Inorganic Crystal Structure Prediction – a Dream Coming True?

Anton Meden

Faculty of Chemistry and Chemical Technology, University of Ljubljana, Aškerčeva 5, SI-1000 Ljubljana

Abstract

Various approaches to gain understanding of crystal building principles are classified in two groups. The first being crystal-chemical analysis of known structures and deriving crystallographic rules thereof, while the second is a quantum-chemical stability calculation, combined with local or global energy minimization. Both approaches are discussed in terms of their applicability to crystal structure prediction of inorganic solids. The meaning of a successful structure prediction is defined and its applicability outlined. Anticipation of further development is given at the end.

Keywords: Inorganic crystal structure prediction, data mining, quantum chemical calculations.

Introduction

Among the existing definitions of crystal structure prediction the preferred one is of LeBail[1] "the final aim of structure prediction should be to announce a crystal structure before any confirmation by chemical synthesis or discovery in nature". It is short and unambiguous, but for the purpose of the present article it needs an addition that our ultimate wish is to announce only those crystal structures that have real chance to be synthesized or found in nature. The term "real chance to exist" is of course insufficiently defined at present and can not serve as an exact definition. But it does indicate the way of the author's thinking and hopefully the future development will bring exact criteria to limit the predicted structures to only those with high probability to be successfully synthesized.

Bearing in mind that one piece of information contained in the "crystal structure" is also chemical formula, the benefit of predicting "really existing" structures is obvious. The synthesis efforts could be focused to produce new compounds with a good chance of success. Even more exciting is the anticipation that from a given (predicted) crystal structure, the physical properties could be reliably calculated. Then the search for suitable materials to use in novel applications could be first done theoretically in a computer and only then the target material would be prepared (Fig. 1). Structure prediction would thus help finding and preparing novel compounds and facilitate their structural analysis, when they are synthesized. Prediction is, however, not going to replace the classical single crystal or powder diffraction based structure analysis. Experimentally obtained crystallographic data will remain one of the most important and wanted items in the characterization process of any novel solid phase, either synthesized or found in nature, no matter whether previously predicted or not.
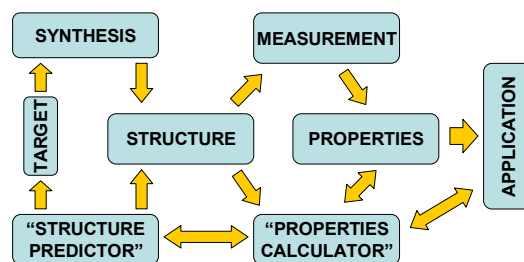

Figure 1: The structure determines the properties and the properties are used by an application. The classic sequence (upper row) of synthesis of many compounds and measurement of their properties to choose one for the application could be faster and cheaper if we had good computation-based structure predictor and properties calculator, as the synthesis could be focused on only a few promising "target" compounds.

Although this scheme may seem too speculative and very far from reality to someone, there are already tools available for both computational "mills" in Fig 1. In the rest of the article the "properties calculator" is left aside and the attention is paid only to the structure predicting part. An illustration of the concepts and

examples of contemporary capabilities of the structure predicting tools is given. There is no attempt to provide an exhaustive review here, just some examples are taken out of many to support understanding the idea of the article and prevent getting lost among various examples and variations of methods.

Two principal ways of structure prediction

From the very beginning of the modern crystallography, the crystallographers were thinking not only on improving methods and instrumentation for crystal structure determination, but also on understanding the principles that guide the highly ordered self-assembly of atoms and molecules, forming crystals. Thus, only about one and a half decade after the first structures were determined, Linus Pauling formulated his well known five rules[2], which were later substantially refined and quantified and even nowadays summarize essential crystallographer's thinking.

Pauling's approach and its derivatives, as well as many that appeared later, are based on the derivation of crystal-chemical rules from the analysis of known structures. This approach will be classified as "data mining" in the following text to underline the fact that conclusions of this approach are based on the existing structural data.

The second way of "understanding" or evaluate a crystal structure in terms of its stability appeared much later – only in the last few decades and is based on the quantum chemical calculations of the periodic atomic arrangements.

Before going into more detail about both principal approaches it is necessary to mention that any self-consistent understanding of crystallographic principles represents not only an aid to overcome difficulties during the structure determination and/or a tool for checking the solved structures for consistency but also a means to predict structures. A short overview of both concepts is given in the following two sections.

Data mining (analysis of known structures)

As mentioned above, the derivation of the five Pauling's rules is a prototype of this approach.[2] It was proved through decades that this five empiric statements (1. coordination polyhedra, 2. bond valences, 3. preference of vertex- over edge- and face-sharing, 4. isolated polyhedra of highly charged and low coordinated cations, 5. small number of different environments of a given ion in a crystal - parsimony) hold rather well for ionic types of structures (mainly oxides and fluorides).

However if one wants to check these rules as a structure prediction tool, giving only an assumed formula, he finds that the rules are just qualitative and rather general – a large number of different structures

closely obeying all the rules can be constructed. Two examples of this fact are given.

The first is DLS-76,[3] which was one of the first computer programs for optimizing crystal structures to achieve the best match with expected bond lengths and angles (average or most common values from known structures). This is possible if the number of restraints (prescribed bonds and angles) exceeds the number of free variables (atomic coordinates). A good example of suitable structures for DLS-76 optimization are zeolite-like materials, consisting of 4-connected tetrahedral frameworks where the Pauling's concept of vertex sharing polyhedra applies. In such cases DLS-76 can produce structures that look very realistic and in cases where the predicted structure corresponds to an existing material, the predicted and actual structures match very closely. It has to be pointed out, however, that this program is more a refinement than a prediction tool. It needs approximate cell dimensions, space group and rough initial atomic coordinates from which the connectivity can be established. So the prediction of the approximate structure for DLS-76 optimization has to come from another source. At the time when this program was written it was not rare that such an input came from manual model building. Nowadays this task can be computerized and the potential structures fed into DLS procedure. The structures, which can not be fitted well, can be ruled out but there remains a large number of structures closely matching the geometric requirements, built into the program.

Second example of the concept of vertex sharing polyhedra is the program GRINSP.[1] It constructs 3D vertex-sharing 3-, 4-, 5- and 6-coordinated frameworks. It also allows combinations of different types of polyhedra. The massive output of this program is very demonstrative – it finds so many feasible predicted structures that they are recorded into an open database (PCOD[4]) with an idea to serve as a resource for structure "identification" of newly synthesized compounds by means of well established qualitative phase analysis using X-ray powder diffraction data. It is an original idea, but it is not the aim of this article to discuss its future development. GRINSP was used here just as a recent and a rather general demonstration that staying with a few general rules (vertex-connected polyhedral frameworks obey Pauling's rules by definition) does not lead to prediction of only a few "really existing" structures. And shows how ambitious such a wish is.

There are various ways to narrow the number of possibilities. Just four of them are mentioned to underline the diversity of possibilities: 1. refining and quantifying the bond-valence rule, 2. calculation of the structure uniformity, 3. exploring the feasibility of different types of tilting in hypothetical perovskites and 4. using larger building blocks instead of individual polyhedra.

The bond-valence approach was gradually improved in terms of quantitative relation between the bond valence and the bond distance[5,6] and successfully used for structure prediction of specific classes of inorganic compounds (for example La_2NiO_4[7]).

The structure uniformity concept, which calculates and prefers most uniform cation and anion arrangement, was successfully used for structure prediction of some new fluorides of the type $M1_nM2_mM3F_6$.[8]

Simulation of different kinds of tilting of the octahedra in perovskites can be used to predict the actual tilting type in the new perovskite structures.[9]

The use of larger building blocks instead of the individual polyhedra also reduces the number of possible structures. Construction of new architectures can be achieved using 2D[10,11] or 3D[12,13] fragments of known structures.

Common feature of all four examples of more refined data mining (known-structures-derived) methods is that they worked on a specific class of compounds with high similarity of the composition and constituent atoms of the known structures, building the "knowledge database", and predicted, not yet synthesized, compounds. Therefore, if we limit ourselves to a specific class of compounds and invest enough effort into detailed analysis and rules-finding (refining) it is already possible to predict a limited number of new structures, which can later be synthesized.

The diversity of the world of inorganic crystal structures is the fact that limits the generalization of this statement to all inorganic structures – the majority of classes of inorganic compounds has not been explored in a sufficient detail yet.

Quantum chemical approach

At least in theory, the calculation of the (electronic) energy of a periodic solid form the "first principles" does not require the database knowledge. From this point of view it can be regarded as more general (not bound to a specific class of compounds). The problem may arise when approximations are applied, for example pseudopotentials or even atomistic force-fields, which are validated by comparison of the computational results to the experimental data for only a small subset of the known structures. However, based on numerous publications of very good results and also personal experience[14] it is believed that the problem of accuracy of the quantum chemical calculations is adequately solved. This statement is based on the applications of the Density Functional Theory (DFT)[15-17] for the calculations of periodic structures. There are many computer programs, which use this approach[18], however, one has to bear in mind, that the "native" use of these programs is not the "structure prediction" but rather "structure refinement". This means that they take a crystal structure and calculate its energy, electronic bands... and can usually also optimize (refine) the structure to slip into the closest local energy minimum.

A natural idea is to combine such a program with one of the global minimum search algorithms (Monte Carlo, genetic algorithm, simulated annealing …) to explore the "whole space" of possible (and impossible) crystal structures at an atomic level. It is clear that without any limitations, the problem is of infinite size – at least the unit cell volume has to be limited. And even then the number of all possible arrangements of atoms, which would have to be calculated and at least partially optimized, can be extremely large.

Thus, the global minimum (or better – the few deepest minima, to see possible polymorphs) search at atomic level using the DFT procedure as the cost function (quantitative criteria for the suitability of a candidate) is not (yet) achievable without simplifying approximations. Namely, according to the experience of an ordinary crystallographer and the computer power to which he has access, it nowadays takes weeks of computation just to optimize and compare a few (i.e. about five) candidate structures. It is clear that doing this on millions of structures is too time consuming even for contemporary supercomputer capabilities and either the number of trial structures has to be reduced significantly or the computation speed has to be increased using approximations (or both).

Reduction of the number of trial structures is achievable in many ways. For example, use of the unit cell and/or possible symmetry obtained by other methods (powder diffraction). However, this is not the structure prediction any more, but structure solution. Another way to reduce the number of possible trial structures is the use of the knowledge from the data mining approach and to take polyhedra or even larger building blocks instead of individual atoms as units to construct the trial structures.[13]

A rather successful example of speeding up the computation while keeping the atomistic "global search" is used in the concept of energy landscapes.[19,20] The energy minima are first identified by a fast calculation, using extensive approximation, and then explored in detail with more accurate (slower) algorithm. Using this approach, kinetically stable phases can be identified as valleys of potential energy at low temperatures, while a more complex minimization of the free energy is required to identify high-temperature phases.

A combination of various methods: reducing the size of the problem using known unit cell, application of genetic algorithm for global search, use of a fast cost function based on bond-valence calculation to identify candidate and a more accurate optimization of the candidates using GULP,[21] can also work.[22]

Structure prediction pathway

Based on the previous sections, a scheme can be constructed to visualize some possible ways leading from a chemical formula of an inorganic compound to a predicted crystal structure (Fig. 2).

The following summary can be made:

1. General crystallographic rules (i.e. Pauling's) lead to prediction of too many structures – as a cost function such rules are not selective enough.

2. More detailed data mining approach (bond-valence, larger building units, uniformity...) can provide more selective predictions that can be verified by subsequent synthesis (i.e. achieve the final goal of structure prediction). Usually these methods are limited to specific classes of compounds.

3. Quantum chemical calculations can refine trial structures with the accuracy comparable to the experiment and quantitatively sorted by the energy (stability), which may be related to the probability that the structure can be synthesized in reality.

4. These methods are, however, too time consuming to be used for global minimization procedures, where very fast evaluation is required to screen large number of candidates and pick up only a few that are possibly existing kinetically stable polymorphs.

5. Two-stage screening is currently applied to overcome the problem. The "rough sieve" (cost function) uses either empirical parameters, derived by data mining or extensive approximation of quantum chemical calculation or both to achieve suitable candidates in a reasonable time. "Fine sieve" of accurate quantum chemical calculation is then applied to evaluate these candidates in detail.
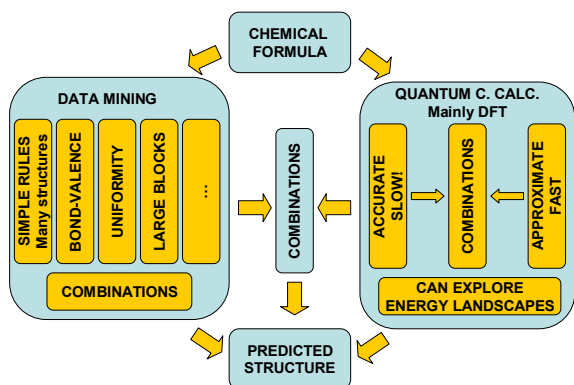

Figure 2: Structure prediction pathway (see text).

Conclusion

Structure prediction of inorganic compounds is not a dream any more. The number of publications reporting successful structure predictions is growing and more and more people contribute to the field with their scientific capabilities. From the experience with similar situations in the past, we can expect an accelerating development of the field in the near future. Obviously any improvement of any method involved is a significant contribution. A good question is, however, which stage of the process is currently the bottleneck, slowing down the expansion of the field. The author's opinion is that this is a cost function for rough screening of the candidates during global search. It has to be general (describing atoms, not being bound to a specific class of compounds) and simple (fast to calculate) but good enough to identify "all and only" suitable candidates for really existing polymorphs.

It is difficult to predict, which way leads to such a cost function. It may be a gradual development or a brilliant breakthrough, it may be based on data mining or quantum chemical approach or a combination thereof, or something different... One idea, which may be worth to explore is a general atom-based data mining approach that would upgrade the sole bond-valence principle. Similar task has recently been applied on organic structures to produce "database-derived" atom-atom potentials[23]

Whatever is going to happen in the field of inorganic crystal structure prediction, it will certainly be exciting science, having a considerable impact on other research fields.

References

1. A. LeBail, *J. Appl. Cryst.* **2005**, *38*, 389-395.

2. L. Pauling, *J. Am. Chem. Soc.* **1929**, *51*, 1010.

3. Ch. Baerlocher, A. Hepp, W. M. Meier, *DLS–76–A Program for the Simulation of Crystal Structures by Geometric refinement*, Institute of Crystallography and Petrography, ETH Zurich, Switzerland, **1977**. & Ch. Baerlocher, A. Hepp, *Z. Kristallogr.* **1976**, *144*, 415–416.

4. A. LeBail, *GRINSP for Windows*, http://www.cristal.org/grinsp/, (accessed: Apr. 10th 2006)

5. I.D. Brown, *J. Appl. Cryst*. **1996**, *29*, 479–480. (and references therein)

6. J. S. Rutherford, *Acta Cryst.* **1998**, *B54*, 204–210.

7. I.D. Brown, *Z. Kristallogra.* **1992**, *199*, 255–272.

8. E.V. Peresypkina, V.A. Blatov, *Acta Cryst.* **2003**, *B59*, 361–377.

9. M. F. Lufaso, P. M. Woodward, *Acta Cryst*. **2001**, *B57*, 725–738.

10. H. Kabbour, L. Cario, F. Boucher, *J. Mater. Chem.*, **2005**, *15*, 3525–3531.

11. H. Kabbour, L. Cario, M. Danot, A. Meershaut, *Inorg. Chem.*, **2006**, *45*, 917–922.

12. C. M. Draznieks, J. M. Newsam, A. M. Gorman, C. M. Freeman, G. Ferey, *Angew. Chem. Int. Ed.* **2000**, *39*, 2270–2275.

13. C. Mellot-Draznieks, S. Girard, G. Ferey, J. C. Schon, Z. Cancarevic, M. Jansen, *Chem. Eur. J.* **2002**, *8*, 4103–4113.

14. A. Meden, A. Kodre, J. P. Gomilšek, I. Arčon, I. Vilfan, D. Vrbanič, A. Mrzel, D. Mihailović, *Nanotechnol.* **2005**, *16*, 1578–1583.

15. P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.

16. W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

17. A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652. (and references therein)

18. *Density Functional Theory*, http://en.wikipedia.org/wiki/Density_functional_theory, (accessed: Apr. 10th 2006)

19. J. C. Schon, M. Jansen, *Z. Kristallogr*. **2001,** *216*, 307–325.

20. J. C. Schon, M. Jansen, *Z. Kristallogr*. **2001**, *216*, 361–383.

21. J. D. Gale, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 629–637.

22. S. M. Woodley, P. D. Battle, J. D. Gale, C. R. A. Catlow, *Phys. Chem. Chem. Phys*. **1999**, *1*, 2535–2542.

23. D.W.M. Hofmann, J. Apostolakis, *J. Mol. Struct.* **2003**, *647*, 17–39.

Povzetek

Različne poti do razumevanja zakonitosti gradnje kristalov so razvrščene v dve skupini. Prva je kristalo-kemijska analiza in izpeljava kristalografskih zakonitosti iz znanih struktur, druga pa kvantno-kemijski izračun stabilnosti skupaj z lokalno ali globalno energijsko minimizacijo. Oba pristopa sta obravnavana s stališča uporabnosti za napovedovanje kristalni struktur anorganskih trdnih snovi. Definiran je pomen uspešne napovedi strukture in poudarjena njena uporabnost. Na koncu je podan pričakovan nadaljnji razvoj.