ORIGINAL PAPER

Bohdan Koudelka · Pavla Capkova

An algorithm to filter out packing arrangements based on steric clashes

Abstract This document outlines the use of an algorithm to filter out impossible crystal-packing arrangements based on steric considerations. Within an exhaustive grid search frame, the space sample is reduced by analysis of spherical areas where atom pairs from different rigid units might clash.

This technique finds areas in the state space where the global energy minimum might lie. The minimum can then be found by the usual methods of molecular modeling restricted to these particular areas.

Only a tiny fraction of atom pair distances need to be tested; usually a single quantity on average per one state of model space! For example, a crystal of three rigid molecules, each containing 12 atoms, has $3 \times 12 \times 12 = 432$ atom pairs just in one unit cell but our method needs to test on average 1 to 4 atom pairs per state.

Using modern computers, about 10^{12-15} models can be tested within several hours or days. For example, a crystal model with six rotational degrees of freedom (two rigid molecules in the unit cell) each with step 3° can be tested in a few hours on a 1-GHz x86 processor-based machine. The method presented here has been implemented in the SUPRAMOL program.

Keywords Molecular modeling · Crystal structure prediction · Energy minimization · Crystal packing

Introduction

The goal of molecular modeling is to predict the particular arrangement of molecules in the crystal unit cell, which it is hoped will to be realized in nature. This is achieved by finding the global energy minimum. Stochastic methods (e.g. Monte Carlo, random kick method, genetic algorithms), molecular dynamics or deterministic grid searches can be used to achieve this requirement [1]. Despite the widely used stochastic approaches, a grid search can examine the whole space of possible crystal arrangements and thus one can be sure to find the global energy minimum. On the other hand, grid search methods are rather time consuming and can be applied only on small crystal structures—up to one hundred atoms per crystal unit cell. Although nowadays it is trivial to solve a single crystal X-ray structure containing one hundred atoms per unit cell, unequivocally predicting such a structure is currently still hopeless [2, 3].

There are a number of different computer programs developed to be used in crystal structure prediction. Most of them use an empirical energy as their function to classify the fit; e.g. CRYSCA [4], DMAREL [5], MOLPAK [6], MSI-PP [7, 8], Zip-Promet [9] along with those that use grid search methods MPA [10, 11], UPACK [12, 13]. Statistical potentials are used in FlexCryst [14] and the statistical fit in PackStar or Rancel [15]. The reader can find more complex descriptions and comparisons of these programs in [2, 3].

Thus having some knowledge about crystal content, one can now try to find the best arrangement of molecules. Atoms in the crystal unit cell are spread into rigid units (molecules or parts of molecules that can be taken as rigid for the first stage), and an attempt is made to find the position and orientation of each part, whereby the total crystal energy reaches the global minimum. This documented method is in fact only a small fraction of the whole process; an extremely fast primary filter that can be implemented to variate grid searches.

To find the global minimum, the whole space of possible crystal states should be searched. Thus, during the first stage this space is defined by setting appropriate ranges of internal coordinates. Secondly, the grid of testing models is spread over it. This method finds areas where the particular rigid units do not overlap, and should be followed up by more sophisticated searches within

B. Koudelka (✉) · P. Capkova
Department of Chemical Physics and Optics,
Faculty of Mathematics and Physics,
Charles University Prague,
Ke Karlovu 3, 121 16 2 Prague, Czech Republic
e-mail: koudelka@mbox.troja.mff.cuni.cz
Tel.: +4221911245

these areas, which are only a small fraction of the whole state space.

Despite the widely used methods relying on interaction lists used in molecular dynamics, as in the Charmm computer program [16] we use a different approach. The main idea allowing the use of interaction lists in molecular dynamics, "The state of crystal changes slowly during the time", is not satisfied here because the grid should cover the whole state space of the crystal. Even though this idea can also be adopted somehow, we can instead gain even more by taking into account the specific properties of the grid itself.

Basic definitions

A crystal unit cell (any space group) can be defined as containing *N rigid units*—molecules or parts of molecules that can be taken as rigid bodies.

Obviously, in the energy minimum the particular rigid units should not overlap! Thus a *minimal acceptable distance* d_0 is set for each relevant atomic pair. This can be the addition of atomic radii defined in some way, or a distance given specifically with respect to interacting molecules. The important fact is that for each atom pair that is combined of atoms from different rigid units we have defined a minimal acceptable distance.

For each rigid unit $i \in \{1, ..., N\}$ a set of *testing positions P_i* is defined. This set should "uniformly" cover (as much as possible) the whole positional space of the *i*th rigid unit within the unit cell of the crystal, and with respect to positions of other parts (if bonded or related by symmetry). This space covering can be characterized by the maximal distance $\Delta d_i^{\max}$ of an arbitrary space point from the nearest testing position. In the same way, a set of *testing orientations O_i* is defined, characterized by the maximal angle difference $\Delta \alpha_i^{\max}$ of any point of the orientation space, from the nearest testing orientation. For more details and definitions of different "best coverings" for spaces and coverings of the sphere see [17, 18].

The "change of orientation" refers to a rotation of a rigid unit around a specified point (e.g. the center of mass or an atom—i.e. bonded to another rigid unit). This point can be chosen freely with respect to the internal coordinates of the rigid unit, so it translates when the position of the rigid unit changes. This point can be called the *rotation center C_I*. The orientation is then given by three coordinates; e.g. Euler angles or spherical coordinates ϑ, ϕ of some other atom of rigid units with respect to the rotation center and the rotation angle ω around the axis given by this atom and the rotation center. The most important thing that has been defined is the rotation center of each rigid unit, and that each atom A^k of the *i*th rigid unit ($k \in \{1...n_i\}$) has a fixed distance from the rotation center that is denoted r_i^k.

Once the sets of testing points of all rigid units have been defined, next comes the total number of testing models

$$n = \prod_{i=1}^{N} n(P_i)n(O_i) \tag{1}$$

and the top limit of maximal displacement of any atom when approximating the arbitrary state of a crystal by testing point

$$\Delta d^{\max}(A_i^k) = \Delta d_i^{\max} + \Delta \alpha_i^{\max} r_i^k \tag{2}$$

Globally acceptable space states

There can be globally three types of rigid units in the model:

– *Static*—the space $P_i \times O_i$ contains only one state (for example the rigid layer in a crystal concerning intercalation, the rigid skeleton of zeolite structure, etc.), or the rigid unit made static during the computation (see. below).
– *Dependent*—the states of $P_i \times O_i$ depend on the positional and orientational states of other rigid units (not by crystal symmetry which is already set when the state space is defined). The most common example is introduced by the chemical bond of a fixed length and the valence angles between two rigid units—only the torsion angle can change. In such a case one part of the molecule is shifted and rotated independently and the other depends on it.
– *Independent*—the states of $P_i \times O_i$ do not depend on states of other rigid units.

While filtering out the overlapping states of the crystal the following steps should be repeated:

1. Test the interaction of independent rigid units at first with their own symmetry equivalents within the crystal and then with all static units. This way, all the unacceptable states are filtered out right from the space $P_j \times O_i$. This saves time wasted in further testing.
2. Test all the mutual interactions of independent rigid units (This is the most time-consuming part of the algorithm and its optimization determines the performance of the whole procedure)—the *key point of this method*, described in the next section. The implementation in Supramol [19] tests about 500,000–700,000 models per second on a 1.6-GHz x86 processor machine (or 120,000–150,000 on a RISC 250-MHz processor).
3. Concerning one particular acceptable state of independent rigid units, these units become static and some other previously dependent rigid units become independent. Thus, the procedure can be repeated from point 1 until all the rigid units are exhausted.

Finally comes the area of all acceptable arrangements of the crystal unit cell with respect to given minimal distances of all atom pairs.

The key idea of the documented method

Up to this point nothing extra was said. The following paragraph describes the key idea of the whole method—the way to test the mutual interaction of two independent rigid units efficiently.

Let us take the ith and jth rigid units and determine the area of acceptable mutual arrangement (the acceptable points of space $O_i \times O_j$) while positions of both rigid units are fixed. Instead of a full test of $n(O_i) \times n(O_j)$ orientation states, helping data is calculated for each state of both rigid units. It is only $n(O_i) + n(O_j)$ states, which is a much smaller number; e.g. for $n(O_i) = n(O_j) = 2,000$ it is 1,000 times smaller! This helps to eliminate the number of atom pairs tested for distance rapidly.

Now concerning the major idea of this method. When the position of the rigid unit is fixed, all its atoms can move only on the spheres with center at the point C_i and radius r_i^k. When considering an arbitrary atom pair with its first atom A^k from the i-th rigid unit and its second atom A^l from the j-th rigid unit, a *restricted area* F_i^k can be defined for the first atom or *restricted area* F_j^l for the second atom wherever

$$\left| r_i^k - d_{ij} \right| < r_j^l + d_0^{kl} \quad \text{or} \quad \left| r_j^l - d_{ij} \right| < r_i^k + d_0^{kl} \qquad (3)$$

holds. From Fig. 1 it can be seen that the atoms can be closer than their minimal distance d_0^{kl} only when both are in their restricted areas. When one of them is out of the restricted area we do not need to test this atom pair for distance. Therefore we associate a *set of atom pairs*, that can cause a need for the distance test, with each state from O_i or O_j respectively.

When these sets are pre-computed for each single state from O_i and O_j, one can simply go through the space $O_i \times O_j$ and test the distance of only these atom pairs that are in the *intersection* of the two appropriate atom pair sets—one corresponding to the particular state of the i-th rigid unit and the second corresponding to the particular state of the jth rigid unit.

Important implementation details

Ordering of atom pairs

The calculation of the interaction of two rigid units may be speeded up by the proper ordering of the tested atom pairs.

The best representation of the above defined sets of atom pairs is a binary array (the intersection of such sets being simply the binary AND). Each mutual atom pair of both interacting rigid units either is or is not there. This
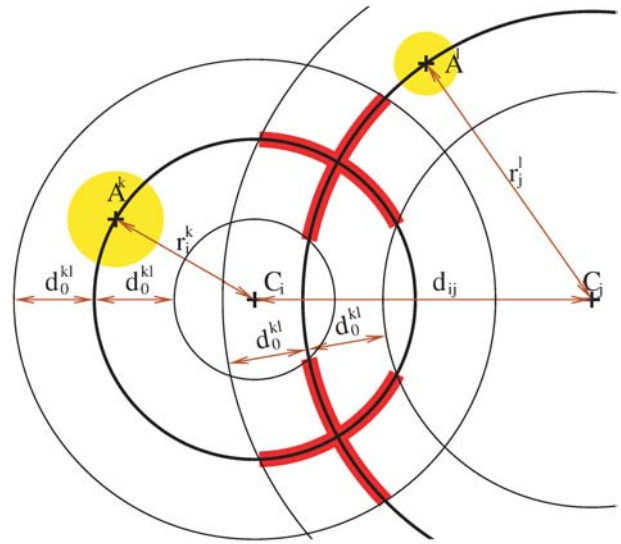

Fig. 1 Restricted area (*thick red lines*) of an atom A^k that belongs to i-th rigid unit and the restricted area of an atom A^l that belongs to j-th rigid unit, while their interaction is considered with the minimal acceptable distance d_0^{kl}. C_i and C_j are rotation centers of both rigid units. When the i-th (or alternatively j-th) rigid unit changes orientation the atom A^k (or A^l) resides on a sphere of radius r_i^k (r_j^l). d_{ij} is the distance between rigid unit centers. If an atom is in its restricted area this does not mean that both atoms are closer than d_0^{kl}—the second atom can be in any other position. Neither do both atoms need to be closer than d_0^{kl}, if they are in their restricted areas. On the other hand, the atoms can be closer than their minimal acceptable distance only when they are both in their restricted areas

way implicit ordering of all the inter-rigid unit atom pairs is defined; the ordering in which they will be tested for distance.

The best way is to order the atom pair with respect to a *probability of collision*. The probability of collision of k-th atom of i-th rigid unit with l-th atom of j-th rigid unit when positions of atoms on spheres (the spatial angle Ω) are random is

$$P_{kl} = \frac{1}{4\pi} \int d\Omega_k \times \frac{1}{4\pi} \int_\pi d\Omega_l \times P_{kl}(\Omega_k; \Omega_l) \qquad (4)$$

where $P_{kl}(\Omega_k; \Omega_l)$ is 1 in collision and 0 elsewhere. The probability (4) can be evaluated by integration over distance x of the l-th atom and the rotation center C_I by

$$P_{kl} = \int_{\left| r_j^l - d_{ij} \right|}^{r_j^l - d_{ij}} \rho(x) dx$$

where

$$\rho(x) = \begin{cases} 0 & |x - r_i^k| > d_0^{kl}\rangle; \quad \text{No collision} \\ \dfrac{\left(d_0^{kl}\right)^2 - (x - d_{ij})^2}{2 r_i^k r_j^l d_{ij}}, & x \in \langle |r_i^k - d_0^{kl}|, r_i^k + d_0^{kl}\rangle; \quad \text{Partial collision} \\ \dfrac{2x}{r_j^l d_{ij}}, & x < d_0^{kl} - r_i^k; \quad \text{Fullcollision} \end{cases}$$

Thus we have the analytical formula

$$P_{kl} = \left[\frac{3x\left(d_0^{kl}\right)^2 - \left(x - d_{ij}\right)^3}{6r_i^k r_j^l d_{ij}}\right]_{\max\left(\left|r_j^l - d_{ij}\right|, \left|r_i^k - d_0^{kl}\right|\right)}^{\min\left(r_j^l + d_{ij}, r_i^k + d_0^{kl}\right)} \tag{5}$$

plus

$$\left[\frac{x^2}{r_j^l d_{ij}}\right]_{d_0^{kl} - r_i^k}^{\min\left(r_j^l + d_i, d_0^{kl} - r_i^k\right)} \quad \text{when} \quad d_0^{kl} - r_i^k > \left|r_j^l - d_{ij}\right|$$

Using this ordering, *the average number of atom pairs tested for distance is on the order of units per one testing state of $O_i \times O_j$.* When having a large number of states (in case d_0 depends on the covering of the state space; see below) of each rigid unit it often becomes very close to 1! This number does not depend much on the number of atoms in the rigid units—it is instead dependent on d_{ij} and how closely packed the crystal is. Fortunately, the large majority of crystals are rather closely packed! Even though the number of atoms in a rigid unit can commonly reach 10, the number of mutual atom pairs (when the crystal symmetry copies are also considered), which is of the order of hundreds or thousands, this method would seem to be amazingly fast!

The choice of a suitable d_0

Let us take the basic d_0^{kl} for all atom pairs given by the chemical properties of given molecules. These parameters can be modified due to the geometry of the model in order to reach some properties for the resulting area of the acceptable arrangement of the crystal unit cell. We have three basic possibilities.

1. No modification—the result ensures that all the tested points are acceptable, but does not say anything about their neighborhood in the space of arrangements.
2. d_0^{kl} is increased by the maximum possible displacement of atoms (2). This way ensures that the global energy minimum will not be skipped, but on the other hand the run will involve many states that will not become acceptable nor within the neighborhood of the testing state and must be discarded later.
3. Decrease d_0^{kl} by the maximum possible displacement (2)—this way an area is found that is surely acceptable as a compact region, but the global minimum can be omitted as well many other acceptable states with the unacceptable neighborhood.

Conclusion

While performing a deterministic grid search to find the global energy minimum of the crystal packing, the method documented here can be used successfully for the first stage—as the primary filter. Although rough, it is very fast (it reduces the number of atom pairs to be tested for distance on an average of 10^2–10^3 times) and efficiently chooses only the states at which the atoms do not overlap, or may not overlap in some place within one mesh of the grid. This method very rapidly decreases the number of states to be further tested by consequent sophisticated but slower methods (especially for closely packed crystals); e.g. the method using the crystal energy computation. In contrast to the widely used stochastic method, the parameters can be chosen in such a way that the global minimum can definitely be found! However, it is likely to take more time. Today's computers limit exploration up to 8- or 12-dimensional spaces of internal parameters.

The algorithm described is implemented in the program SUPRAMOL [19].

References

1. Comba P, Hambley TW (1995) Molecular modelling of inorganic compounds. VCH, Weinheim, pp 1–58
2. Lommerse JPM, Motherwell WDS, Ammon HL, Dunitz JD, Gavezzotti A, Hofmann DWM, Leusen FJJ, Mooij WTM, Price SL, Schweizer B, Schmidt MU, van Eijck BP, Verwer P, Williams DE (2000) Acta Crystallogr B 56:697–714
3. Motherwell WDS, Ammon HL, Dunitz JD, Dzyabchenko A, Erk P, Gavezzotti A, Hofmann DWM, Leusen FJJ, Lommerse JPM, Mooij WTM, Price SL, Scheraga H, Schweizer B, Schmidt MU, van Eijck BP, Verwer P, Williams DE (2002) Acta Crystallogr B 58:647–661
4. Schmidt MU, Englert U (1996) J Comput Chem 17:1–18
5. Willock DJ, Price SL, Leslie M, Catlow CRA (1995) J Comput Chem 16:628–647
6. Holden JR, Du ZY, Ammon HL (1993) J Comput Chem 14:422–437
7. Karfunkel HR, Leusen FJJ, Gdanitz RJ (1993) J Comput Aid Mater Des 1:177–185
8. Verwer P, Leusen FJJ (1998) Rew Comput Chem. Lipkowitz KB, Boyd DB (eds) Wiley-VCH, New York, pp 327–365
9. Gavezzotti A (1999) Zip-Promet with Minopec. University of Milan, Italy
10. Williams DE (1999) J Mol Struct 321–347
11. Williams DE (1999) J Mol Struct 485–486
12. van Eijck BP, Kroon J (1999) J Comput Chem 20:799–812
13. Mooij WTM, van Eijck BP, Kroon J (1999) J Phys Chem 103:9872–9882
14. Hofmann DWM, Lengauer T (1997) Acta Crystallogr A 53:225–235
15. Motherwell WDS (1999) Nova Acta Leopoldina 79:89–98
16. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
17. Hardin RH, Sloane NJA, Smith WD (1999) Spherical coverings,http://www.research.att.com/~njas/coverings/
18. Hardin RH, Sloane NJA, Smith WD (2000) Tables of spherical codes with icosahedral symmetry,http://www.research.att.com/~njas/icosahedral.codes/
19. Koudelka B, Čapková P (2002) J Mol Model 8:184–190