FORCE FIELDS AND CRYSTAL STRUCTURE PREDICTION

Bouke P. van Eijck (vaneijck@xs4all.nl)
Utrecht University (Retired)
Department of Crystal and Structural Chemistry
Padualaan 8, 3584 CH Utrecht, The Netherlands

Originally written in 2003
Update blind tests 2017

Contents

1 Introduction

Everybody who looks at a crystal structure marvels how Nature finds a way to pack complex molecules into space-filling patterns. The question arises: can we understand such packings without doing experiments? This is a great challenge to theoretical chemistry.

Most work in this direction uses the concept of a *force field*. This is just the potential energy of a collection of atoms as a function of their coordinates. In principle, this energy can be calculated by quantumchemical methods for a free molecule; even for an entire crystal computations are beginning to be feasible. But for nearly all work a parameterized functional form for the energy is necessary. An *ab initio* force field is derived from the abovementioned calculations on small model systems, which can hopefully be generalized to other related substances. This is a relatively new development, and most force fields are *empirical*: they have been developed to reproduce observed properties as well as possible. There exists a number of more or less time-honored force fields: MM3, CHARMM, AMBER, GROMOS, OPLS, DREIDING...

In the present notes it is assumed that the basic aspects of force fields are known. Normally there will be intramolecular energy terms for bond stretching, angle bending and internal rotation, although people still use models based on rigid molecules as building blocks for a crystal. Most essential is always the intermolecular energy, consisting of van der Waals terms (attractive as well as repulsive) and Coulomb energy. For an introduction the reader is referred to a recent review chapter [1].

2 Lattice Energy

The *packing energy P* is defined as the energy needed to break up a hypothetical non-vibrating crystal into its constituent non-vibrating free molecules. Often the *lattice energy L* is defined, which is just minus the packing energy:

$$L = -P \tag{1}$$

The packing energy is almost, but not quite, the same as the *sublimation enthalpy* ΔH. For nonlinear molecules an approximate relation is given by:

$$\Delta H = P - 2RT \tag{2}$$

Further complications arise if the molecular geometry in the gaseous state and in the crystalline state are appreciably different. Moreover, experimental uncertainties in sublimation enthalpies may be in the order of 5 kJ/mol [2]. So it is not really simple to correlate the temperature-independent packing energy with the observed temperature-dependent sublimation enthalpy.

If relaxation effects upon sublimation are neglected, the lattice energy can be calculated from the force field as the intermolecular part of the potential energy. In the usual

formulation the calculation involves a summation over distance-dependent atomic interactions $u(r)$ for pairs of atoms:

$$L = \tfrac{1}{2} \sum_k \sum_j \sum_{k'}{}' u_{kk'}(r_{k,jk'}) \tag{3}$$

where k indicates an atom in an asymmetric unit, j indicates a symmetry-related set of atoms (including the lattice translations), and $r_{k,jk'}$ denotes the distance between an atom k and an atom k' in set j. The prime signifies that intramolecular contributions must be omitted from the summation.

The simplest model of a crystal is a collection of rigid molecules stacked together in the best possible way. Sometimes such a description is adequate, but usually it is much better to take intramolecular degrees of freedom into account. This is especially important for flexible molecules, where the molecular conformation is intimately linked to the crystal packing. For instance, in carbohydrate crystals the hydroxyl group conformations must be such that intermolecular hydrogen bonds can be formed.

A severe limitation of most force fields is that electrostatics is introduced by means of partial charges on the atoms. These charges can be obtained empirically, as an integral part of the force field, by fitting to experimental data. In some force fields they must be found from fitting to an ab initio molecular potential. But for an accurate description of the electrostatic energy a model of point charges is insufficient. A better description of the molecular charge distribution is given by a multipole expansion which is distributed over various sites, usually the atomic positions [3]. Such a distribution can be found directly from the wave function for a free molecule or, preferably, from fitting to the electrostatic potential calculated from that wave function.

The assumption of pairwise additive energy contributions is liable to break down for partially charged atoms. Consider, for instance, two dipoles that polarize each other. The charges in the second dipole are enhanced by the first one, and so its interaction with a third dipole is not independent of the position of the first one. This is the origin of cooperative phenomena in hydrogen bonding. As discussed in Section 4.4, an approach where polarization is included explicitly can lead to considerable improvement, but at the cost of great complications in the calculation.

In practice, of course, only a finite number of terms can be taken into account in the calculation of the lattice energy (Eq. 3). The simplest solution is to introduce a maximum interatomic distance, the *cutoff* radius. However, for Coulomb interactions this may introduce unacceptable errors as illustrated in Figure 1, where the + $\cdots$ – attraction is included but the nearly compensating + $\cdots$ + repulsion is discarded.

So the cutoff must not break a neutral unit into charged parts. To this end it may be useful to divide a molecule into *charge groups* with zero or small total charge, and to base the cutoff criterion on the centers of these charge groups rather then on the atomic positions. Even then, convergence is bad and an acceleration procedure as detailed in section 2.2 is often preferable.

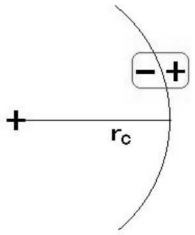

Figure 1: Breaking charge groups by the cutoff sphere.

2.1 Polar crystals

For Coulomb interactions Eq. 3 reduces to the direct summation:

$$U_{\text{direct}} = \frac{1}{8\pi\varepsilon_0} \sum_{k=1}^{N} \sum_{j=1}^{\infty} \sum_{k'=1}^{N}{}' \frac{q_k q_{k'}}{r_{k,jk'}} \tag{4}$$

where q_k is the partial charge on atom k. Here and in the next sections we consider the energy of an entire crystal cell, containing N atoms and having a volume V.

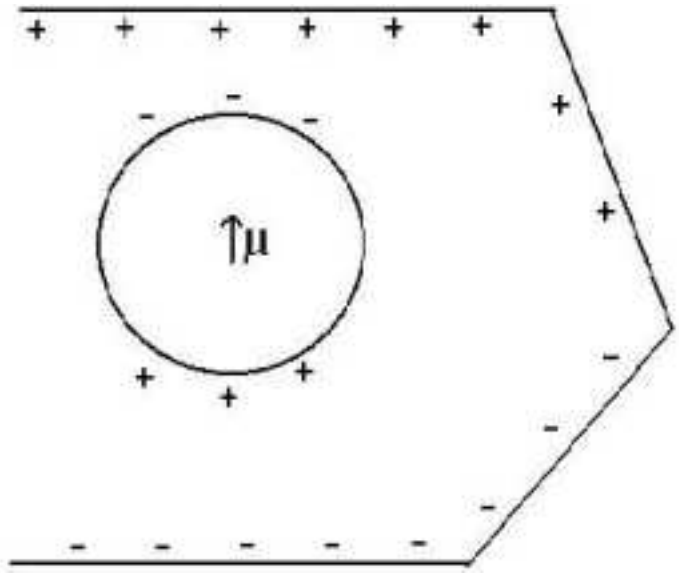

Figure 2: Surface charge in polar molecules.

Here a complication arises for polar crystals, i.e. crystals with a dipole moment μ in the unit cell. Figure 2 shows how charges will accumulate on the outside of the cutoff sphere as well as on the inside of the outer boundary of the crystal. Their influence on a charge in the center of the cutoff sphere can be calculated by classical electrostatics.

For the first contribution this gives a simple closed formula, but for the second one the result depends on the external shape of the crystal:

$$U_{\text{Coulomb}} = U_{\text{direct}} - \frac{\mu^2}{6\varepsilon_0 V} + U_{\text{surface}} \tag{5}$$

For spherically or cubically shaped crystals the surface term is given by:

$$U_{\text{surface}} = \frac{\mu^2}{6\varepsilon_0 V} \tag{6}$$

so, not unexpectedly, $U_{\text{Coulomb}} = U_{\text{direct}}$. For a needle or a platelet with the dipole moment perpendicular to a small face the surface term is zero. Formulae for any parallelepiped have been published [4].

This theory is valid for a hypothetical isolated polar crystal. However, it is very doubtful whether the surface term should be included in energy calculations for the real world. It would imply that lattice energies for polar crystals cannot be tabulated without knowing the crystal form. It has been suggested that crystals will tend to find a form where the surface term is zero, or that external charges will accumulate on the surface to annul the surface charge exactly ("tin foil boundary conditions"). We suggest to omit the surface term from energy calculations.

2.2 Convergence Acceleration

Apart from the complications for polar crystals, the direct summation (Eq. 4) would require a very large cutoff radius to obtain anything resembling an acceptable convergence. In an empirical force field this problem may not be so serious if the same cutoff is used as was done during the parameterization. But in non-empirical force fields a careful calculation of the Coulomb energy is important. Following methods devised by Madelung and Ewald it can be shown that (for non-polar crystals) the exact Coulomb energy is given by:

$$\begin{aligned} U_{\text{Ewald}} \;=\; & \frac{1}{8\pi\varepsilon_0} \sum_{k=1}^{N} \sum_{j=1}^{\infty} \sum_{k'=1}^{N}{}' \frac{q_k q_{k'} \operatorname{erfc}(\alpha r_{k,jk'})}{r_{k,jk'}} + \\ & \frac{1}{8\pi^2 V \varepsilon_0} \sum_{\mathbf{h}\neq\mathbf{0}} \frac{|F_{\mathbf{h}}|^2 \exp(-\pi^2 h^2/\alpha^2)}{h^2} - \frac{\alpha}{4\pi^{3/2}\varepsilon_0} \sum_{k=1}^{N} q_k^2 \end{aligned} \tag{7}$$

The parameter α can be chosen to obtain optimal convergence. N is the number of atoms in the unit cell, V is the cell volume, $\mathbf{h}$ is a vector in reciprocal space and $F_{\mathbf{h}}$ is defined as:

$$F_{\mathbf{h}} = \sum_{k=1}^{N} q_k \exp(2\pi i \mathbf{h}\cdot\mathbf{r}_k) \tag{8}$$

where $\mathbf{r}_k$ is the position of atom k in the unit cell. This functional form is formally identical to the structure factor in diffraction theory.

A careful derivation of Eq. 7 shows that the omission of the terms with $\mathbf{h} = \mathbf{0}$ is not allowed for polar crystals. Here a fourth term arises, which turns out to be exactly the same as U_{surface}:

$$U_{\text{Coulomb}} = U_{\text{Ewald}} + U_{\text{surface}} \tag{9}$$

Comparison with Eq. 5 shows:

$$U_{\text{Ewald}} = U_{\text{direct}} - \frac{\mu^2}{6\varepsilon_0 V} \tag{10}$$

It is important to realize that this equation is valid independent of the question whether or not the surface energy is included (see Section 2.1). That decision must be based on physical phenomena, and has nothing to do with the use of Ewald summation which is just a mathematical trick.

Equations for convergence acceleration have also been developed for other inverse powers of the interatomic distances. This is especially useful for the dispersion interactions, which depend on r^{-6}. These terms converge much faster than for the Coulomb energy, but they are all attractive and the sum of many small contributions from outside the cutoff sphere may not be negligible. Explicit expressions for energies and forces have been given by Karasawa and Goddard [5].

2.3 Energy Minimization

In the static model a crystal structure corresponds to a minimum in the potential energy with respect to a set of parameters: the cell axes, cell angles, and the atomic coordinates. Their exact number depends on the space group symmetry. One of the insights obtained from attempts at crystal structure prediction is that the number of potential energy minima is enormous. Each minimum corresponds to a possible polymorph. The *global minimum* has the lowest energy, all other ones are *local minima*. Figure 3 illustrates a small portion of one of the many dimensions of the potential energy surface. The deepest minimum (C) could correspond to a low-temperature polymorph, the broader one (A) could become more favourable at higher temperature because it is favoured by entropy.

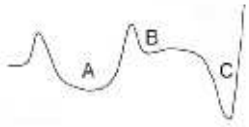

Figure 3: Minima in the potential energy function.

There is no general solution to the problem of finding all minima, or even only the global one. All one can do is to start with a certain set of parameters, and then apply numerical methods to find the nearest local minimum. Computationally this is not a trivial problem; for details we refer to the literature [6, 7].

One application is to start from the experimental crystal structure. This should correspond to an energy minimum, but in an actual force field it never does. The parameters will show a shift upon energy minimization, which is a measure of the inadequacy of the force field used. Force fields are often developed with the objective to minimize these shifts, while also reproducing observed lattice energies (and sometimes also vibrational frequencies). For such an *empirical* force field one expects deviations of a few % in cell axes and a few degrees in cell angles. Also the positions and orientations of the molecules in the unit cell should be reproduced with comparable accuracy, as well as intermolecular contact distances. Especially important is the reproduction of torsional angles in flexible molecules, since the crystal packing is very sensitive to the molecular conformation. Sublimation enthalpies can hardly be expected to be better reproduced than within, say, 5 - 10 kJ/mol [2].

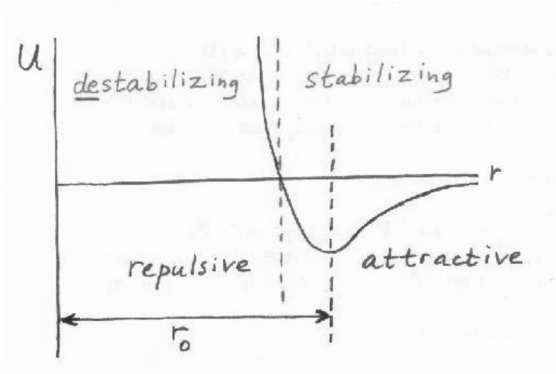

Figure 4: Nonbonded interaction curve.

A few remarks must be made about van der Waals distances. Figure 4 shows the nonbonded interaction curve. Note the attractive and repulsive regions (separated by the minimum distance r_0) compared to the stabilizing and destabilizing regions (separated by $u = 0$). Although r_0 is sometimes interpreted as a van der Waals distance, that is not the definition adopted by crystallographers who take it to be the distance between two "touching" atoms in neighboring molecules. This distance is always smaller than r_0 since there are many other pairs of atoms further away, whose attraction must press the touching atoms into the repulsive region.

3 Temperature effects

Observed cell dimensions and structural details depend on the temperature. But a force field, which is a function of only the coordinates, cannot provide the energy as a function of T. If the force field was parameterized against observed data, it will hopefully be able to reproduce similar data at some average temperature. If it was parameterized against quantumchemical data, it will refer to a hypothetical vibrationless state which does not even exist at 0 K. Then the calculated cell volume is expected to be, say, 5% smaller than observed at room temperature.

Corrections for temperature effects can be calculated under the assumption that deviations from the equilibrium geometry are small enough to allow the use of the harmonic approximation. Then the well-developed theory of lattice vibrations can be used to calculate the thermal effects on energy, entropy and free energy. This approach is summarized in the next section.

3.1 Lattice Vibrations

Any textbook on statistical thermodynamics explains how the free energy of a harmonic oscillator can be found from its vibrational frequency (ν), which in its turn can be found from the mass and the force constant:

$$4\pi^2\nu^2 = \frac{1}{m}\frac{\partial^2 U}{\partial r^2} \tag{11}$$

The same principle applies to harmonic lattice vibrations in a crystal, only the computation is more complex. It can be shown [8] that the normal vibrations for each *wave vector* ($\mathbf{q}$), which is a vector in reciprocal space, can be treated separately. To this end we must find the eigenvalues of the *dynamical matrix*:

$$D_{kg,k'g'}(\mathbf{q}) = \frac{1}{\sqrt{m_k m_{k'}}}\sum_{j'}\left(\frac{\partial^2 U}{\partial r_{jkg}\partial r_{j'k'g'}}\right)\exp[i\mathbf{q}\cdot(\mathbf{r}_{j'k'} - \mathbf{r}_{jk})] \tag{12}$$

Here k, k' label the atoms in a unit cell, j, j' label unit cells, and $g, g' = x, y, z$. Note that, due to translational symmetry, this matrix does not depend on j. The essential quantity to be calculated is the second derivative of the potential energy with respect to the atomic coordinates.

There are $3N$ eigenvalues of the dynamical matrix, each corresponding to $4\pi^2\nu^2(\mathbf{q})$. They lead directly to the free energy [9, pg. 197]:

$$A = L + < \sum_{n=1}^{3N}[\tfrac{1}{2}h\nu_n(\mathbf{q}) + kT\ln(1 - e^{-h\nu_n(\mathbf{q})/kT}) >_{\mathbf{q}} \tag{13}$$

Here the average is over a suitable number (10 - 100) of wave vectors, chosen systematically or randomly within a unit cell of reciprocal space. The calculation can

be done for different polymorphs, with individual frequency spectra and lattice energies. Note the zero-point vibrational energies which may have an important effect at all temperatures.

This theory is incomplete since only harmonic terms in the potential energy are considered. Its most obvious defect is that there is no thermal expansion [8, 10]. Extending the theory to anharmonic effects is rather difficult. Numerically, the calculations outlined above can be carried out at several geometries around the energy minimum, after which the structure with lowest free energy can be selected. This should give the desired temperature dependence of the crystal structure.

4 Prediction of Crystal Structures

As discussed above, important observable quantities that should be reproduced after energy minimization in a certain force field are the sublimation enthalpy and the crystal geometry. It was also noted that it is essential to have a rough starting geometry before one can start an energy minimization at all. For force field development this starting geometry is generally the experimental crystal structure. Sometimes several data on one structure are available, for instance at different temperatures, and their differences immediately indicate the best that can be expected from a force field without some explicit temperature dependence.

Even more interesting is the case where more than one polymorph is known. It is important to remember that the energy landscape for a crystal has many local minima, each corresponding to a possible polymorph. Energy minimization will only lead to the nearest local minimum. This phenomenon reminds us that one should not be content with reproduction of the observed geometry, but that the real goal of crystal modeling should be *crystal structure prediction*. This is an exciting field of research that has been explored during the last decade. A few review papers are available [11, 12, 13].

The problem of crystal structure prediction can be divided in two stages. In the first stage we have to generate possible crystal structures. In the second stage we have to find out which ones correspond to observable polymorphs. Contrary to what one might expect, the second stage is usually the more difficult one. I would like to make the point that, if one could predict crystal structures reliably, one would have not only a very good force field but also a fundamental understanding of crystallization phenomena and crystal physics in general.

4.1 Stage 1: generation of possible structures

The first stage of crystal structure prediction is to make a list of structures that could possibly occur. Various strategies to this end have been published. Some of these have used database statistics, but most have been based on energies calculated by an empirical force field. The currently available programs and methods can be found in

the reports on the two Cambridge blind tests on crystal structure prediction [14, 15], summarized in Section 4.3.

Quite a few unknown parameters are involved: up to six for the unit cell, and for each independent molecule three positional parameters, three orientational parameters, and possibly a number of parameters that determine the molecular conformation. One approach [16] is to start with clusters of molecules and then to combine these to form a three-dimensional lattice. Alternatively [17], the space group is chosen in advance and possible structures are generated by one of various search strategies: random search, grid search, Monte Carlo or molecular dynamics.

The first decision is the number of independent molecules in the asymmetric unit. Let us call Z the number of residues in a unit cell, Z' the number of residues divided by the multiplicity of a general position, and Z'' the number of crystallographically independent molecules. The difference between Z' and Z'' is subtle: a molecule on a special position may have $Z' = 1/2$ but $Z'' = 1$; likewise, for a hydrate Z' may be equal to 1 but Z'' must be larger. The CSD contains about 10% hydrates and 6% other solvates; for homomolecular structures 8% of the structures has $Z' > 1$. Yet in crystal structure prediction it is nearly always assumed that $Z'' = 1$.

The second decision concerns the space groups to be studied. In principle it would be sufficient to assume $P1$ with various values for Z'', which should lead to every conceivable crystal structure. In practice crystal structure prediction tends to become impossible for $Z'' > 2$, so it is necessary to fix the space groups to be studied. Fortunately, 95% of the molecules crystallize in only 8 space groups (and less for chiral molecules). Most programs do not allow for molecules on special positions, but in that case the structure can often be found in a space group with lower symmetry.

The third decision concerns the geometry of the crystal building blocks, usually one individual molecule. Many studies concern the methodology of crystal structure prediction, so they are performed on substances with known structures. One should then resist the temptation to take the molecular building block from the observed structure, which is quite unfair - especially for flexible molecules. One should use model building and energy minimization for the free molecule. For flexible molecules several conformations are possible. Some of them may have a relatively high intramolecular energy, but that may be compensated by a favourable packing energy.

A complete search usually finds the important structures more than once. The equivalence is not always easily recognized, because different space group settings are possible (Figure 5). An algorithm to cluster equivalent structures can be based on a comparison of interatomic distances, which should be exactly the same after full energy minimization.

The final result of the first stage is a list of unique hypothetical structures. This list is often surprisingly long, with many structures in an energy range of only a few kJ/mol. Not all of these correspond to possible polymorphs. Some may be intrinsically instable because they correspond to saddle points in the energy. This can be verified by inspecting the normal vibrations for imaginary frequencies, or by repeating the energy

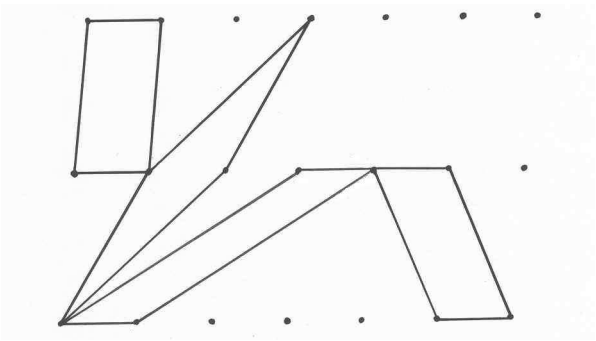

Figure 5: Equivalent settings for the same structure.

minimization after expansion to one or more unit cells in space group $P1$ to remove restrictions in the parameter space. Other minima may be thermally instable because they can easily convert into more favourable structures; see for example minimum B in Figure 3. Such structures can be eliminated by molecular dynamics simulations.

Nevertheless, it is unlikely that the numbers of distinct hypothetical structures will ever become really low. This suggests an enormous potential for polymorphism which, theoretically, should be the rule rather than the exception. Indeed, according to McCrone [18] the number of polymorphs of a given substance is proportional to the effort that one is prepared to put into crystallization experiments. Considering that Sarma and Desiraju [19] have identified only 3.5% of the entries in the CSD as polymorphic, a fertile field of research should still be open here. In any case there is always the possibility that hypothetical structures with low energy can actually be realized in nature, but have never been observed.

4.2 Stage 2: selection of the right structure(s)

In the second stage of the structure prediction the most probable candidates for experimentally observable structures must be selected. Unfortunately, the thermodynamical approach (looking for the structure with the lowest free energy) may not be sufficient. The crystallization process is a complex combination of thermodynamics and kinetics. There are cases known where different polymorphs crystallize even simultaneously from the same solution! But as such phenomena are difficult to understand quantitatively, this possibility is generally disregarded.

Furthermore, many studies are limited to energy calculations. So, in fact, the calculated structures refer to a static structure, and no transition temperatures between polymorphs can be predicted. Thus we are neglecting the entropy, the vibrational contributions to the energy, and the thermal expansion. The situation is illustrated in Figure 6. Some of these corrections may be important, and the only possible justification for their neglect is that errors in the force field are probably even larger. The effects of

pressure are easily incorporated by adding a PV term to the energy, thus changing to enthalpy.

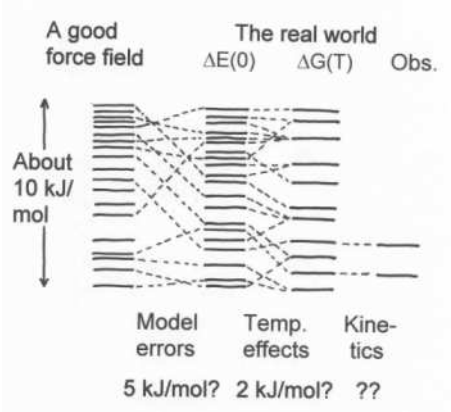

Figure 6: Energies and free energies from modeling compared with the real world.

At the end of the first stage the key question, of course, is whether or not the observable polymorph(s) are present in the final list of structures. Assuming that the search has covered the right space groups, the right Z'' and the right molecular conformations, experience shows that the chances of success decrease rapidly for 20 or more degrees of freedom. The answer to the question is roughly [20]:

- "usually", for rigid molecules with $Z'' = 1$
- "often", for flexible molecules with $Z'' = 1$ or simple rigid molecules with $Z'' = 2$
- "seldom", for flexible molecules with $Z'' > 1$

It is not always trivial to recognize an observed polymorph in the list of possible structures: what comes out of the calculations can only be an energy-minimized structure. The observed structure will have a different geometry, and quite probably a different space group setting. A pragmatic solution is to compare the powder diffraction diagrams. If no similarity can be found, the prediction may be said to have failed. But the yes or no decision is not always straightforward. A more objective solution is to search the list for the energy-minimized experimental structure, which should be retrieved in every possible detail. However, if this structure is greatly deformed by the energy minimization we can hardly call the structure prediction really successful. Important measures of success are further the energy difference (ΔE) and the ranking (R) with respect to the global energy minimum (Figure 7). A sloppy search will find only a few structures. With luck, the experimental structure is one of them, and it will have a good ranking because there are so few structures. A careful search will find more structures and so the ranking and energy difference can only get worse. Not all referees appreciate this fact.

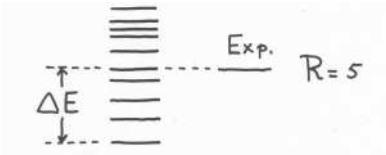

Figure 7: The energy difference (ΔE) and ranking (R) of an energy-minimized experimental structure with respect to the global energy minimum.

As an example of the difficulty of crystal structure prediction in flexible molecules we refer to our study [21] of 32 small anhydrous pyranoses, selected to have $Z'' = 1$ in space groups $P1$, $P2_1$ or $P2_12_12_1$. One of them was missed in the search. The results for the other substances are summarized in Table 1, where two force fields (UNITAT and OPLS) are compared. These force fields should have been specially suitable for this class of compounds, using modeling by united atoms and all atoms, respectively. Of course, the energy ordering is affected by the choice of the force field. In this study the RMS difference between all hypothetical structures in the two force fields was 10 kJ/mol .

Table 1: Results of crystal structure prediction for 31 pyranoses in two empirical force fields.

Force field	UNITAT	OPLS
Worst R	192	529
Median R	16	10
Successful predictions ($R = 1$)	4	5
Worst ΔE (kJ/mol)	16.2	34.1
Average ΔE (kJ/mol)	6.5	8.9

Gavezzotti [22] has evaluated the energy differences between known polymorphs, and estimated that they very seldom exceed 20 kJ/mol. Assuming polymorphic equilibrium, this must also be the order of magnitude of the entropic effects:

$$\Delta U(T) = T\Delta S(T) \tag{14}$$

So it should not be surprising that we cannot expect reliable results in standard structure prediction: 10 kJ/mol is the uncertainty of the force field as well as the effect of the neglected entropy differences. So for progress we need better force fields, but also better understanding of thermal effects.

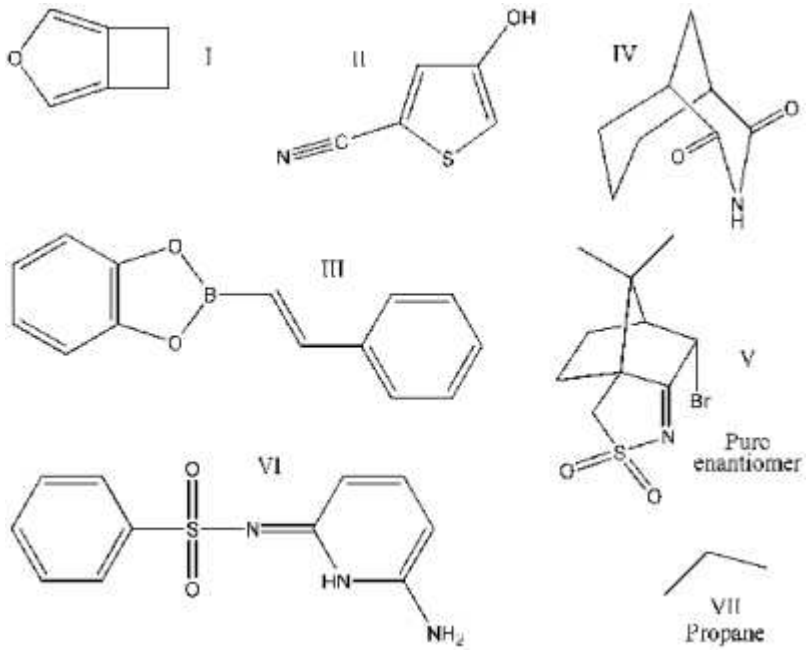

Figure 8: The molecular diagrams given to the participants in the blind tests 1999 (I-III, VII) and 2001 (IV-VI).

4.3 Blind tests

As noted above, many publications on crystal structure prediction have known structures as subject. Even for the most honest investigator there is the temptation to continue the work until the experimental structure is at least found, albeit possibly with a disappointingly high ranking. Therefore it was a very good initiative of the Cambridge Structural Database to organize two blind tests, in 1999 [14] and in 2001 [15]. There were 11 researchers invited to participate in the first test and 17 in the second. In both tests the chemical diagrams of three compounds were given, with the structure of propane (which was already in the press) as a bonus in the first test. The only further information was that there was one independent molecule in one of the most populated space groups. A maximum of three predictions per molecule was allowed.

The seven molecules are shown in Figure 8. Not every participant tried to predict every structure. The rankings of the correct predictions are given in Table 2. It is seen that there were 13 submitted structures that were accepted according to the rules of the blind test, but that there were only 8 genuine predictions ($R = 1$). The case of substance I is interesting: during the data collection the cooling broke down and the crystal melted. After recrystallization only another polymorph (Ib) could be obtained, so this is presumably the stable one. But only the first polymorph was predicted, notably by four participants.

It is clear that the present possibilities of crystal structure prediction are rather limited. Especially for flexible molecules (III and VI) the results are poor. For details of all methods we must refer to the original papers. Here we just note that the only successful result for propane (structure VII) was obtained by a purely ab initio method, as discussed in the next section.

Table 2: Rankings of correct predictions in the first two Cambridge blind tests

Test	Molecule	Participants	Rankings
1999	Ia	11	1, 1, 1, 3
	Ib	11	–
	II	8	2
	III	11	1
2001	IV	15	2, 3
	V	15	1, 1, 1, 2[†]
	VI	11	–
1999	VII	6	1

[†] Purists might reject one of these entries because the structure is rather deformed and another one because it was submitted after the experimental structure was revealed.

4.4 Beyond empirical force fields

The importance of better force fields has been shown by the group of Price [3], who has always stressed the importance of good electrostatics. Not only charges, but also dipoles and quadrupoles on the atoms are needed. These are obtained by a distributed multipole analysis (DMA) of the charge density calculated from an ab initio wave function. Apart from electrostatics, there are also the Van der Waals terms: attraction at long distances, repulsion at short distances. This part of the force field was empirical, although in recent work more sophisticated approaches were developed [23]. The results are quite impressive. For rigid hydrogen-bonded molecules the rankings are nearly always one. A weak point is the use of rigid models, which may explain a certain noise in the energies: there are differences between the energies of structures that should be identical.

In our group we have followed this approach, and extended it to the development of a completely ab initio intermolecular force field [24] for aliphatic molecules containing only C, H and O. This was done by parameterization on quantumchemical energies for methanol dimers and trimers. Moreover, first-order intermolecular polarization was included. This is essential to obtain transferability of the force field, from gas molecules to condensed phases and from methanol to larger molecules. For rigid molecules very promising results were obtained [25]: it was possible to predict the observed crystal

structures of several small rigid molecules, including one case (propane, see above) where the structure was really unknown in advance.

For nonrigid molecules the intramolecular energy is the next bottleneck. Again, empirical force fields were found to be insufficiently reliable. For small molecules it is now feasible to perform ab initio energy minimizations "on the fly" by calculating the energy of a molecule in the geometry imposed by its crystal surroundings. By also calculating the first and second derivatives of the energy, a local harmonic force field can be created [26] which is then combined with the intermolecular contributions. Of course, this must be redone when the geometry changes too much during the energy minimization, and the calculations are rather time-consuming. After introduction of thermal corrections, calculated in the standard harmonic approximation (Section 3.1), good rankings were found for glycol and glycerol [27]. Even better results were obtained for six hexapyranoses [28], as shown in Table 3. The structures were energy-minimized at the SCF/6-31G* level for intramolecular energies, and recalculated at that geometry with a more sophisticated technique (DFT/PW91-EXT). The RMS differences of the ab initio intramolecular energies with the values found earlier for two empirical force fields (see Section 4.2) were around 9 kJ/mol for UNITAT and 11 kJ/mol for OPLS. The improvement on changing from energies (ΔE) to free energies (ΔA) is seen to be significant. It was found that the temperature dependence of the energy and the thermal expansion do not have a large influence, but the entropy and the zero-point energy can contribute decisively. Because of the zero-point vibrations, consideration of only energy would be insufficient even at absolute zero.

Table 3: Ab initio relative energies for six hexapyranoses.

	SCF/6-31G*				DFT/PW91-EXT				
	ΔE	R_E	ΔA	R_A	ΔE	R_E	ΔA	R_A	D
α-D-galactose	(16.8)	1	(11.7)	1	(14.5)	1	(10.9)	1	4.0
α-D-glucose	(4.8)	1	(7.3)	1	(4.7)	1	(7.4)	1	4.0
α-D-talose	3.2	2	2.4	2	(0.7)	1	(1.5)	1	2.4
β-D-allose	4.5	3	0.5	2	2.9	5	(0.4)	1	3.1
β-D-galactose	6.4	3	1.0	2	4.7	3	(0.7)	1	3.5
β-D-glucose	1.9	3	(3.6)	1	6.2	5	1.3	2	3.1

(Free) energy differences ΔE, ΔA (kJ/mol) and rankings R_E, R_A refer to the experimental structure with respect to the global (free) energy minimum at 300 K, as found from 20 selected structures for each substance. In case of a ranking $R = 1$ the entry in parentheses refers to the (free) energy difference with the second best structure. Intermolecular energies from SCF/DZ^O multipoles, intramolecular energies from levels as given, D being the root mean square difference (kJ/mol) between the two sets.

4.5 Conclusions

These results suggest that better calculation of thermodynamic quantities, especially energies, is a prerequisite for progress in crystal structure prediction. Of course, kinetic effects will occasionally be responsible for the observation of thermodynamically metastable structures. Indeed, "Ostwald's rule" says that these structures are even preferentially formed, at least in the initial stages of crystallization. So let us propose a more modest demand for crystal structure prediction: the free energy of an experimental structure should not be unreasonably higher than that of any other possible structure. In a careful study it should be possible to produce a fairly limited list of possible structures. From the point of view of force field development, a force field that calculates a high relative free energy for an observed structure (say, over 7 kJ/mol) should be regarded as inadequate.

It might be expected that some representation of entropic effects, or even the crystallization process, would be possible by closer inspection of the structure generation. For example, looking at Figure 3 one would expect that a random search method would find minimum A more often than minimum C, despite its higher energy. But in our experience experimentally observed structures are not found significantly more frequently than others.

It is interesting to note that, for a given substance in a given force field, it is possible to calculate energy and density without knowledge of the crystal structure! This is because all the most promising hypothetical structures usually have rather similar values for these properties.

Completely reliable *ab initio* crystal structure prediction may remain an illusion. In 1994 Gavezzotti wrote a much-quoted article [22] titled "Are crystal structures predictable?", and suggested that the answer might very well be "No". But even if that were true, trying to improve the present possibilities provides very exciting research - which one can do at present on computers as encountered in many private homes.

4.6 Update 2017

After writing these notes in 2003, four more blind tests for crystal structure prediction have been held in Cambridge [29, 30, 31, 35]. Some results are given in Table 4. In the last test [35] more than 3 submissions were allowed and not all participants worked independently, so the present comparison with the previous tests is somewhat arbitrary. The interested reader should consult the paper.

In the 2007 test [30] all four structures were correctly predicted by one group of participants (Neumann, Leusen and Kendrick). They developed a sophisticated and computationally highly expensive method based on DFT calculations for the entire crystal [32, 33, 34]. In the last two tests this approach gave again excellent results, and it is obviously the direction into which further research should proceed.

Table 4: Rankings of correct predictions in the last four Cambridge blind tests

Test	Molecule	Participants	Rankings
2004	VIII	15	1, 1, 1, 2[†]
	XIX	10	1
	X	12	–
	XI	13	– ($Z'' = 2$)
2007	XII	13	1, 1, 2, 2
	XIII	12	1, 1, 1, 1
	XIV	11	1, 1, 1
	XV	9	1, 3
2010	XVI	15	1, 2
	XVII	13	1, 2
	XVIII	13	1
	XIX	11	2, 3
	XX	10	1, 1
	XXI	10	–
2015	XXII	21	1, 1, 2, 2, 3, 3, ...
	XXIII[#]	14	1, 2, ...
	XXIV	8	2
	XXV	14	1, 1, 1, 2, ...
	XXVI	12	1, 2, ...

[†] Purists should reject these entries because the structure turned out to have been published previously.
[#] For XXIII 5 polymorphs were known, two of them with $Z'' = 2$. Altogether only two were found with ranking less than 4.

References

[1] B. P. van Eijck, L. M. J. Kroon-Batenburg, and J. Kroon, Energy minimization and molecular dynamics calculations for molecular crystals, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, edited by A. Gavezzotti, pages 99–146, John Wiley and Sons, Chicester, 1997.

[2] A. Gavezzotti and G. Filippini, Energetic aspects of crystal packing: Experiment and computer simulations, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, edited by A. Gavezzotti, chapter 3, pages 61–97, John Wiley & Sons, 1997.

[3] D. S. Coombes, S. L. Price, D. J. Willock, and M. Leslie, J. Phys. Chem. **100**, 7352 (1996).

[4] B. P. van Eijck and J. Kroon, J. Phys. Chem. **B101**, 1096 (1997).

[5] N. Karasawa and W. A. Goddard III, J. Phys. Chem. **93**, 7320 (1989).

[6] U. Burkert and N. L. Allinger, *Molecular Mechanics*, American Chemical Society, Washington, 1982.

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes, the Art of Scientific Computing*, Cambridge University Press, Cambridge, 1989.

[8] B. T. M. Willis and A. W. Pryor, *Thermal Vibrations in Crystallography*, Cambridge University Press, Cambridge, 1975.

[9] D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, 1976.

[10] C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1966.

[11] R. J. Gdanitz, *Ab initio* prediction of possible molecular crystal structures, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, edited by A. Gavezzotti, chapter 6, pages 185–201, John Wiley and Sons, Chicester, 1997.

[12] P. Verwer and F. J. J. Leusen, Computer simulation to predict possible crystal polymorphs, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd, volume 12, pages 327–365, Wiley-VCH, New York, 1998.

[13] T. Beyer, T. Lewis, and S. L. Price, CrystEngComm **3**, 178 (2001).

[14] J. P. M. Lommerse et al., Acta Cryst. **B56**, 697 (2000).

[15] W. D. S. Motherwell et al., Acta Cryst. **B58**, 647 (2002).

[16] A. Gavezzotti, J. Am. Chem. Soc. **113**, 4622 (1991).

[17] H. R. Karfunkel and R. J. Gdanitz, J. Comput. Chem. **13**, 1171 (1992).

[18] W. C. McCrone, Polymorphism, in *Physics and Chemistry of the Organic Solid State*, edited by D. Fox, M. M. Labes, and A. Weissberger, volume II, pages 725–767, Interscience, New York, 1965.

[19] J. A. R. P. Sarma and G. R. Desiraju, in *Crystal Engineering. The Design and Application of Functional Solids*, edited by K. R. Seddon and M. J. Zaworotko, Kluwer, Dordrecht, 1998 (in press).

[20] B. P. van Eijck and J. Kroon, Acta Cryst. **B56**, 535 (2000).

[21] B. P. van Eijck and J. Kroon, J. Comput. Chem. **20**, 799 (1999).

[22] A. Gavezzotti, Acc. Chem. Res. **27**, 309 (1994).

[23] J. B. O. Mitchell and S. L. Price, JPC **A104**, 10958 (2000).

[24] W. T. M. Mooij, F. B. van Duijneveldt, J. G. C. M. van Duijneveldt - van de Rijdt, and B. P. van Eijck, J. Phys. Chem. **A103**, 9872 (1999).

[25] W. T. M. Mooij, B. P. van Eijck, and J. Kroon, J. Phys. Chem. **A103**, 9883 (1999).

[26] B. P. van Eijck, W. T. M. Mooij, and J. Kroon, J. Comput. Chem. **22**, 805 (2001).

[27] B. P. van Eijck, J. Comput. Chem. **22**, 816 (2001).

[28] B. P. van Eijck, W. T. M. Mooij, and J. Kroon, J. Phys. Chem. **B105**, 10573 (2001).

[29] G. M. Day et al., Acta Cryst. **B61**, 511 (2005).

[30] G. M. Day et al., Acta Cryst. **B65**, 107 (2009).

[31] D. Bardwell et al., Acta Cryst. **B67**, 535 (2011).

[32] M. A. Neumann and M.-A. Perrin, J. Phys. Chem. **B109**, 15531 (2005).

[33] M. A. Neumann, J. Phys. Chem. **B112**, 9810 (2008).

[34] A. Asmadi et al, J. Phys. Chem. **B113**, 16303 (2009).

[35] A. M. Reilly et al., Acta Cryst. **B72**, 439 (2016).

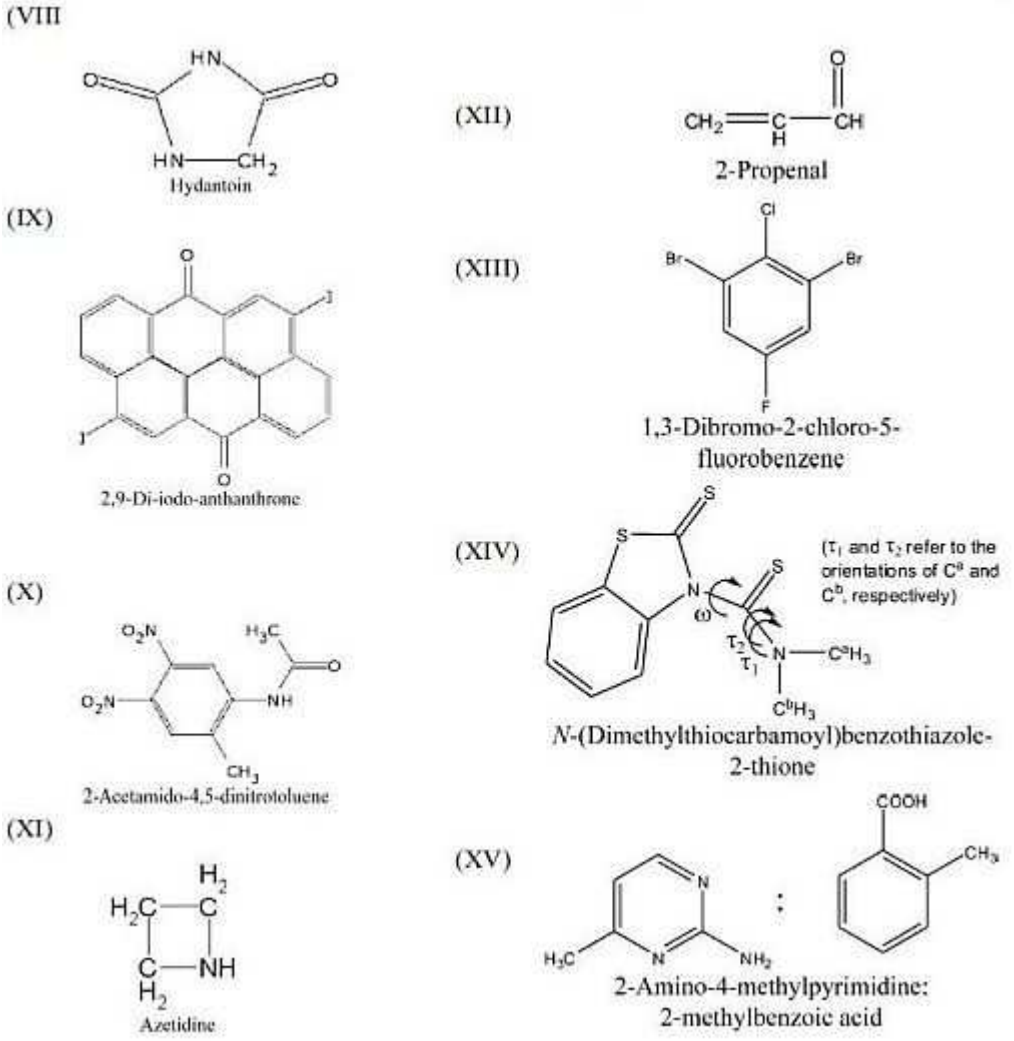

Figure 9: The molecular diagrams given to the participants in the blind tests 2004 (VIII-XI) and 2007 (XII-XV)

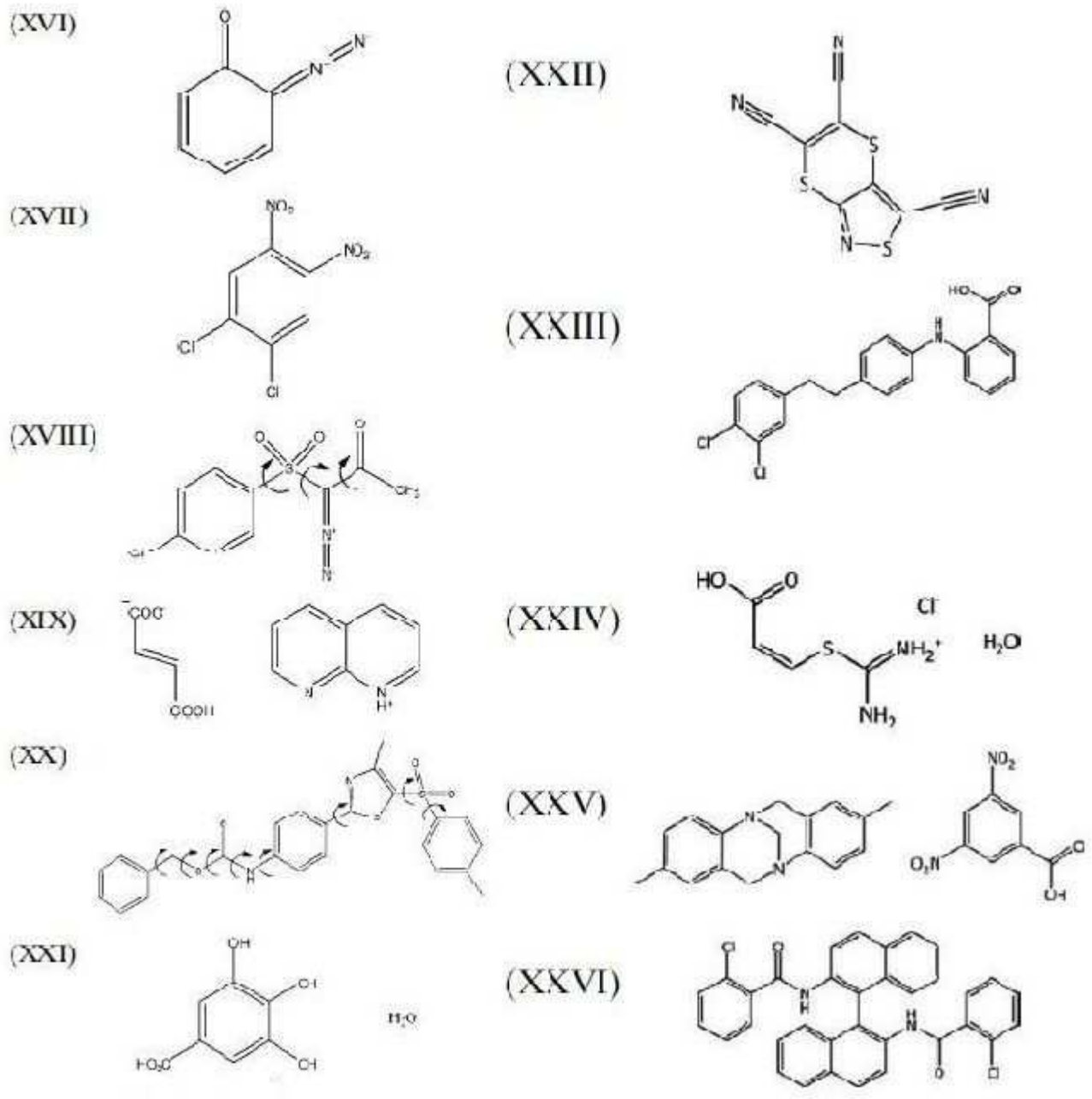

Figure 10: The molecular diagrams given to the participants in the blind tests 2010 (XVI-XXI) and 2017 (XXII-XXVI)