these methods to more complex molecules with additional degrees of conformational freedom and also to co-crystals, including hydrates and other solvates. Thus, given a molecular formula, we should be able to obtain a list of perhaps 10–20 crystal structures within an energy window of a few kJ mol^{-1}, among which all observable polymorphs are likely to be found. Since energy differences between polymorphs are indeed often very small, we could then claim that the crystal structure problem is more or less solved. However, the ideally periodic structures that serve as models for the energy calculations are far from the physical reality. The free energies of real crystals depend on strains associated with defects and imperfections, accidents of their growth. To go any further with our predictions we would need to ask exactly what we are supposed to predict. How are the crystals to be prepared? By sublimation? By cooling the melt? From solution? In what solvent? At what temperature? At what cooling rate? Besides, it is well known that crystallization experiments are sometimes not reproducible, as witnessed by the phenomenon of 'disappearing polymorphs'.[12] We have a long way to go before we can think of predicting which polymorph will be obtained under any given circumstances.

There are obvious similarities between the crystal-packing problem and the protein-folding problem. Both involve delicate balances between attractions and repulsions at the atomic level, between potential energy and entropic contributions to the free energy, and between thermodynamic and kinetic factors. There are even analogies between the use of 'tailor-made additives' to inhibit formation of a more stable polymorph[13] and Nature's use of chaperone proteins to prevent improper associations and thus ensure biologically correct behavior of their charges.[14] If anything, the protein-folding problem may be even more difficult since it involves interactions between different side chains rather than between identical molecules. Moreover, to allow the flexibility necessary for function, the packing density in a protein molecule must be somewhat less than in a crystal, so even more structures may be attainable within a narrow energy range. The biggest difference, perhaps, is that because of its bio-medical importance vastly more money and effort is being invested in the protein-folding problem.

References

1 A. Gavezzotti, *Acc. Chem. Res*, 1994, **27**, 309.
2 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 and D. E. Williams, *Acta Crystallogr., Sect. B*, 2000, **56**, 697.
3 W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, 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 and D. E. Williams, *Acta Crystallogr., Sect. B*, 2002, **58**, 647.
4 G. J. Jeffrey, J. R. Ruble, R. K. McMullan and J. A. Pople, *Proc. R. Soc. London, Ser. A*, 1987, **414**, 47.
5 G. J. Piermarini, A. D. Mighell, C. E. Weir and S. Block, *Science*, 1969, **165**, 1250.
6 B. P. van Eijck, A. L. Spek, W. T. M. Mooij and J. Kroon, *Acta Crystallogr., Sect. B*, 1998, **54**, 291.
7 J. D. Dunitz, G. Filippini and A. Gavezzotti, *Helv. Chim. Acta*, 2000, **83**, 2317.
8 G. R. Desiraju, *Angew. Chem. Int. Ed. Engl.*, 1995, **34**, 2311.
9 S. L. Price, *J. Chem. Soc., Faraday Trans*, 1996, **92**, 2997.
10 A. Gavezzotti, *J. Phys. Chem., Sect. B*, 2002, **106**, 4145.
11 A. Gavezzotti, personal communication.
12 J. D. Dunitz and J. Bernstein, *Acc. Chem. Res*, 1995, **28**, 193.
13 R. J. Davey, N. Blagden, G. D. Potts and R. Docherty, *J. Am. Chem. Soc*, 1997, **119**, 1767.
14 See, for example, D. Voet and J. G. Voet, *Biochemistry, 2nd Edition*, Wiley, New York, 1995, p. 201.