Journal of
Applied Crystallography
ISSN 0021-8898

Received 16 December 2004
Accepted 21 January 2005

Inorganic structure prediction with *GRINSP*

Armel Le Bail

Université du Maine, Laboratoire des oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France. Correspondence e-mail: alb@cristal.org

A new computer program is described, *GRINSP* (geometrically restrained inorganic structure prediction), which allows the exploration of the possibilities of occurrence of 3-, 4-, 5- and 6-connected three-dimensional networks. Hypothetical (as well as known structure) models for binary compounds are produced with exclusive connection of polyhedra by corners, such as $[MX_3]$ triangles in M_2X_3 formulation, $[MX_4]$ tetrahedra in MX_2 (zeolites or dense SiO_2 polymorphs), $[MX_5]$ polyhedra in M_2X_5, and finally $[MX_6]$ octahedra in MX_3 polymorphs. Moreover, hypothetical ternary compounds are built up by combinations of either two different polyhedra or two different radii for two different cations adopting the same coordination. The cost function is based on the agreement of the model interatomic distances with ideal distances provided by the user. The Monte Carlo algorithm first finds structure candidates selected after the verification of the expected geometry, and then optimizes the cell parameters and the atomic coordinates. A satellite software (*GRINS*) uses the predicted models and produces the characteristics of isostructural compounds which would be obtained by cationic substitutions. A huge list of CIF files of hypothetical boron oxide polymorphs (including nanotubes), zeolites, aluminium and 3*d*-element fluorides, fluoroaluminates, borosilicates, titanosilicates, gallophosphates *etc.*, is freely available at the PCOD (Predicted Crystallography Open Database).

1. Introduction

The final aim of structure prediction should be to announce a crystal structure before any confirmation by chemical synthesis or discovery in nature. In a lead article entitled *Stuctural Aspects of Oxide and Oxysalt Crystals*, Frank C. Hawthorne (1994) stated, ten years ago, that: ‘The goals of theoretical crystallography may be summarized as follow: (1) predict the stoichiometry of the stable compounds; (2) predict the bond topology (*i.e.* the approximate atomic arrangement) of the stable compounds; (3) given the bond topology, calculate accurate bond lengths and angles (*i.e.* accurate atomic coordinates and cell dimensions); (4) given accurate atomic coordinates, calculate accurate static and dynamic properties of a crystal. For oxides and oxysalts, we are now quite successful at (3) and (4), but fail miserably at (1) and (2)’. This seems in contradiction with a previous statement by Catlow & Price (1990), four years earlier, that ‘computational methods can now make detailed and accurate predictions of the structures of inorganic materials’. The fact is that predictions of inorganic compounds mentioned in a recent book about computer modelling in inorganic crystallography (Catlow, 1997) are very few, if one excludes hypothetical zeolites. Moreover, in the case of organic molecules, the predictions do not appear to be any more brilliant, based on the results of a recent blind test (Motherwell *et al.*, 2002). If the state of the art had dramatically evolved in the past ten years, we should have a huge database of predicted compounds, and no new crystal structure would surprise us since it would correspond to an entry in that database. Moreover, we would have obtained in advance the physical properties and we would have preferably synthesized those interesting compounds. Of course, this is absolutely not the case, unfortunately. However, two databases of hypothetical compounds were built in 2004. One is exclusively devoted to zeolites (Foster & Treacy, 2004); the other includes zeolites as well as other predicted oxides (borosilicates, titanosilicates, gallophosphates *etc.*) and fluorides (Le Bail, 2004). Such databases will play a role analogous to databases of actually existing structures: in principle they preclude the prediction of a structure that has already been predicted, or the redetermination/republishing of a known structure. Moreover, calculated powder patterns from these databases would be useful at the identification stage, provided that the accuracy level of prediction is high (observed and predicted cell-parameter differences smaller than 2%).

Let us cite a few of the computer programs and methods producing predictions in the inorganic world. *CASTEP* uses the density functional theory (DFT) for *ab initio* modelling, applying a pseudopotential plane-wave code (Payne *et al.*, 1992). The structures gathered in the database of hypothetical zeolites (Foster & Treacy, 2004) are produced from a 64-processor computer cluster, grinding away non-stop, generating graphs and annealing them, the selected frameworks being then re-optimized using the *General Utility Lattice Program*, *GULP* (Gale, 1997), using atomic potentials. *GULP* itself is able to predict crystal structures (TiO_2 polymorphs). Recently, a genetic algorithm was implemented (Woodley, 2004) in *GULP* in order to generate crystal framework structures from the knowledge of only the unit-cell dimensions and constituent atoms (however, according to the definitions above, this is structure determination, not prediction); the structures of the better candidates produced are relaxed by minimizing the lattice energy, which is based on the Born model of a solid. The concept of ‘energy landscape’ of chemical systems is used

doi:10.1107/S0021889805002384

by Schön & Jansen (2001*a*,*b*) for structure prediction with their computer program *G42*. Another package, *SPuDS*, is dedicated especially to the prediction of perovskites (Lufaso & Woodward, 2001). The AASBU method (automated assembly of secondary building units) is developed by Mellot-Draznieks *et al.* (2000, 2002), using *Cerius2* (2000) and *GULP* in a sequence of simulated-annealing plus minimization steps for the aggregation of large structural motifs. This list of software is rather small considering the fact that structure and properties prediction is obviously an unavoidable part of our future in crystallography and chemistry.

Possibilities for structure prediction which would be easily available freely to academic users appear to be limited somewhat. Apart from the broadly explored zeolite subject, one cannot find many atomic coordinates of hypothetical compounds in databases. Moreover, it seems better to gather hypothetical compounds in a specific database, different from those of determined crystal structures, because predictions will be much more numerous than confirmations. This, combined with the fact that we ought no longer to 'fail miserably' at predicting the stoichiometry and the approximate atomic arrangement of stable compounds (see above), prompted the development of new software, *GRINSP* (geometrically restrained inorganic structure prediction). This computer program is described below, enabling the exploration of hypothetical 3-, 4-, 5- and 6-connected three-dimensional networks, in binary and ternary inorganic compounds, using a Monte Carlo approach.

2. *GRINSP* algorithm

2.1. Monte Carlo generation of structure candidates

With *GRINSP*, the occurrence of $M_uM'_vX_w$ or M_vX_w models depends on a drastic selection when trying to build the net of M/M' atoms. First, a space group and the M/X or $M/M'/X$ corner-sharing system to be explored are chosen; then a single initial M or M' atom (selected at random) is placed at random coordinates (at one Wyckoff position, itself selected at random) in a box, the dimensions of which are again selected at random. The next M or M' atoms are placed randomly in delimited volumes close to the M or M' atoms already positioned (these volumes are restricted by the range of provided interatomic distances). Generally, 300 000 Monte Carlo tests for placing atoms are realised before a new series of tests is started with different cell parameters. At this stage, in order to be retained, an M/M' model should exactly correspond to the geometrical specifications (with exact coordinations, though the distances can vary: for instance, if M is decided to be in sixfold coordination, one has to find six M or M' atoms around it at the end of the process). The fact that distances are given a large tolerance range allows the capture of many solutions which may not correspond to regular polyhedra. In other words, the Monte Carlo random walker may stay far above the deep local minimum of interest. In this first step, atoms do not move: their possible positions are only tested and checked; then they are retained or discarded. If the process fails before the end of the allowed series of tests (the number of tests for positioning a new atom inside of the defined restricted volumes is limited by the use of 'insistence factors'), a new initial M or M' atom is placed without changing the cell *etc*. The cell is progressively filled up to respect the geometrical restraints completely, if possible. The number of M/M' atoms placed is not predetermined. The process is thus different from the AASBU approach, or from the simulated-annealing approach used in pioneering studies on zeolites (Deem & Newsam, 1989, 1992; Newsam *et al.*, 1992), since *GRINSP* explores a large range of cell parameters for a given space group instead of concentrating on known cell parameters with a given number of M atoms moving up to find some energy cost function minimum. It is, however, obvious that *GRINSP* can also be used as a structure solution tool for corner-sharing systems of polyhedra, including zeolites, if the cell parameters are known (but such structure solution takes us beyond the realms of structure prediction).

2.2. Model optimization

In a second step, the X atoms are added between the (M/M')–(M/M') first neighbours, at the midpoints, and it is verified by distance and cell improvements (using a Monte Carlo approach as well) that regular $[(M/M')X_n]$ polyhedra can really be built, *i.e.* that there is a deep local minimum existing close to this previously selected rough arrangement of (M/M') atoms. The cost function enabling the finding of a minimum R is based on the verification of ideal (M/M')–(M/M'), (M/M')–X and X–X first-neighbour distances, provided by the user. The total R factor is defined by

$$R = [(R_1 + R_2 + R_3)/(R_{01} + R_{02} + R_{03})]^{1/2},$$

where R_n and R_{0n} for n = 1, 2, 3 are defined by the expressions

$$R_n = \textstyle\sum[W_n(d_{0n} - d_n)]^2$$

and

$$R_{0n} = \textstyle\sum[W_n d_{0n}]^2,$$

where the d_{0n} values for n = 1 to 3 are the ideal first interatomic distances (M/M')–X (n = 1), X–X (n = 2) and (M/M')–(M/M') (n = 3), whereas the d_n values are the corresponding observed distances in the structure model for these atom pairs. The weights retained (w_n) are the same as those used in the *DLS* software (Baerlocher *et al.*, 1978) for the calculation of idealized framework data (w_1 = 2.0, w_2 = 0.61 and w_3 = 0.23). The ideal distances are to be provided by the user for pairs of atoms supposed to form polyhedra (for instance in the case of [SiO_4] tetrahedra, one expects to have d_1 = 1.61 Å, d_2 = 2.629 Å and d_3 = 3.07 Å). The similarity of the cell parameters estimated by *GRINSP* for zeolites with the idealized cell constant listed at the official zeolite Web site (Database of Zeolite Structures, http://www.iza-structure.org/databases/) is thus not fortuitous, since these idealized values are calculated by using the *DLS* software applying a similar cost function during the distance least-squares refinements. Some differences may come from the space-group constraint (always *P*1 with *GRINSP*).

The strategy for the model optimization is first to allow 1/4 of the Monte Carlo events (NA × 10–20 000 events, generally, where NA is the total number of atoms in the cell) only to move randomly the $M/M'/X$ atoms; then another 1/4 are exclusively devoted to random cell-parameter changes, and finally the remaining Monte Carlo events are used for both kind of changes, chosen randomly. A smooth quenching is imposed: the maximum amplitudes of the changes are progressively reduced during the optimization process.

For ternary compounds, the M–M' ideal distances are calculated by *GRINSP* as being the average of the M–M and M'–M' distances. During this second step of optimization, all the atoms can move, but no jump is allowed because a jump would break the coordinations established at the first step. The change in the cell parameters from the structure candidate to the final model may be quite considerable (up to 30%); this explains why some models may show parameters that are larger or smaller than the limits defined at the beginning of the runs, these limits being applied only to the results of the first step (when placing the M/M' atoms). During the optimization, the original space group selected for placing the M atoms may not be conserved

Table 1
Comparison of predicted cell parameters with observed or idealized ones for a few selected zeolites and dense SiO_2 phases.

	Predicted				PCOD entry	Observed or idealized		
	a (Å)	*b* (Å)	*c* (Å)	*R*		*a* (Å)	*b* (Å)	*c* (Å)
Dense SiO_2								
Quartz	4.958		5.364	0.0006	1000001	4.912		5.404
Cristobalite	5.010		6.855	0.0010	1000003	4.969		6.926
Tridymite	5.048		8.382	0.0043	1000002	5.052		8.270
Keatite	7.525		9.066	0.0046	1000037	7.456		8.604
Zeolites								
ABW	9.878	5.129	8.547	0.0034	1000011	9.9	5.3	8.8
ACO	9.890			0.0048	1000009	9.9		
AFI	13.788		8.514	0.0045	1000025	13.8		8.6
AFY	12.322		8.599	0.0074	1000046	12.3		8.6
AHT	15.722	9.372	8.430	0.0088	1000041	15.8	9.2	8.6
ANA	13.555			0.0025	1000012	13.6	13.6	13.6
APD	8.131	17.581	10.566	0.0080	1000044	8.7	20.1	10.2
AST	13.601			0.0059	1000013	13.6		
ASV	8.641		13.709	0.0052	1000034	8.7		13.9
ATT	9.588	7.499	9.538	0.0041	1000040	10.0	7.5	9.4
ATV	8.394	15.349	9.441	0.0056	1000042	8.6	15.3	9.7
AWW	13.654		7.671	0.0033	1000033	13.6		7.6
BIK	7.513	15.830	5.129	0.0049	1000008	7.5	16.2	5.3
CAN	12.459		5.221	0.0057	1000020	12.5		5.3
CAS	4.995	13.890	16.434	0.0063	1000045	5.3	14.1	17.2
CHA	13.293		15.376	0.0054	1000047	13.7		14.8
EAB	13.154		15.028	0.0036	1000023	13.2		15.0
EDI	6.921		6.410	0.0044	1000000	6.926		6.410
ERI	13.022		15.298	0.0059	1000027	13.1		15.2
GIS	9.778		10.165	0.0027	1000028	9.8		10.2
GME	13.625		9.916	0.0028	1000022	13.7		9.9
JBW	5.139	7.950	7.484	0.0035	1000004	5.3	8.2	7.5
LOS	12.504		10.333	0.0052	1000021	12.6		10.3
LOV	7.165		20.819	0.0059	1000036	7.2		20.9
LTA	11.907			0.0033	1000016	11.9		
MEP	13.683			0.0077	1000018	13.7		
MER	13.996		10.017	0.0027	1000031	14.0		10.0
MON	7.124		17.780	0.0051	1000030	7.1		17.8
NAT	13.827		6.424	0.0050	1000029	13.9		6.4
OFF	12.943		7.718	0.0048	1000024	13.1		7.6
OSI	18.363		5.136	0.0045	1000035	18.5		5.3
OSO	10.148		7.624	0.0123	1000026	10.1		7.6
PHI	9.993	13.897	13.877	0.0034	1000043	9.9	14.1	14.0
RHO	14.918			0.0023	1000019	14.9		
SAS	14.031		10.364	0.0039	1000032	14.3		10.4
SOD	8.881			0.0045	1000010	9.0		
THO	13.837	6.923	6.409	0.0045	1000039	14.0	7.0	6.5
WEI	11.786	10.303	9.966	0.0068	1000038	11.8	10.3	10.0

after having added the X atoms, so that the final structure is always proposed in the $P1$ space group. The final cell characteristics and atomic coordinates are presented in a CIF file. An ultimate check of the real symmetry has to be performed by using a program like *PLATON* (Spek, 2003).

The models produced by *GRINSP* may need further optimization by using bond valence rules, or energy calculations; however, in many cases the predicted cell parameters differ by less than 2% from the real ones, when the real compounds are built up from ideal polyhedra, which is the case with dense SiO_2 polymorphs or zeolites (Table 1) and fluoroaluminate phases (Table 2). Choosing to use one precise ideal M–M first-neighbour distance, depending on the M–X–M angles (even if coming from an average value produced by data mining), will produce the smaller R values for particular models. In Table 1, the quartz structure is clearly favoured (R = 0.0006). In Table 2, the smaller R value corresponds to the HTB model, not to the perovskite one. Modifying the Al–Al distance in order to have an Al–F–Al angle of 180° would of course have favoured a small R value for the perovskite structure, but without obtaining cell parameters closer to the observed ones in the $R\bar{3}c$ space group (a cubic space group would have been obtained instead). This shows that no confidence can be given to a precise classification by R values in a range of, say $0 < R < 0.01$. Moreover, values $0.01 < R < 0.02$ may well correspond to existing compounds (R = 0.0159 for τ-AlF_3 in Table 2, offering a large distribution of Al–F–Al angles).

2.3. The *GRINS* satellite program

Searching for the characteristics of isostructural hypothetical compounds obtained by cation substitution (FeF_3 or GaF_3 *etc.*, instead of AlF_3 for instance), it is not necessary to run again the structure prediction software *GRINSP*. A satellite program named *GRINS* was developed, including a modified version of the structure optimization part (Monte Carlo adjustment of the atomic coordinates and cell parameters). This software uses the desired starting M/M' positions and cell parameters, and finds the minimum R factor corresponding to any new set of ideal interatomic distances for new cation/anion pairs selected by the user.

3. Results

3.1. Binary compounds

Formulations M_2X_3, MX_2 and MX_3 were partly examined (not yet M_2X_5 which would occur for M cations in fivefold coordination, since $[MX_5]$ polyhedra cannot be regular).

Table 2
Comparison of predicted cell parameters with observed ones for 6-connected three-dimensional aluminium fluorides.

Observed parameters are given below each row of predicted parameters. FD = framework density (number of M = Al/Ca/Na atoms reported to a volume of 1000 Å^3). SG = space group of the real structure. Z = number of (Al/Na/Ca)F$_3$ formula per cell. N = number of Al/Na/Ca atoms with different coordination sequences. R = quality factor regarding the ideal (Al/Ca/Na)–F, F–F and (Al/Ca/Na)–(Al/Ca/Na) first-neighbour interatomic distances. perov = perovskite; HTB = hexagonal tungsten bronze; pyr = pyrochlore; TTB = tetragonal tungsten bronze.

	Predicted/observed									
	a (Å)	b (Å)	c (Å)	β (°)	R	SG	FD	Z	N	PCOD entry (reference)
α-AlF$_3$ (perov)	5.111		12.504		0.0062		21.21	6	1	1000048
	4.931		12.446			$R\bar{3}c$				(Daniel *et al.*, 1990)
β-AlF$_3$ (HTB)	6.984	12.107	7.213		0.0035		19.67	12	1	1000049
	6.931	12.002	7.134			*Cmcm*				(Le Bail *et al.*, 1988)
η-AlF$_3$ (pyr)	9.667				0.0046		17.71	16	1	1000017
	9.749					$Fd\bar{3}m$				(Fourquet *et al.*, 1988)
κ-AlF$_3$ (TTB)	11.539		3.615		0.0098		20.78	10	2	1000050
	11.403		3.544			$P4/mbm$				(Herron *et al.*, 1995)
τ-AlF$_3$	10.210		7.241		0.0159		21.17	16	3	1000014
	10.184		7.174			$P4/nmm$				(Le Bail *et al.*, 1992)
Na$_4$[Ca$_4$Al$_7$F$_{33}$]	10.876				0.0122		23.27	22	3	1000015
	10.781					$Im\bar{3}m$				(Hemon & Courbion, 1990)
Rb$_2$[NaAl$_6$F$_{21}$]	12.103	6.986	10.651	111.52	0.0088		16.71	14	2	1000051
	12.075	6.972	10.214	113.2		*C*2				(Le Bail *et al.*, 1989)

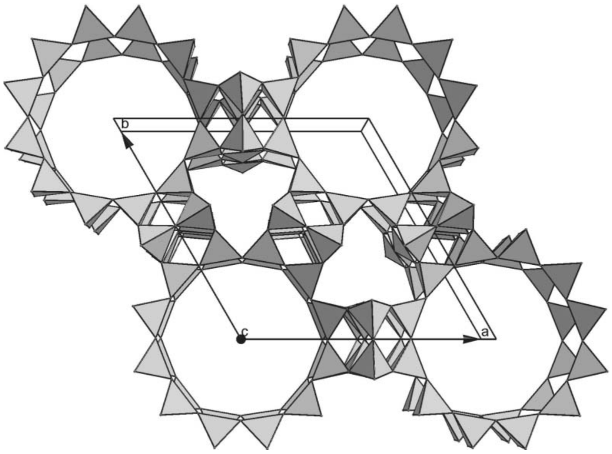

Figure 1
Hypothetical zeolite: space group $P6/mmm$, a = 15.60, c = 7.13 Å, R = 0.0085, PCOD1030081.

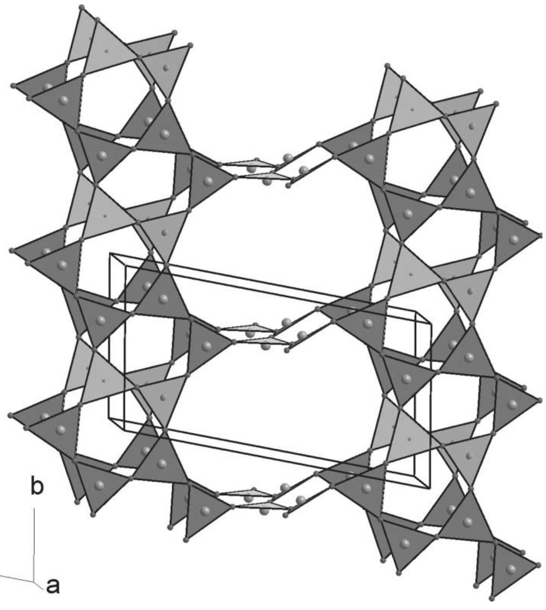

Figure 2
Hypothetical boron oxide B$_2$O$_3$: space group $P1$, a = 4.616, b = 6.609, c = 12.480 Å, α = 80.47, β = 104.94, γ = 90.00°, R = 0.0057, PCOD1062004.

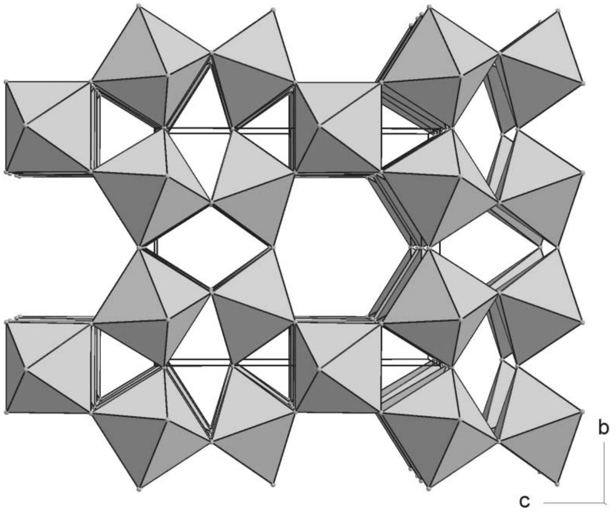

Figure 3
One of the yet to be synthesized virtual AlF$_3$ pyrochlore/perovskite intergrowths: space group $P\bar{4}m2$, a = 6.876, c = 8.258 Å, R = 0.0054, PCOD1020402.

The complete exploration of the zeolites is still not finished. More than a thousand models are expected to be produced by *GRINSP* with $R < 0.01$ and cell parameters <16 Å. The PCOD database (Le Bail, 2004) already contains more than 300 models, mainly in cubic and hexagonal symmetry. Examples establishing the quality of the predictions are presented in Table 1, showing some of the already known zeotypes retrieved by the program. The CIF files can be obtained by consulting the PCOD, giving the entry number provided with the figure captions, for instance PCOD1030081 (Fig. 1).

Not many crystalline varieties are known for the B$_2$O$_3$ composition. Too many were proposed by *GRINSP*, even reducing the limit to $R < 0.006$ (see an example in Fig. 2).

Apart from the well known perovskite structure type, which can be retrieved in almost all space groups during the exploration of the 6-connected three-dimensional nets with *GRINSP*, all the known structure types with AlF$_3$ formulation were retrieved (Table 2), including the most complex one recently discovered, τ-AlF$_3$ (Le Bail *et al.*, 1992). A series of 'yet to be synthesized' AlF$_3$ polymorphs were also proposed, one example being presented in Fig. 3. A detailed study of the hypothetical MF$_3$ phases (M = Al, Cr, V, Fe, Mn, Ga) will be published elsewhere (Le Bail, 2005).

3.2. Ternary compounds

For ternary compounds, M and M' cations are considered. They could have the same coordination but different ionic radii (enabling the exploration of ordered aluminosilicates or aluminophosphates *etc.*) or different coordination (exploring calcium–aluminium fluorides, titanosilicates, gallophosphates, borosilicates *etc.*), but the current limitation with *GRINSP* is that the connections by X atoms will only be by corner sharing: all X atoms should be connected to exclusively two M atoms or two M' atoms or one M and one M' atom. As a consequence, only some formulations can occur which fulfill these conditions. Moreover, if M or M' are not able to form electrically neutral binary compounds with corner-sharing only, then the built ternary compound will also not be electrically neutral. All the borosilicates formed with *GRINSP* are automatically electrically neutral. There is only one hit in the ICSD database for this kind of compound. A strange result is that *GRINSP* produces a huge quantity of hypothetical borosilicates, showing exclusively [BO_3] triangles and [SiO_4] tetrahedra linked by corners. Limiting $R < 0.006$, and working in cell symmetry higher or equal to monoclinic, but using the general Wyckoff position of the $P1$ space group, 57 different models were found with SiB_2O_5 formulation, 32 $Si_3B_4O_{12}$ models, 28 $Si_2B_6O_{13}$ and $Si_4B_2O_{11}$ models, 24 $Si_2B_2O_7$ models, 18 for SiB_6O_{11}, 17 for SiB_4O_8, 14 for $Si_3B_2O_9$, six for $Si_6B_2O_{15}$, and two $Si_3B_6O_{15}$ models. Moreover, 369 different additional models were disclosed in triclinic symmetry! The number of these models would probably explode if a complete search was done in the 230 space groups, since the introduction of Wyckoff positions having more than one equivalent boosts the capacity of the *GRINSP* software when experiencing difficulties to find structures more complex than 10–20 independent M/M' atoms in a triclinic cell. Those hypothetical borosilicates are not all yet included in the PCOD. One example is shown in Fig. 4.

Explorations in the titanosilicates domain (in fact a part of that domain where octahedra and tetrahedra are exclusively corner-linked) are in progress. The models are not electrically neutral so that the frameworks would have to accept some additional cations or charged molecule to exist in reality. One example is shown in Fig. 5.

3.3. By-products of the search with *GRINSP*

Other sixfold polyhedra than octahedra can be obtained: trigonal prisms or pentagonal-based pyramids. Since they do not correspond to unique ideal X–X or M–X distances, they are ranked with high R values. Aluminium is not known in solid fluorides with coordination other than very regular octahedral, so that such predictions are very probably useless, at least for an AlF_3 formulation. However, from the point of view of the structures, surprisingly some presented very small framework densities, showing large tunnels, and may be of interest. Two examples are shown in Figs. 6 and 7.

Moreover, many two-dimensional compounds can be formed which correspond to all polyhedral connections being satisfied by corners. In such cases, *GRINSP* has no way of making any correct estimate of the intersheet distance at the optimization stage, and thus these models are not collected (they frequently correspond to extremely small FD values). Some one-dimensional models have even been built (nanotubes with B_2O_3 formulation for instance; Fig. 8), but again, the distances between the rods could not be estimated and the cell parameters are fanciful.

4. Prediction confirmation

More difficult even than structure prediction would be the prediction of the synthesis conditions for realising these hypothetical crystal structures. However, if the chemical composition is complex enough (at least ternary or quaternary), one may first try the battery of solid-state classical synthesis routes with the suggested compositions (this being of no help at all for binary compounds). For instance, the calcium and sodium fluoroaluminates were only partly explored by *GRINSP* up to now, combining octahedra with different sizes (AlF_6 with CaF_6 or NaF_6). Some known 6-connected frameworks were retrieved, such as $[Ca_4Al_7F_{33}]^{4-}$, which actually exists as $Na_4Ca_4Al_7F_{33}$ (Hemon & Courbion, 1990), or $[NaAl_6F_{21}]^{2-}$, known in $Rb_2NaAl_6F_{21}$ (Le Bail *et al.*, 1989). One of the latest discovered metastable τ-AlF_3 variety (Le Bail *et al.*, 1992) was obtained from the thermolysis of either an organometallic compound $[(CH_3)_4N]AlF_4.H_2O$, or amorphous $AlF_3.xH_2O$ ($x < 0.5$). Thus, if a *GRINSP* version had existed before 1990, it would possibly have helped to solve the τ-AlF_3 structure, the solution of which was long delayed until a pure and sufficiently well crystallized powder could be obtained (no single crystal of suitable size available), or the synthesis of $Na_4Ca_4Al_7F_{33}$ may have been suggested sooner. Another hypothetical framework suggested by *GRINSP* in this series, which could well be viable, is that of $[Ca_3Al_4F_{21}]^{3-}$ (Fig. 9). Consequently, the idea

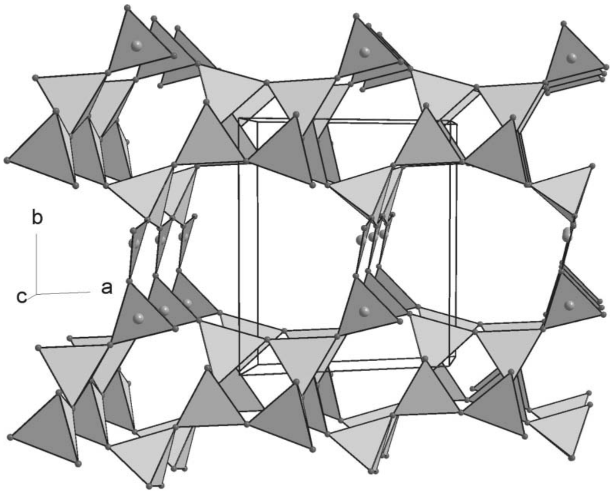

Figure 4
Combination of [SiO_4] tetrahedra and [BO_3] triangles connected by corners. Hypothetical $Si_5B_2O_{13}$: space group $P1$, a = 9.108, b = 9.602, c = 4.952 Å, α = 90.00, β = 123.92, γ = 90.00°, R = 0.0055, PCOD2050102.

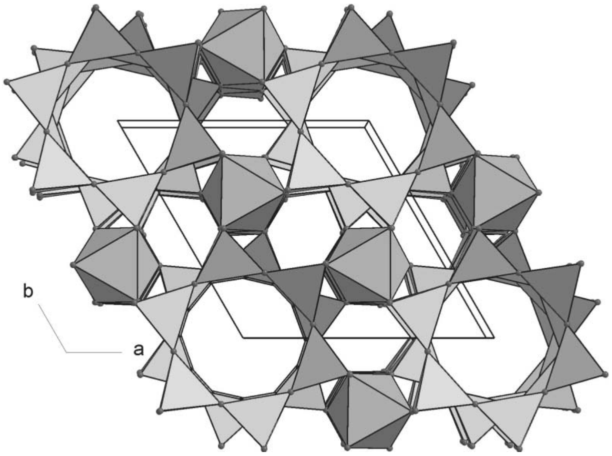

Figure 5
Combination of octahedra [TiO_6] and tetrahedra [SiO_4] connected by corners. Hypothetical titano-*cyclo*-silicate $[Si_3TiO_9]^{2-}$: space group $P6cc$, a = 9.411, c = 9.757 Å, R = 0.0047, PCOD2030304.

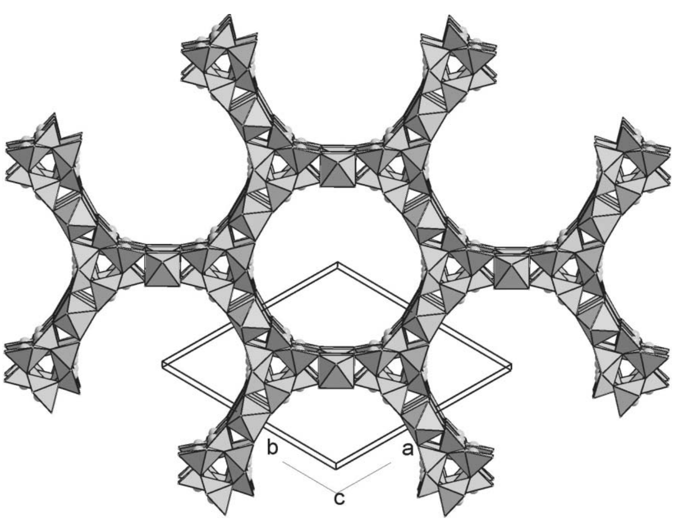

Figure 6
Framework built up from octahedra and pentagonal pyramids: space group $P6_3/mcm$, a = 14.708, c = 6.861 Å, R = 0.045, PCOD9000001.

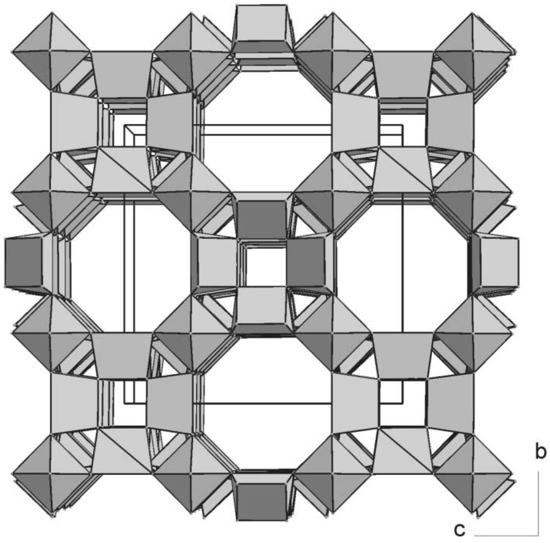

Figure 7
Framework built up from octahedra and trigonal prisms: space group $Im\bar{3}m$, a = 13.371 Å, R = 0.048, PCOD9000002.

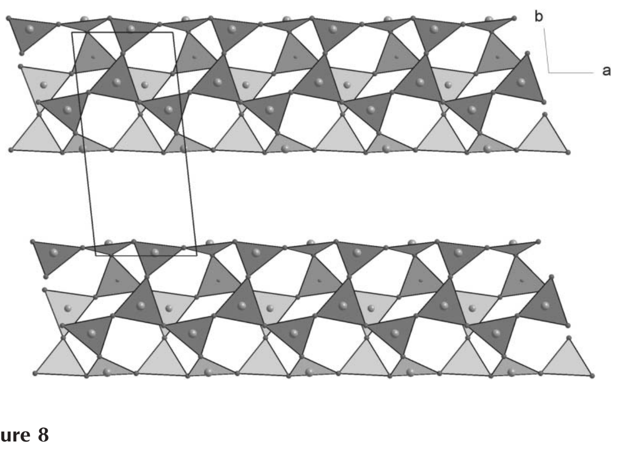

Figure 8
Triangles [BO_3] connected by corners. Hypothetical nanotubes with B_2O_3 formulation: space group $P1$, a = 4.663, b = 10.249, c = 9.794 Å, α = 89.59, β = 81.71, γ = 98.04°, R = 0.0058, PCOD1062005.

was to try to synthesize compounds with formulations $M_3Ca_3Al_4F_{21}$ (M = Li, Na, K, Rb, Cs). Unfortunately, attempts (using the solid-state route) have failed to produce the desired structure. Attempts at confirming the hypothetical titanosilicates predicted by *GRINSP* could be worth pursuing (for instance $M_2Si_3TiO_9$ shown in Fig. 5).

We can already be sure that most predictions will be in vain, and never confirmed, because the synthesis route may depend on a precursor (organometallic, hydrate, amorphous compound) which itself is yet unknown, or because the prediction is simply false. For the confirmation of some of the predictions gathered in the PCOD database, we may have to wait for decades or centuries. Nevertheless, structure prediction is an unavoidable part of our future in crystallography and chemistry. A further prediction is that the accuracy of the structure prediction methods will considerably improve.

5. Further planned improvements

The introduction of more complexity in the predictions can be readily imagined, by authorizing the connection of polyhedra by corner-, edge- and face-sharing, altogether, and by enabling the automatic re-establishment of electrical neutrality by the detection of holes and the filling of these holes by appropriate cations.

It is clear that the R factor considers only the X–X intrapolyhedra distances, neglecting any X–X interpolyhedra distances. This cost function, R, could possibly be better defined differently, for instance by using the bond valence sum rules, or energy calculations.

The way *GRINSP* recognizes an already existing or predicted structure is by comparison of the coordination sequence (Meier & Moeck, 1979) of any model with a list of previously established ones (as well as with the other coordination sequences already stored during the current run). This method is in fact insufficient because it may occur (scarcely) that the coordination sequences of two different models can be identical up to the tenth order. Therefore, other means are needed in order to differentiate structures (vertex symbol for instance).

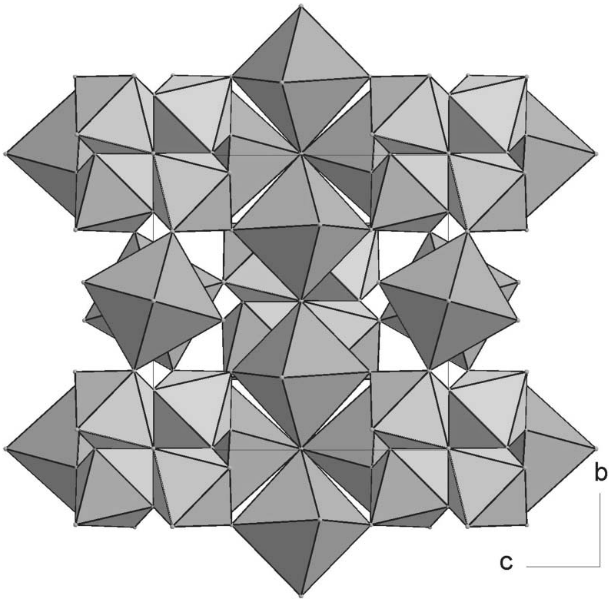

Figure 9
Combination of octahedra with two different sizes. Hypothetical $[Ca_3Al_4F_{21}]^{3-}$: space group $P\bar{4}3n$, a = 9.160 Å, R = 0.0127, PCOD1010005. One can distinguish the tetrahedra of [AlF_6] octahedra existing in the τ-AlF_3 variety and in the pyrochlore structure type.

A problem is the long calculation time. For instance, installed on a single-processor PC running at 2 GHz, the *GRINSP* software needs one day to examine one set of chemical elements in one space group (realising 20 000 to 200 000 runs of 300 000 Monte Carlo tests in each run), for random search of composition and random cell parameters (<16 Å), so that the full exploration needs 230 days! Moreover, one given model can be retrieved in different space groups with slightly different R values. Exploring the 230 space groups is a tedious task. One can imagine using a parallel computer, or grid computing, with a *GRINSP* version which would allow also the random selection of a space group, so that one run would provide the optimal model for each structure type, the best results being sorted out only at the end of such a global process.

6. Conclusions

Combining accurate structure and properties predictions would provide inorganic chemists with invaluable information enabling them to concentrate their synthesis efforts on compounds of interest. The *GRINSP* and *GRINS* computer programs are a small step in the direction of such an ambitious vision. They are potentially able to suggest thousands of hypothetical inorganic structures with complex formulations (ternary and quaternary compounds).

7. Program features

7.1. Hardware and software environment

The executable program was built by using the Compaq Visual Fortran compiler. It runs on a PC under Windows 9x/2000/Me/NT/XP. No DLL is necessary. There would be no serious problem in installing *GRINSP* on Unix platforms by using a different Fortran 77 compiler (though a few compiler-dependent subroutines would have to be adapted, mainly those calculating the elapsed CPU time).

7.2. Program specifications

The maximum number of M/M' atoms is 64. *GRINSP* explores the 3-, 4-, 5- or 6-connected nets leading to corner-sharing polyhedra, in binary (M_2X_3, MX_2, M_2X_5, MX_3) or ternary ($M_uM'_vX_w$) compounds. A file (Wyckoff.txt) contains the general and special positions of the 230 space groups. The user provides his or her own set of ideal interatomic distances inside of a text file (distgrinsp.txt). The coordination sequences avoiding the proposal of already predicted structures are gathered inside a text file (connectivity.txt). The parameters describing the conditions for a run need to be prepared in a short entry file (with .dat extension), containing a title, the space group, the choice of $M/M'/X$ atoms, the minimum and maximum cell parameters, the minimum and maximum framework density, the number of independent tests, the number of Monte Carlo events in each test, the maximum R value for retaining a model, the number of Monte Carlo events at the cell and atomic coordinate optimization stage, and the initial file name. Output files contain the atomic coordinates in the $P1$ space group, in CIF format as well as in a .dat file, the latter being directly readable by the structure drawing software *STRUPLO*/*STRUVIR*, producing VRML files which can be displayed in three dimensions by visualizer software (*CosmoPlayer*, *VrWeb etc.*). A series of test file examples are provided.

7.3. Program availability

GRINSP is available *via* http://www.cristal.org/grinsp/. The software is free of charge for non-profit organizations, and is delivered with the Fortran source code under the GNU Public Licence. The installation instructions and the user manual are accessible *via* the Web in HTML format, as well as included in the package.

7.4. PCOD database

Most of the hypothetical structures predicted by *GRINSP* were included in the PCOD (Predicted Crystallography Open Database), freely available *via* http://www.crystallography.net/pcod/. The search by elements, formula or/and cell parameters is possible through an Apache/MySQL/PHP server, delivering directly the CIF and VRML files. The search can also be performed by using the PCOD entry number, as given in the above figure captions. The database accepts the upload of any new hypothetical structure, organic as well as inorganic.

References

Baerlocher, Ch., Hepp, A. & Meier, W. M. (1978). *DLS76, A Program for the Simulation of Crystal Structures by Geometric Refinement*. Lab. f. Kristallographie, ETH, Zürich.

Catlow, C. R. A. (1997). *Computer Modelling in Inorganic Crystallography*. New York: Academic Press.

Catlow, C. R. A. & Price, G. D. (1990). *Nature (London)*, **347**, 243–248.

Cerius2 (2000). Version 4.2. Molecular Simulations Inc., Cambridge, UK.

Daniel, P., Bulou, A., Rousseau, M., Nouet, J., Fourquet, J. L., Leblanc, M. & Burriel, R. (1990). *J. Phys. Condens. Matter*, **2**, 5663–4677.

Deem, M. W. & Newsam, J. M. (1989). *Nature (London)*, **342**, 260–262.

Deem, M. W. & Newsam, J. M. (1992). *J. Am. Chem. Soc.* **114**, 7189–7198.

Foster, M. D. & Treacy, M. M. J. (2004). *Hypothetical Zeolites*, http://www.hypotheticalzeolites.net/.

Fourquet, J. L., Riviere, M., Le Bail, A., Nygrens, M. & Grins, J. (1988). *Eur. J. Solid State Inorg. Chem.* **25**, 535–540.

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

Hawthorne, F. C. (1994). *Acta Cryst.* B**50**, 481–510.

Hemon, A. & Courbion, G. (1990). *J. Solid State Chem.* **84**, 153–164.

Herron, N., Thorn, D. L., Harlow, R. L., Jones, G. A., Parize, J. B., Fernandez-baca, J. A. & Vogt, T. (1995). *Chem. Mater.* **7**, 75–83.

Le Bail, A. (2004). Predicted Crystallography Open Database, http://www.crystallography.net/pcod/.

Le Bail, A. (2005). *Z. Kristallogr.* Submitted.

Le Bail, A., Fourquet, J. L. & Bentrup, U. (1992). *J. Solid State Chem.* **100**, 151–159.

Le Bail, A., Gao, Y. & Jacoboni, C. (1989). *Eur. J. Solid State Inorg. Chem.* **26**, 281–288.

Le Bail, A., Jacoboni, C., Leblanc, M., De Pape, R., Duroy, H. & Fourquet, J. L. (1988). *J. Solid State Chem.* **77**, 96–101.

Lufaso, M. W. & Woodward, P. M. (2001). *Acta Cryst.* B**57**, 725–738.

Meier, W. M. & Moeck, H. J. (1979). *J. Solid State Chem.* **27**, 349–355.

Mellot-Draznieks, C., Girard, S., Férey, G., Schön, C., Cancarevic, Z. & Jansen, M. (2002). *Chem. Eur. J.* **8**, 4103–4113.

Mellot-Draznieks, C., Newsam, J. M., Gorman, A. M., Freeman, C. M. & Férey, G. (2000). *Angew. Chem. Int. Ed. Engl.* **39**, 2270–2275.

Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Lommerse, J. P. M., Mooij, W. T. M., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2002). *Acta Cryst.* B**58**, 647–661.

Newsam, J. W., Deem, M. W. & Freeman, C. M. (1992). *Accur. Powder Diffr. II, NIST Spec. Publ.* **846**, 80–91.

Payne M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. (1992). *Rev. Mod. Phys.* **64**, 1045–1097.

Schön, J. C. & Jansen, M. (2001*a*). *Z. Kristallogr.* **216**, 307–325.

Schön, J. C. & Jansen, M. (2001*b*). *Z. Kristallogr.* **216**, 361–383.

Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.

Woodley, S. M. (2004). *Application of Evolutionary Computation in Chemistry*, Vol. 110, edited by R. L. Johnston, pp. 95–132. Berlin: Springer-Verlag.