CrystEngComm

RSC Publishing

HIGHLIGHT

View Article Online
View Journal | View Issue

The "sceptical chymist": intermolecular doubts and paradoxes

Cite this: *CrystEngComm*, 2013, **15**, 4027

Angelo Gavezzotti*

Received 10th January 2013,
Accepted 21st March 2013

DOI: 10.1039/c3ce00051f

www.rsc.org/crystengcomm

Physico-chemical doubts and paradoxes abound in the description of forces and in the molecular simulation of the crystalline state, as they did in Boyle's times about the inner structure of matter. Solid-state structuring and bonding are still characterized in sometimes doubtful, sometimes paradoxical, and sometimes distorted terminologies. Phase transitions are mostly considered as just relationships between the two termini, without an operational understanding of the in-between transition mechanisms. Drawing from personal experience and recent computational results, this highlight provides a few answers, many caveats, and some suggestions for a better handling of these tortuous matters.

He also told them a parable: "Can a blind man lead a blind man? Will they not both fall into a pit? ... Or how can you say to your brother, 'Brother, let me take out the speck that is in your eye,' when you yourself do not see the log that is in your own eye?" (Luke 6: 39–42).

Introduction

In the opening gambit of one of his famous Pink Panther movies, Inspector Clouseau (Peter Sellers) harasses a poor accordion player in the streets of Paris for not having a license (which, in his irresistible *franglais*, he pronounces *lee-s-uns*). Meanwhile, behind his shoulders two brutes maim an old lady while fleeing from a bank robbery. Deservingly, Clouseau, disguised as a street artist or peddler while stalking the wrong suspect, gets twice arrested by his colleagues for not having a *lee-s-uns*. In another story almost as famous, two gentlemen in a balloon land in a cornfield out of thick fog and ask a peasant, where are we? The man answers, well, you're in a balloon. The two gentlemen correctly retort that the answer is completely rigorous, fully correct, and totally irrelevant. And even an atheist as the writer of the present paper is struck by Jesus' insight into human nature, testified by the Gospel excerpt that serves here as epigraph.

Sound science thrives on skepticism, that is, running after the most outstanding culprits first, taking care of logs long before specks, and pursuing relevance before, and even sometimes at the expenses of, rigorousness. Otherwise, even careful operators may end up producing papers that "are not even wrong" (a stingy epithet whose invention has been attributed in turn to several great scientists). Or less cautious writers may indulge in the dangerous kind of sleight of hand that makes easily lifted specks more important than immovable logs, launching wrong ideas that sail unchallenged in the open seas of present-day publishing policies.

The study of intermolecular interaction enjoys rich contributions from experiment, nowadays performed by apparatus and techniques that were unheard of ten years ago and expand and improve at an impressive pace. X-ray structure analysis for gaseous substances by *in situ* crystallization is now almost routine, and one can measure in attojoules per meter the force holding a single molecule on a templating surface. On the computational side, the frontier of petaflops has been crossed, and the barrier of million-atom simulations has been broken. Theory, or more humbly, the ability to separate scientific wheat from chaff, seems to lag behind. This Highlight leads

Department of Chemistry, University of Milano, Via Venezian 21, Milano 20133, Italy. E-mail: angelo.gavezzotti@unimi.it

Angelo Gavezzotti

Angelo Gavezzotti's graduation thesis in Chemistry consisted of the determination of one crystal structure, in 1968. He has been working on crystal structure determination and analysis ever since. His interests in the physics of intermolecular interactions led him to stray into theoretical chemistry and to write computer programs for a variety of crystal-related tasks. He is presently a retired Professor of Physical Chemistry and he is still active in research as a guest of the Chemistry Department in Milano. He is also associated with PolyCrystalLine in Bologna (www.polycrystalline.it).

the reader in a trip across the field using worked examples and specific case studies: it is not a review, and the material is almost entirely taken from unpublished data. Accordingly, there is no literature survey and citations are restricted to the essentials.

Typical questions and a key number, 1389.34

Chemical bonding is a matter of favorable arrangement of nuclei and electrons in a complex building. A chemist by training tends to isolate atoms, or better their nuclei, identified by a graphic device like the symbol of the element or a colored ball, and tends to draw single, double, triple lines, dots, or even curved arrows between these *ersatz* atoms. For some purposes this is an essential exercise in chemical understanding; for intermolecular interactions, this is often a subjective affair between a researcher and his or her favorite hobbies. The Atoms in Molecules (AIM) theory[1] is completely rigorous, fully correct, but hardly useful in interpreting organic crystal constitution, producing as it does long lists of very small numbers associated with hardly significant bond paths. Some of the work in this direction is at risk of severely downgrading the excellent reputation of AIM in the description of intramolecular bonding.

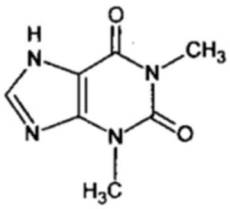
In a typical situation, one has good single-crystal diffraction data for a compound of crucial solid-state properties. Typical questions are: how do I understand how the crystal forms? What are the crystal structure determinants? These questions are important if one wants to proceed from analysis to prediction and control of crystal formation and constitution, for which theoretical leverage is indispensable. For sure, structure determinants are many, especially in multiform organic crystals. The matter is priority, and quantitative ranking is all there is to it. In many excellent crystallographic papers the analysis of crystal packing has lots of specks and forgets the logs, with discussion of weak interactions sorted out by visual inspection of that most misleading by-product of X-ray diffraction experiment – packing diagrams. These are indeed weak potentials, worth less than a handful of kJ mol^{-1}, while electrostatic interactions two orders of magnitude larger are peacefully neglected. The number appearing in the title of this section is the conversion factor from electrons squared over Ångstrom to kJ mol^{-1} and tells that two unit charges 10 Å apart still suffer from attraction or repulsion from a potential energy of 140 kJ mol^{-1}! There exist nowadays cheap, accessible, easy-to-use methods for the quantitative evaluation of crystal potentials and forces. The following paragraphs give a (non-exhaustive) variety of such applications.

Theophylline (TH) solids are known in anhydrous and hydrate form, in a host of one-to-one complexes with urea, DMSO, salicylic, malonic, maleic and *p*-hydroxybenzoic acids, acetaminophen, saccharin, sulfathiazole, *p*-nitroaniline, and even in *S*-substituted form (*S*-theophylline). Table 1 has some numbers that can be obtained by freely available software, in a few hours of computing time and in a couple of days of man-time of a moderately trained operator. These numbers show some valuable information: complexation brings about a large disruption of TH–TH interaction in favor of strong host–guest interaction, not surprisingly since all guests form strong hydrogen bonds to the TH host, as semiquantitatively confirmed by the overwhelming predominance of Coulombic interaction in crystals of the complexes. The *p*-hydroxybenzoic acid co-crystal almost completely segregates the host–guest manifold out of spectator host–host and guest–guest relationships.

Table 1 Lattice energies of theophylline and its complexes[a]

	Host–host		Host–guest		Guest–guest	
	Coul	Disp	Coul	Disp	Coul	Disp
TH anhydrous	−95	−105	—	—	—	—
TH–urea	−24	−60	−162	−85	−40	−11
TH–DMSO	−13	−16	−101	−89	−30	−75
TH–OHBz acid	−23	−38	−219	−151	−7	−18
TH–malonic acid	−51	−19	−143	−98	−58	−56

[a] From PIXEL[2] calculations (kJ mol^{-1} units); lattice sums partitioned over the indicated fragments. Polarization/repulsion contributions (not shown) add up to total lattice energies of about 150 kJ mol^{-1}. See ref. 12 for formulae and literature citations.



theophylline

Lattice energies are univocally defined and univocally correspond to a thermodynamic concept, the energy released when one mole of matter goes from the gaseous to the crystalline state (or the energy absorbed in the inverse process). Lattice energies, however, condense a wide spectrum of distributed potentials into just one number. Chemists want to localize energies into easily recognizable ''contact points''. But how to localize? In an intramolecular reactive process there are obvious transfers of electrons from one bond center to another, but in the formation of a non-ionic solid (ionic solids are considered in a separate section of this account) the ''intermolecular'' electrons are evanescent, and localizing on atoms and into bonds between atoms is often impossible. Much less criticizable is the localization on molecules, on whose identity there is less disagreement. In analyzing the crystal structure of anhydrous theophylline an obvious priority goes to hydrogen bonding, but a simple calculation of molecule–molecule interaction energies shows that this priority is seriously challenged by aromatic stacking (Fig. 1) in a ratio of just 2 : 1.

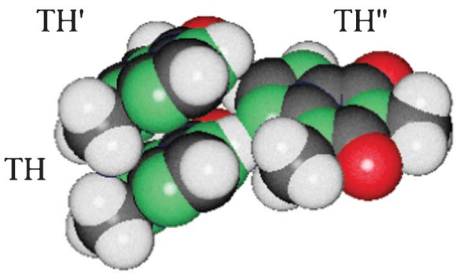


Fig. 1 Molecular pairs in the crystal structure of theophylline: TH–TH′, stacking, $E = -24$ kJ mol^{-1}; TH–TH′′, hydrogen bonding, $E = -49$ kJ mol^{-1}. N, green, O, red, H, white, C, black. All molecular drawings are by Schakal.[3]

Fig. 2 shows the detail of the massive structural rearrangement that accompanies the co-crystal formation, already guessed from the overall data in Table 1. Urea is so H-bond avid that it keeps hydrogen bonds to itself and prevents the formation of H-bonds between guest theophylline, which however preserves its stacking pattern. Curiously then the weaker stacking potential of Fig. 1 survives better than the stronger hydrogen bond. The host–guest hydrogen bond is 30% stronger than the guest–guest hydrogen bond, and one is tempted to speculate that urea–theophylline pairs must exist already in solution. *p*-Hydroxybenzoic acid occupies both the double acceptor–donor functionality and the single N-acceptor functionality of theophylline (binding energy 70 and 36 kJ mol^{-1}, respectively) preventing any host–host liaison (incidentally, note that the carboxylic oxygen between the two methyl groups is unavailable for H-bonding for steric reasons). One is tempted to generalize: host–guest interaction must win out if a complex is to be formed. Not always so: in the TH–malonic acid complex (Fig. 3) the formation of one very strong host–guest hydrogen bond turns the host into adopting the cyclic double N–H···O hydrogen bond pattern never appeared in the former examples.

Crystals and co-crystals so far considered were formed out of non-charged units, with lattice energies of the order of 150 kJ mol^{-1}. When organic ions are present the numbers change by doubling or even by an order of magnitude. The total lattice energy of ethanolammonium theophyllinate is 393 kJ mol^{-1},

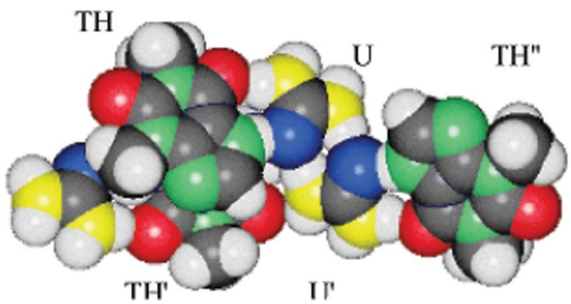


Fig. 2 Molecular pairs in the crystal structure of urea–theophylline: TH–U double N–H···O, −78 kJ mol^{-1}; U–U′ double N–H···O, −60 kJ mol^{-1}; TH···TH′ stack, −30 kJ mol^{-1}. The urea molecule is coded with blue O and yellow N.

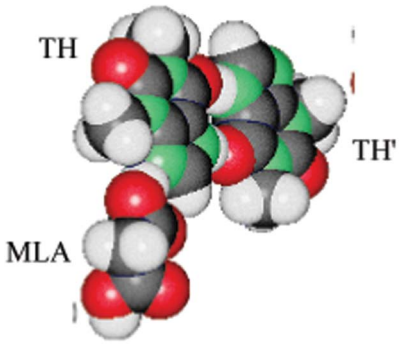


Fig. 3 Molecular pairs in the crystal structure of theophylline–malonic acid: TH–TH′ double N–H···O, −79 kJ mol^{-1}; TH–MLA single O–H···N, −50 kJ mol^{-1}.

and the molecule–molecule energy plot is shown in Fig. 4. Coulombic interactions at 30 Å are still as energetic as a strong hydrogen bond at 1.6 Å. Anion–anion and cation–cation interactions are strongly destabilizing and repulsive, and down to a center of mass separation of 10 Å, deviations from the energy of the bare Coulomb's law are small, *i.e.* molecules behave like point charges. Below 10 Å structuring and specific interaction are evident, and at the limit of shortest distances, where a hydrogen bond is formed, there is very large stabilization with respect to point-charge behaviour. Anion–cation interactions are always attractive and stabilizing and more than compensate the double influence of repulsion from like charges; the total anion–anion, cation–cation, and anion–cation Coulombic energies are 3800, 3760 and −8335 kJ mol^{-1}, respectively. The very popular Mercury[4] software searches for atom–atom distances below the sum of atomic radii, and a couple of (CH_2)···O=C distances appear. Short distances of that kind pertain to interaction energies of a handful of kJ mol^{-1}, and are nevertheless often pointed out as structurally relevant, although, in view of the numbers shown above, they are not even specks against logs, but barren leaves swept along by the tornado represented by the 1389.34 factor. Note that the responsibility for quoting weak features as

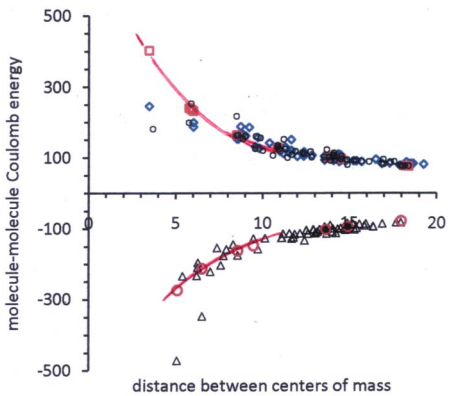


Fig. 4 A plot of PIXEL[2] Coulombic energy between pairs of molecules in the crystal of the ethanolammonium theophyllinate organic salt. The solid lines mark the bare Coulomb-law energies for unit point charges.

Table 2 Binding energy (kJ mol^{-1}) and anharmonicity index for the potential energy curves of some typical organic molecular pairings

Chemical description	E_0	Anh. index
Pair of organic ions	200–400	—
Pair of stacked cholesterols	40–60	—
Acetic acid single OH···O	32	1.0
Benzamide single NH···O	27	3.8
Acetic acid–acetamide OH···O (plus NH···O)	63	1.0
Pyrazole single linear N–H···N	39	1.6
Phenol OH···O	25	3.3
But-1-en-3-one *cis* dimer, double CH···O	9	8.0
Benzene offset stacked π···π	6–10	Very large
Benzene–hexafluorobenzene offset stacked	17	5.3
Benzene T-shaped C–H···π	11	10

significant rests on authors using the Mercury software, not on the authors of the software.

Examples of the negative spinoff of preconceived ideas on the relative importance of cohesive factors abound, and there is severe danger of contamination of otherwise important matters. Picking more or less at random, one may cite a display of extreme ingenuity and dedication to assay the differences in binding energies between hydrogen- and halogen-containing moieties in DNA strands.[5] These rather minor differences (3–20 kJ mol^{-1}) are then attributed to the fashionable ''halogen bond'', whereas the difference between the steric and electronic demands of hydrogen and a bromine atoms might require a somewhat more complex analysis. In a breakthrough paper[6] reporting the neutron diffraction determination of hydrogen positions in the interaction of acetazolamide and carbonic anhydrase, there is a discussion of ''very weak hydrophobic interactions'', and of ''somewhat distorted C–H···π interaction'' in a complex between a molecule carrying a NH^- anionic form, a sulfone group, and a zinc cation, where Coulombic energies of hundreds of kJ mol^{-1} are certainly at stake.

Physics or geometry: which comes first?

In science, God (if anything) would be a physicist, not a topographical surveyor. X-ray crystallography buries the physics into the data collection and data processing software, and produces geometry. Many excellent crystallographers seem to think that one could dip a polystyrene model of the space group into a solution, and molecules would oblige by crystallizing around its symmetry elements. In a similar fashion, the standard discussion of molecular interaction after X-ray data implies that atoms attract each other because they are neighbors, and a dotted line is promptly drawn between pair of nuclei that are closer than ''the sum of van der Waals radii'' – a mythic expression that is usually followed by citation of some prehistoric data. *Vice versa*, atoms (better, molecules) become neighbors when they attract, and one must explain neighborhood by the physics of attraction, rather than infer dubious physics from proximity. We may only offer here (Table 2) a semiquantitative list of relative cohesion energies and anharmonicity indices,[7] which may help avoiding some misunderstandings in crystal structure analysis. Perhaps even more important than binding energies are anharmonicity indices. They are a measure of the resistance of bonding to disruption: a value of 1 indicates almost harmonic behaviour, a very high value indicates that the dissociation side of the potential energy curve is flat, so that binding is labile and may not resist ordinary thermal strain.

And a *caveat*: in many cases, with the exception of hydrogen bonding, molecular pairings responsible for the largest part of the interaction energy in a crystal show no particular atom–atom feature, no easily identifiable ''bond'', not even aromatic stacks, or the like; they stick together by compatibility of minor and diffuse features in the electrostatic potential, that defy recognition and, *a fortiori*, classification. Only a quantitative calculation of cohesion energies can reveal true crystal structure determinants. A typical example is provided by the crystal structures of a series of cholesterol derivatives. The top contribution to crystal packing is in most cases by pairing of the cholesterol cores at 6 Å translation, irrespective of the nature of substituents. Table 3 shows the energies; the dispersion energy in the top pair is a large part of the total lattice energy, and quite often the dispersion contribution to pairing is larger than the total pairing energy, other electronic factors being actually antagonists in the molecule–molecule cohesion. Fig. 5a shows the typical 6 Å translation pairing mode; Fig. 5b shows instead a main determinant consisting of

Table 3 Largest molecule–molecule pairing energies and total lattice energies in crystals of cholesterol derivatives

CSD refcode[a]	Type of pairing symmetry	Molecule–molecule total[b]	Molecule–molecule dispersion[b]	Total lattice[b]
AXSCHO	Translation	−51	−48	−195
BZCHOL03	Translation	−53	−61	−195
CHOLOL	Translation	−56	−62	−212
COMYEI	Screw	−42	−45	−175
HINGAM	Screw	−49	−55	−168
NAKMIV	Screw	−58	−62	−178
PUXHAR	Translation	−36	−32	−174
QULWOJ	Translation	−53	−40	−227
SEGPOJ	Screw	−55	−67	−228
SOHVOA	Screw	−40	−47	−207

[a] Formulae and literature citations in ref. 12. [b] From AA-CLP[8] calculations (kJ mol^{-1} units).

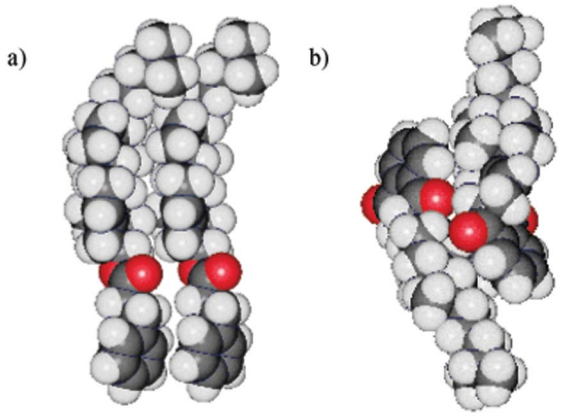


Fig. 5 (a) The most stable molecular pair in cholesterol crystals: lineup of cholesterol cores (CHOLOL). (b) Lineup of cholesterol core to an aromatic core (NAKMIV). In both cases the bulk of the interaction energy is unspecific dispersion (refer to Table 3).

a match of a cholesterol core with an aromatic core. A neologism for these recognition patterns awaits to be proposed. Possible C–H···O, C–H···π or other ''weak'' contacts are, as suggested in the opening sentences, specks in presence of logs.

Periodicity, order and disorder

Loosely speaking, a crystal is said to be ''ordered'' matter. Order is however a subjective concept; a better definition reads: at ordinary laboratory conditions, a crystal is a solid mostly but not always exhibiting time-averaged, periodic translational symmetry. This definition copes with the fact that molecules are librating, that lateral chains may be wagging or even from time to time changing conformation, and so on. No crystal modeling so far shown in this Highlight did include thermal motion; rich as these descriptions may be in energetic detail, they lack a most important physicochemical factor and are therefore largely incomplete. Fig. 6 shows two extreme views, the static model obtained by repeating the asymmetric unit according to crystal symmetry, and a picture of what one would actually see if he or she were able to take an instant snapshot inside a crystal. The latter picture was obtained by computational ''thermalization'', *i.e.* running a few hundred thousand Monte Carlo (MC) steps starting from the static crystal structure. This is no place to give a detailed account of the method;[8] suffice it to say that Monte Carlo is an astute machinery that systematically changes molecular coordinates according to a tolerance level of increase in energy, until a Boltzmann equilibrium state is achieved. Each MC snapshot is then an instant view on the phase space spanned by the system. In this sense, Fig. 6a is a physicochemical fiction, while Fig. 6b clarifies what is meant by ''time averaged periodicity'' and is much more realistic than ordinary packing diagrams.

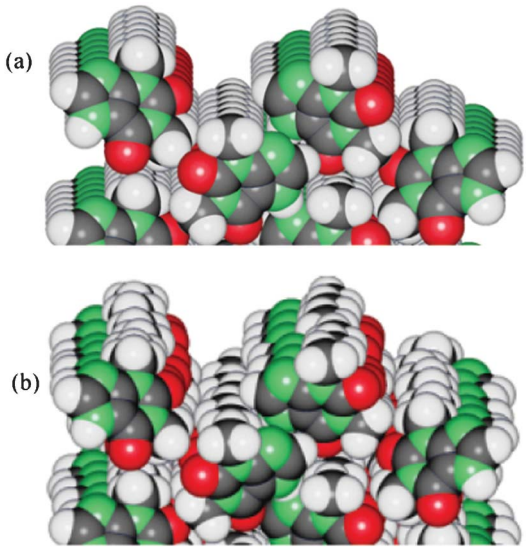


Fig. 6 (a) A ''static'' view of a surface slab of the theophylline crystal. (b) A Monte Carlo snapshot of the same at 400 K. Notice the misalignments.

Thermal motion is just a prototypical form of what may be called ''disorder'' in a crystal. But in crystal structures that include some solvent molecules, true loss of long-range periodicity is commonplace, because the host molecule often forms cages or channels that are loosely filled by the guests. For obvious reasons, a typical case in pharmaceutical practice is the occurrence, wanted or unwanted, of hydrates. The X-ray time-averaged picture of the theophylline monohydrate crystal is in terms of a water molecule distributed over two half-positions. A quick Monte Carlo modeling reveals what is really going on. If one builds a static model of the hydrate crystal using the full symmetry, the two half-water molecules overlap and clash into one another (Fig. 7a). Monte Carlo treatment patiently shifts the water molecules around in the channel to Boltzmann equilibrium, revealing the incommensurate nature of the host–guest complex, in which water molecules are more

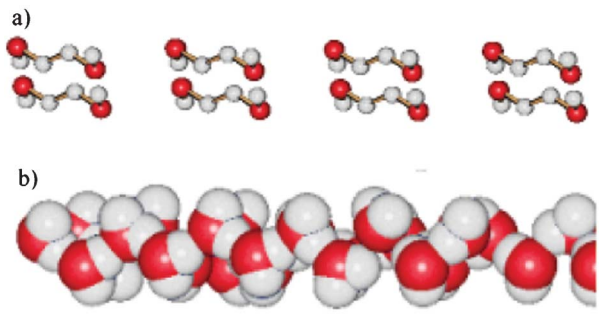


Fig. 7 (a) Beads of water molecules in a static model of the theophylline hydrate crystal: using X-ray coordinates and symmetry[12] water hydrogens overlap. (b) The same after 2 million Monte Carlo steps: water molecules arrange themselves in the channel in an unsymmetrical fashion, but are regularly hydrogen-bonded.

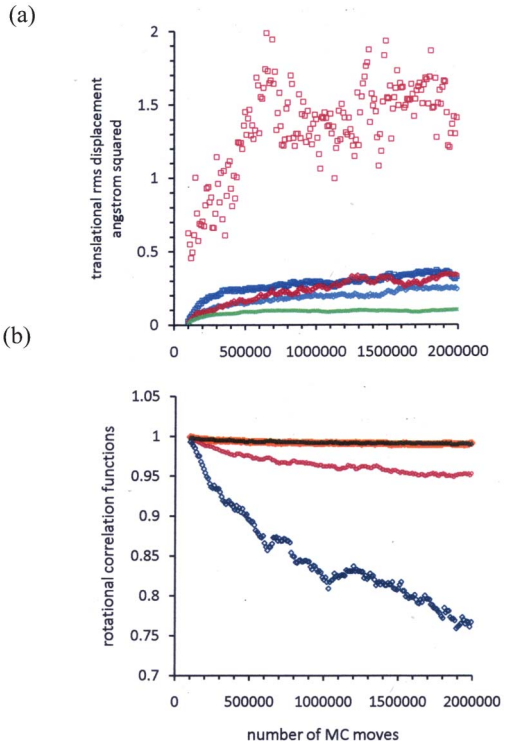


Fig. 8 (a) Translational rms displacement (red squares) of the guest water molecules in the simulation of the theophylline hydrate crystal. Zero is the static model of Fig. 7a. The lower lines are the rms displacement of host molecules in the hydrate and in other solvates. (b) The rotational correlation functions over the same simulation. Blue is guest water, other lines are host molecules, showing zero or minor loss of correlation.

or less randomly hydrogen-bonded, as shown in the snapshot of Fig. 7b.

Much more on the dynamic nature of the structure is revealed by inspection of correlation functions.[8] Briefly, these functions describe the translational and rotational ''memory'' of an intermolecular ensemble: a completely static crystal will have a zero rmsd displacement of the centers of mass, and a 100% retention of reciprocal orientation (rotational correlation = 1). A liquid has a rmsd displacement according to its viscosity, and the rotational correlation function drops quickly to zero from any reference frame. Fig. 8 shows the case for theophylline hydrate crystal. Water molecules soon leave the unrealistic placement in the static model and after a displacement of some 1.5 Å, bounce back and forth in the channel with a displacement amplitude range of about 1 Å. Rotational correlation (Fig. 8b) decreases steadily and is quickly lost, for an indication that water molecules still flip around and hydrogen bonding is fluxional. So the X-ray structural model which is able to bring the conventional *R*-factor down to 4.5% is an excellent starting point, but a dim representation of the reality. A cheap simulation offers an invaluable amount of additional information.

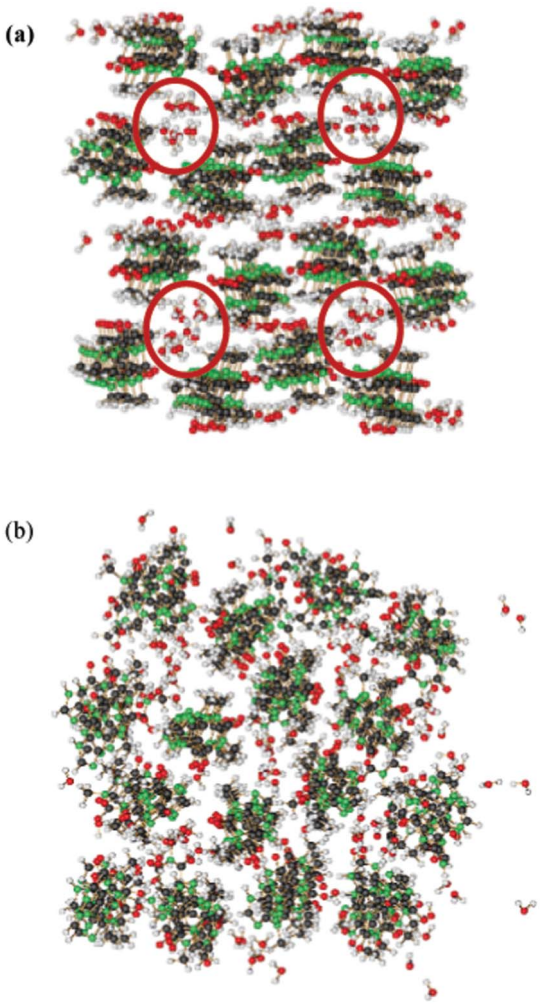


Fig. 9 (a) A cubic chunk of theophylline hydrate crystal in a simulation at 293 K: water molecules stay in channels (red ovals) and stick to the crystalline phase. (b) Simulation at 450 K: water molecules leave the system and channels begin to collapse.

Phase transitions: the ultimate source of doubt

Chemists and crystallographers who see neat and clean crystal specimens which in turn afford clean and neat X-ray diffraction pictures of the interior tend to forget the complex and laborious processes that preside over molecular aggregation in solution, over nucleation and growth. Conversely, little is known on the dynamics of loss of periodicity, disaggregation and possible release of inclusion compounds and, ultimately, melting. Once again, the very appealing picture of molecular packing offered by single-crystal (and, nowadays, also powder) diffraction experiments tends to give an oversimplified impression of reality.

The availability of a computational temperature and pressure in molecular simulation allows an excursion into the land of crystal evolution. Crystal melting can be easily simulated. A computer quickly finds one of the billion paths

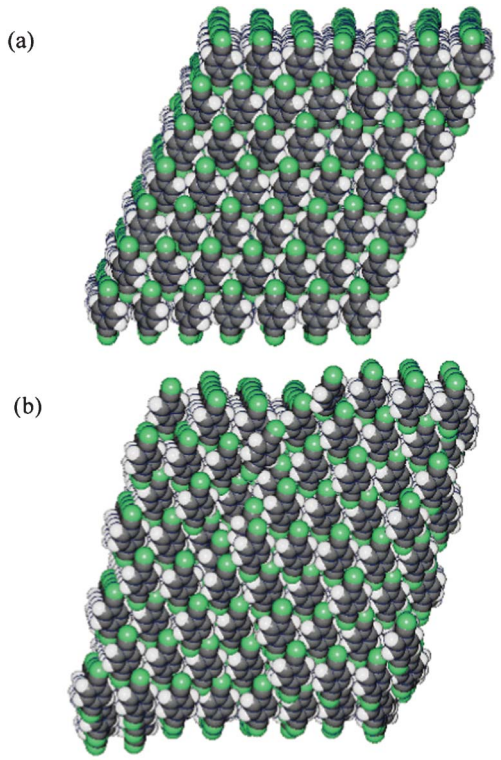


Fig. 10 (a) A slab of the experimental 1,4-dicyanobenzene crystal after heating to 450 K in a Monte Carlo run (tiny molecular misalignments differentiate from the static picture). (b) The final result of Monte Carlo-biased simulation of the transition from the melt at the same temperature, followed by energy minimization.

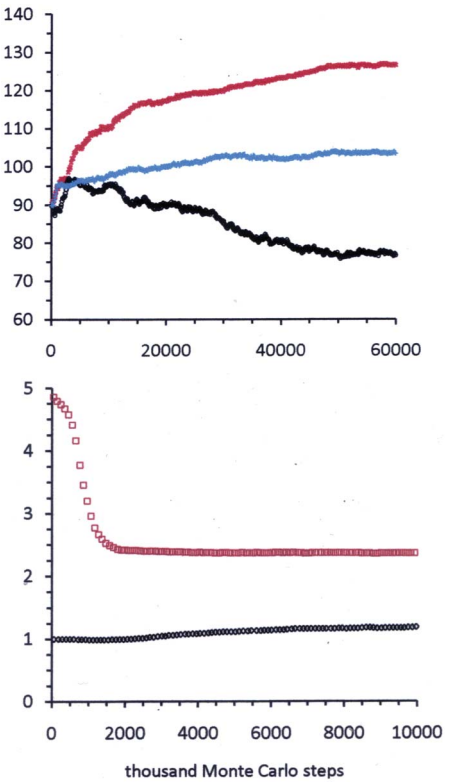


Fig. 11 A polygraph showing the structural trajectory of the TEPN crystallization simulation. Top: very fast anisotropic evolution of the three box angles from cubic. Bottom: red: evolution of the asymmetry index; black: evolution of density. A small dip in the 0–2 million steps region is barely visible.

through phase space that lead from crystal to liquid. As the computational temperature is raised, density decreases, since potential functions are anisotropic as they should be. At some point the crystal collapses to an isotropic liquid by chaotic molecular rearrangement.[9]

For a simple example of a different evolutionary process, Fig. 9 shows the result of a molecular simulation of a small chunk of theophylline hydrate crystal, without periodic boundary conditions, where it is seen that at 450 K water molecules are leaving out of the channels, while there is major restructuring inside the crystal with partial collapse of the channels. Notably, the experimental dehydration temperature of theophylline hydrate is 350 K, or some 100 degrees lower. The simulation needs some extra computational-thermodynamic "drive" to make things happen, but the indication that water loss will occur only at high temperature (and before melting) is clear, although deriving the exact number in a routine simulation is out of question.

The present level of understanding of the process of crystal formation from solution is very low, not to say zero. A much simpler process, barely amenable to molecular simulation, is crystallization from the melt. Monte Carlo is a computational machine that can be easily biased by the operator; in this case, one can introduce a computational bias that tells the machine to accept changes only if a properly designed asymmetry parameter decreases, or increases within a selected threshold ("asymmetry tolerance"). This parameter can be easily written for the case of $P\bar{1}$ crystals and rigid molecules:[10] it is just a number that counts the sum of distances between corresponding atoms, when molecules are computationally overlapped by bringing them to a common origin in the center of mass. This asymmetry parameter is zero for a perfect crystal model and is a very large number in a liquid. Without this bias, the simulation would take the age of the universe before finding the unique or one of the very few paths that lead from liquid to crystal.

The Monte Carlo-biased simulation starts from an isotropic liquid phase at or just below the crystallization temperature, and runs until a stationary state hopefully similar to the experimental structure is reached. In later stages, energy optimization cycles may be necessary, which are carried out as Monte Carlo runs at formally zero temperature. In the case of 1,4-dicyanobenzene (terephthalonitrile, TEPN) the procedure is completely successful (Fig. 10) and the experimental crystal structure is reproduced almost perfectly; had it not been known, one could have claimed here a really *ex novo* dynamic crystal structure prediction – except that a parallel orientation of rigid molecules obtains only one space group, namely, $P1$, which is therefore predetermined.

Monte Carlo offers quick access to configurational space, but obviously such physics-based, random walk techniques are far from granting efficient structure prediction. More interesting is however the analysis of the crystallization trajectories: in fact, Fig. 11 shows a very fast evolution from isotropic (all angles 90°) to anisotropic box dimensions, together with a very fast decrease of the asymmetry index and, contrary to expectation, a slight drop in density, barely visible on the scale of the figure. The total energy in this stage rises by some 5 kJ mol^{-1}. Taken together, these features outline an early crystallization stage in which there is a very fast reshuffling of molecular orientation, concurring to a line-up proper of the $P\bar{1}$, $Z = 1$ crystal structure. Isotropy is quickly lost as the parallelization of the TEPN molecules requires slanted ribbons and layers. This first stage might lead to something similar to a liquid-crystal state, as shown by intermediate snapshots along the crystallization path. Evolution toward full long-range periodicity requires fine adjustment and is then much, much slower.

The ''velocity'' of the process can be regulated by the value of the asymmetry tolerance: when this is very low, the bias pulls the system very fast through phase space, there is little simulation time for relaxation, density drops and energy rises as molecules are forced to climb energy hills to find the new alignment. When the tolerance is increased, the drive to symmetrization is less strict and molecules have time to relax. So to some extent – as is proper of all molecular simulations – the results depend on the choice of the simulation parameters.

Note that the introduction of the term ''velocity'' implies that the Monte Carlo procedure has been contaminated and has been forced into temporary deviation from Boltzmann conditions. In any case, the suggestion from this computational experiment, along with many others conducted on different molecules, is that crystallization requires a volume and energy activation; when molecular shape is more complex than just cylinder-like, molecules behave like passengers on a crowded subway car trying to reach the alignment that allows increased density. In the process some compression arises and energy rises. The Monte Carlo-bias simulation is an example of non-rigorous procedure that brings in significant new facts or, at the very least, some valuable suggestions. The fact that the real crystal structure is eventually reproduced rather well gives substantial confidence in the performance of the force field.

There is a long way to routine application of this procedure, if only because writing asymmetry parameters for space groups with symmetry operators other than translation or inversion is extremely difficult. How does one write in numbers the deviation from a precise screw-axis relationship between two complex organic molecules? The method thus now works only for the two triclinic space groups, leaving out all more common space groups for organic compounds.

The above approach to crystal generation may be compared to crystal structure prediction methods which consist of drawing a molecular structure and generating thousands of crystal packings by anonymous computer power. The lattice energies are evaluated and the most stable structure is the predicted structure.[11] This brand of CSP is slowly becoming more and more successful, having nevertheless more a flavor of statistical expert system than the look of a sound and progressive chemical tool. While we had to adopt a mathematical bias to overcome the molecules' reluctance to crystallize, the above described molecule-to-crystal CSP certainly brings to the fore the inevitable fact, that many almost equi-energetic structures exist for a given molecule and that many routes to low-energy crystal structures must exist.

Final remarks

The main aim of this Highlight is the description of a number of computational techniques that are nowadays becoming more and more available to chemists for whom the structural point of view is indispensable. It now quite easy to have quantitative estimates of separate factors that concur into the formation of a crystal structure, or for that matter, of any other molecular aggregate. The proper way to go to about the very complicated business of analyzing a crystal structure is to recognize stronger influences first, and then, if necessary, proceed to investigate subtler factors. All packing features, not only fashionable coupling modes, should be considered for this purpose; crystal structures have many and many more facets than the few patterns one may try to restrict them to. The study of weak interactions is a well cultivated field for theoretical specialists, but the fact that they are very much studied does not *per se* imply that they are dominant. The fact that they can be easily but roughly pulled out by looking at interatomic distances is even less of a justification to put them in the forefront. Skepticism and doubt help in avoiding the paradox of describing an aggregate whose cohesion energy is 100 by looking at bricks whose aggregation power is 2. No sensible journal should any more accept sentences like ''the crystal structure consists of pairs/chains/layers held together by such and such interactions...'' and similar assertions, when they are not supported by reliable energy numbers.

Admittedly, the weak point of molecular simulation is the assessment of the reliability and applicability of methods and force fields. Many good ones are now available; most if not all of the results presented here do not change substantially on changing procedures. Besides, if a simulation result is easily overthrown by a small change in parameterization, chances are that it is an unreliable result anyway.

The exploration of phase space is no longer limited to that isolated point which is the perfect static model of a crystal; we now can map the surroundings, we can see vibrational amplitudes and molecular rotation and diffusion, and we can even simulate with some success the molecular transition path for some simple phase transitions. Mature crystal engineering cannot dispense with these fundamental physicochemical aspects. For looking at the constellation of crystal structure, we now have a Galilean telescope: why not look into it?

Acknowledgements

Fairly priced access to software provided by the Cambridge Crystallographic Data Centre is gratefully acknowledged. Reference to original crystallographic work is also given for each CSD refcode.[12]

References

1 R. F. W. Bader, *Atoms in Molecules: a Quantum Theory*, Oxford University Press, 1990.
2 A. Gavezzotti, *Mol. Phys.*, 2008, **106**, 1473–1485. As often discussed in the literature, the partitioning offered by the PIXEL method is one of the many possible. While Coulombic energies are physically meaningful, being the numerical integration equivalent of the exact quantum chemical operators, polarization, dispersion and repulsion terms stem from parametrization. See *e.g.* J. D. Dunitz and A. Gavezzotti, *Angew. Chem., Int. Ed.*, 2005, **44**, 1766–1787.
3 E. Keller, *SCHAKAL92, A Program for the Graphic Representation of Molecular and Crystallographic Models*, University of Freiburg, 1993.
4 C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Woo, *J. Appl. Crystallogr.*, 2008, **41**, 466–470.
5 S. Z. Fisher, M. Aggarwal, A. Y. Kovalevsky, D. N. Silverman and R. McKenna, *J. Am. Chem. Soc.*, 2012, **134**, 14726–14729.
6 M. Carter and P. S. Ho, *Cryst. Growth Des.*, 2011, **11**, 5087–5095.
7 A. Gavezzotti and J. D. Dunitz, *Cryst. Growth Des.*, 2012, **12**, 5873.
8 A. Gavezzotti, *New J. Chem.*, 2011, **35**, 1360–1368.
9 A. Gavezzotti, *Molecular Aggregation*, Oxford University Press, Oxford, 2007, ch. 13.
10 A. Gavezzotti, *CrystEngComm*, 2011, **13**, 3573–3579. Note that 'biased' refers to the Monte Carlo steps, not to the energy states.
11 D. A. Bardwell *et al.*, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2012, **B67**, 535–551, and references therein.
12 (*a*) BAPLOT01, anhydrous theophylline: Y. Ebisuzaki, P. D. Boyle and J. A. Smith, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1997, **53**, 777; (*b*) THEOPH01, theophylline hydrate: C. Q. Sun, D. L Zhou, D. J. W. Grant and V. G. Young Jr, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2002, **58**, o368; (*c*) KIGLOC, theophylline-*p*-hydroxybenzoic acid: S. L. Childs, G. P. Stahly and A. Park, *Mol. Pharmaceutics*, 2007, **4**, 323; (*d*) WUYRIR, ethanolammonium theophyllinate: J. Madarasz, P. Bombicz, K. Jarmi, M. Ban, G. Pokol and S. Gal, *J. Therm. Anal. Calorim.*, 2002, **69**, 281; (*e*) DUXZAX, theophylline–urea, H. Wiedenfeld and F. Knoch, *Arch. Pharm.*, 1986, **319**, 654; (*f*) RIGYEM, theopylline–DMSO, C. Cardin, Y. Gan and T. Lewis, *Acta Crystallogr., Sect. E: Struct. Rep. Online*, 2007, **63**, o3175; (*g*) XEJXAM, theophylline–malonic acid: A. V. Trask, W. D. S. Motherwell and W. Jones, *Int. J. Pharm.*, 2006, **320**, 114; (*h*) AXSCHO, (*E*)-3*a*-acetoxy-5,10-seco-cholest-1(10)-en-5-one: H. Fuhrer, L. Lorenc, V. Pavlovic, G. Rihs, G. Rist, J. Kalvoda and M. L. Mihailovic, *Helv. Chim. Acta*, 1979, **62**, 1770; (*i*) BZCHOL03, cholesteryl benzoate: A. P. Polishchuk, V. I. Kulishov and Yu T. Struchkov, *Mol. Cryst. Liq. Cryst.*, 1985, **128**, 367; (*j*) CHOLOL, cholest-5-en-3-ol dihydrocinnamate: A. A. Polishchuk, M. Yu Antipin, R. G. Gerr, V. I. Kulishov, Yu T. Struchkov and L. G. Derkach, *Cryst. Struct. Commun.*, 1980, **9**, 263; (*k*) COMYEI, (22*S*,23*R*)-22,23-methylene-cholest-5-en-3*b*-ol acetate: M. P. Zimmerman, H.-T. Li, W. L. Duax, C. M. Weeks and C. Djerassi, *J. Am. Chem. Soc.*, 1984, **106**, 5602; (*l*) HINGAM, (22*S*,23*S*,25*R*)-3*b*-methoxy-16*b*,23:22,26-diepoxy-5*a*-cholestane: C. Betancor, R. L. Dorta, R. Freire, A. Martin, T. Prange and E. Suarez, *J. Org. Chem.*, 1998, **63**, 6355; (*m*) NAKMIV, 2,3-benzo-19,28,29-trinorcholest-2,7-diene-1,4-dione: F. De Riccardis, I. Izzo, C. Tedesco and G. Sodano, *Tetrahedron Lett.*, 1997, **38**, 2155; (*n*) PUXHAR, 3*b*-acetoxy-5*a*-cholestan-6-one: V. Rajnikant, V. K. Gupta, S. Shafi, E. H. Khan, J. Firoz, Shafiullah and R. Gupta, *Cryst. Res. Technol.*, 2001, **36**, 215; (*o*) QULWOJ, (20*R*,22*S*,24*R*)-3*b*-acetoxy-20,22-epoxy-24-methyl-5*a*-cholestan-6-one: B. Voigt, A. Porzel, C. Wagner, K. Merzweiler and G. Adam, CSD private communication, 2001; (*p*) SEGPOJ, cholesterol *p*-ethoxybenzoate: A. P. Polishchuk, M. Yu. Antipin, T. V. Timofeeva, O. D. Lavrentovich, A. S. Kotel'chuk, L. A. Taraborkin and Yu. T. Struchkov, *Kristallografiya*, 1988, **33**, 1134; (*q*) SOHVOA, 2*a*,3*a*-diacetoxycholest-4-en-6-one: A. A. Akhrem, N. M. Galitskii, G. A. Yasnitskii, V. I. Gritsuk, F. A. Lakhvich, V. A. Khripach and N. V. Kovganko, *Zh. Org. Khim.*, 1990, **26**, 771.