Inorganic Structure Prediction: Too Much And Not Enough

Armel Le Bail

Université du Maine, Laboratoire des Oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France

alb@cristal.org

Abstract. Elementary to sophisticated approaches allow for crystal structure prediction from empirical and theoretical principles since a few decades. Some recent efforts in inorganic chemistry for the production and classification of structure-candidates are described. The discussion is focused on an incredibly huge list of hypothetical zeolites, more modest series of aluminum fluoride polymorphs, or titanosilicates (etc) build up from the exploration of a special topological class, the N-connected 3D frameworks.

Introduction

Whether it is possible to predict the existence of crystal structures is a major scientific question. Due to the obvious interest in relying more on planning than on serendipity for chemical synthesis, and thanks to the continuing increase in computer power, times are coming for the systematic prediction of the crystal structures of inorganic compounds. It is enlightening to cite shortly 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 pseudo-potential plane-wave code [1]. The structures gathered in the database of hypothetical zeolites [2] are produced from a 64-processor computer cluster grinding away non-stop, generating graphs and annealing them, the selected frameworks being then re-optimized with the General Utility Lattice Program, *GULP* [3] using atomic potentials. *GULP* itself is able to predict crystal structures (TiO_2 polymorphs). Recently, a genetic algorithm was implemented [4] in *GULP* in order to generate crystal framework structures from the knowledge of only the unit cell dimensions and constituent atoms (this being closer to structure determination than to 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. A concept of 'energy landscape' of chemical systems is used by Schön and Jansen [5-6] for structure prediction with their computer program *G42*. Another package, *SPuDS*, is dedicated especially to the prediction of perovskites [7]. The *AASBU* method (Automated Assembly of Secondary Building Units) [8-9] is using *Cerius2* [10] and *GULP* in a sequence of simulated annealing plus minimization steps for the aggregation of large structural motifs. This list of software is finally small owing to the fact that structure and properties prediction is obviously an unavoidable part of our future in crystallography and chemistry. From these approaches, if zeolites are excluded, the number of inorganic crystal predictions of new probable structures is of the order of a few dozens of compounds, including still unknown varieties of carbon, sodium chloride, AB_2 compounds (etc). Approaches can be elementary up to highly sophisticated (*ab initio*), the latter option being still computer time consuming. On the elementary side, the new *GRINSP* (Geometrically Restrained INorganic Structure Prediction) code [11] for the building of N-connected 3D nets (N = 3, 4, 5, 6 and binary combinations) allows for the exploration of single or mixed frameworks, extending a lot the domain of possible investigations. Zeolites and some *GRINSP* predictions are discussed below.

## Four-connected 3D nets: zeolites

The more fabulous example of crystal structure prediction consists in the > 1.000.000 zeolite models gathered into the hypothetical zeolites database [2], though it includes many similar models for a

same zeotype (of which less than 200 are really observed). Let us recall that the Inorganic Crystal Structure Database (*ICSD*) contains less than 100.000 inorganic crystal structures, as a whole. This zeolite database should facilitate the identification and structure determination of any new real zeolite by a simple search-match against the calculated powder patterns. In that case, the predicted/observed (P/O) ratio > 5.000 is looking quite too much (though one would need for the effective P/O calculated for identified unique predicted zeotypes, of which the number is not yet clearly provided). Among hypothetical compounds are those already observed, those which will be confirmed (C) later, hopefully, and those being wrong. Efficient predictions should lead to a ratio P/(O+C) ~1. Unfortunately, C is hardly exactly predictable up to now, suggesting that our theory of materials is quite imperfect (ranking by energy level and quantum mechanics calculation). For the building of useful databases, one needs first to identify the O among the P, and this is yet not a simple task [12]. The number of new real zeotypes identified is less than ten per year. In the current state of the hypothetical zeolites database, only half of these newly discovered zeolites (last year) was predicted [13], so that the predictions will continue up to attain several more million models.

The P/O ratio (zeotypes only) for zeolites predicted by *GRINSP* is currently close to 50, but the study was limited to cell parameters smaller than 16 Å, and only 1/3 of the known zeotypes was retrieved. If several millions zeolites is probably too much, some sure way to remove wrong models having clearly to be found, other structural families are certainly not enough explored yet.

Six-connected 3D nets: MX$_3$ formulation

For MX$_3$ compounds with corner-sharing octahedra, the total number of structure-types predicted by *GRINSP* after a one-year first calculation campaign is quite small (a dozen), including only 5 new virtual models [14]. In the case of AlF$_3$, the five known varieties were reproduced (α-, β-, η-, κ- and τ-AlF$_3$) and seven hypothetical models were predicted. Among these still to be synthesized AlF$_3$ phases are recognized two known structure types (TlCa$_2$Ta$_5$O$_{15}$, Ba$_4$CoTa$_{10}$O$_{30}$) and some easy to imagine intergrowths, however, a few others are completely unexpected, though simple. Energy calculations using *WIEN2K* show that they all seem to have chances to be synthesized some day. A second campaign of calculations by using *GRINSP* has already doubled this number of hypothetical models. One of the new predictions is represented on Fig. 1. The frequency of discovery of new real MX$_3$ polymorphs is much smaller than for zeolite. The latest structure type (τ-AlF$_3$) was determined from powder diffraction data in 1992 [15], so that, presumably, we will have to wait for a long time before confirmations of these predictions, if any.
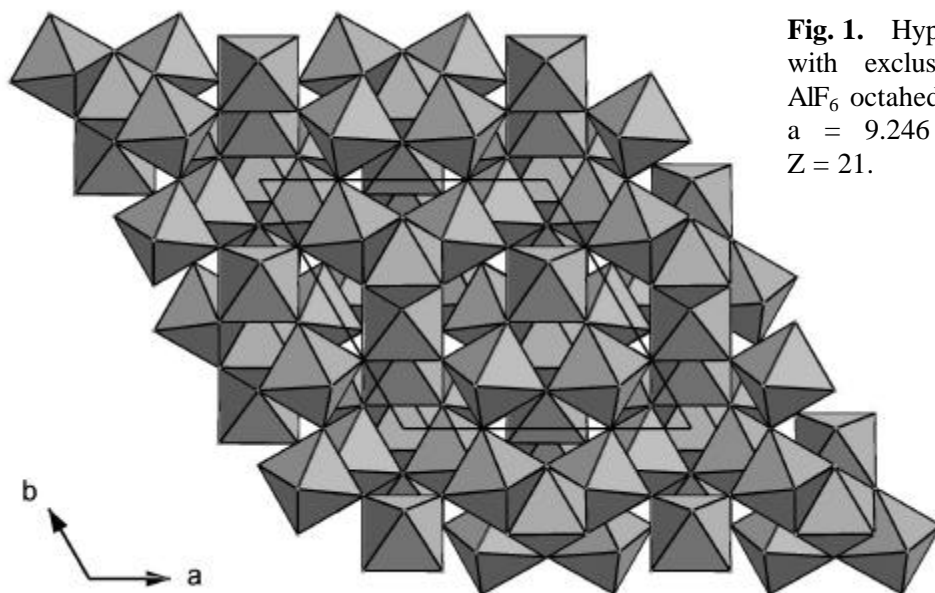
Fig. 1. Hypothetical AlF$_3$ model with exclusive corner-sharing of AlF$_6$ octahedra. Space group R-3m, a = 9.246 Å, c = 13.337 Å, Z = 21.

Mixed frameworks (4,6)-3D nets : titanosilicates and others

More than 1000 titanosilicates built up by *GRINSP* from corner-sharing TiO_6 octahedra and SiO_4 tetrahedra were recently included into the *PCOD* (Predicted Crystallography Open Database) [16]. A large majority of these models has the general formula $[TiSi_nO_{(3+2n)}]^{2-}$. The most numerous models are those with n = 1, 2, 3, 4 and 6, with respectively 93, 179, 174, 205 and 158 models corresponding to the satisfaction of a reliability criterion R < 0.02 (based on the close respect of ideal interatomic distances, and used as a cost function during the Monte Carlo prediction process). Examination of the *ICSD* database leads to establish a list of at least 2581 structures with similar formulations (Table 1). The 20 "best" models (with smallest R factors) is provided on Table 2, with partial identification.

Table 1. Numbers of compounds in *ICSD* version 1-4-1, 2005-2 (89369 entries) potentially fitting structurally with the $[TiSi_nO_{(3+2n)}]^{2-}$ series of *GRINSP* predictions, adding either C, C_2 or CD cations for electrical neutrality.

|	+C	+C_2	+CD	Total	*GRINSP*
ABX_5	300	495	464	35	1294	93
AB_2X_7	215	308	236	11	770	179
AB_3X_9	119	60	199	5	383	174
AB_4X_{11}	30	1	40	1	72	205
AB_5X_{13}	9	1	1	0	11	36
AB_6X_{15}	27	1	13	1	42	158
Total					2581	845

Table 2. The 20 hypothetical titanosilicate frameworks with smallest R factors as produced by *GRINSP*, with their *PCOD* entry numbers.

Formula	*PCOD*	R(%)	FD	a	b	c	α/β/γ	Z	SG	Identification
Si_2TiO_7	2200001	0.30	20.3	5.168	12.421	5.172	117.0	2	$P2_1/m$	SiP_2O_7-I
$SiTiO_5$	2200002	0.30	21.9	7.481	7.500	7.534	119.7	4	$P2_1/c$	titanite
$SiTiO_5$	2200003	0.32	24.0	6.544	6.544	7.793	90.	4	$P-42_1c$	α-VPO_5
Si_2TiO_7	2200004	0.33	20.0	6.587	7.314	6.583	108.6	2	$P2/m$	
$SiTi_2O_8$	2200005	0.33	19.1	9.822	9.824	6.522	90.	4	Pnma	
Si_2TiO_7	2200006	0.35	19.8	7.141	5.174	8.965	113.7	2	$P2_1/c$	keldyshite
$SiTiO_5$	2200007	0.37	21.7	7.561	7.484	6.522	90.	4	$P2_12_12_1$	β-VPO_5
Si_3TiO_9	2200008	0.37	18.0	9.751	6.491	7.008	90.	2	Pma2	
Si_3TiO_9	2200009	0.38	18.0	9.755	14.012	6.512	90.	4	Pnna	
Si_6TiO_{15}	2200010	0.41	22.8	10.049	10.049	14.066	120.	4	P6/mcc	sogdianite
Si_2TiO_7	2200011	0.43	20.9	7.626	8.642	8.701	90.05	4	$P2_1/c$	
$SiTiO_5$	2200012	0.44	24.2	6.513	6.513	7.783	90.	4	P-4c2	
Si_6TiO_{15}	2200013	0.44	19.3	14.070	7.266	7.082	90.	2	Pmma	elpidite
Si_2TiO_7	2200014	0.44	19.1	7.277	7.277	10.505	90.	4	$P4_2/mmc$	
$SiTiO_5$	2200015	0.44	24.2	5.004	9.212	7.782	113.0	4	$P2/m$	
Si_6TiO_{15}	2200016	0.45	22.2	12.776	12.776	6.690	120.	3	R-3	
Si_3TiO_9	2200017	0.45	19.3	7.000	7.000	9.776	120.	2	$P6_3/m$	wadeite
$SiTiO_5$	2200018	0.46	20.2	11.361	6.630	5.260	90.	4	$P2_12_12_1$	Na_2ZrSiO_5
Si_4TiO_{11}	2200019	0.46	22.4	15.511	7.565	7.622	90.	4	Pbcn	
Si_3TiO_9	2200020	0.47	21.4	9.411	9.411	9.757	120.	4	P6cc	

Those models presenting the smallest R values are built up from perfect octahedra and tetrahedra. Not all of them were identified as corresponding to a real titanosilicate (or to a given structure-type, whatever the formula) counterpart. Along this survey, it appeared that a lot of minerals with different names are indeed topologically equivalent, presenting the same coordination sequence (CS). For instance, this is the case with kieserite, lacroixite, montebrasite, amblygonite, titanite, etc, a lists to be

completed with many synthetic compounds such as $Ni_{0.5}TiPO_4$, $NaVOPO_4$, $VPO_4 \bullet H_2O$, etc, having the same CS as the *PCOD*2200002 model. The identification problem, manually simple for only one structure (an extended search in ICSD by using cell and formula criteria), becomes extremely time-consuming when more than 1000 structures have to be analyzed. Most of these "best" models are simple, with a small cell volume, they present a small number of different Si or Ti nodes (with common coordination sequences), generally a single node for each of them or 2 Si nodes and one Ti node, maximum. Their framework density (FD) is in the range 18 to 24 Si/Ti per 1000 $Å^3$. Two of these "best" models are presented on Fig. 2, still unidentified, however, more efforts are needed. They contain generally small cages or not more then 6-rings tunnel section. Models showing the largest cavities and/or tunnels were previously analyzed [17], corresponding to FD close to 10, as small as for the best zeolites.
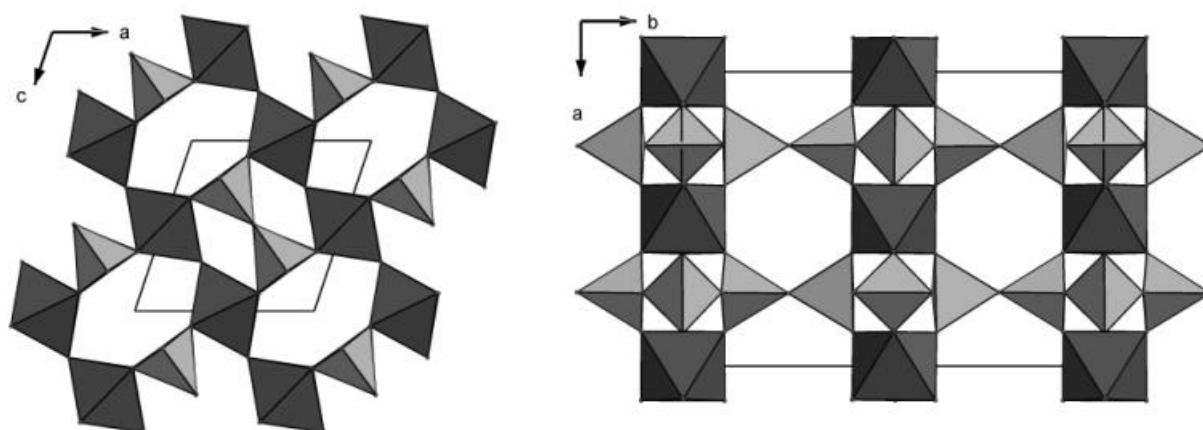
Fig. 2. Two of the "best" titanosilicates predicted by *GRINSP*, with *PCOD* entry numbers 2200004 (left) and 2200009 (right).

Using *GRINS*, a satellite program inside of the *GRINSP* package, the feasibility of isomorphous compounds was tested. *GRINS* can read a multiple *CIF* issued from *GRINSP* and try the cationic/anionic replacement much faster (a few hours for one thousand models) than if a full prediction was undertaken. Series of zircono-, niobio-, vanadyl-silicates as well as gallo-, vanadyl-, and titano-phosphates, or sulfates, were built starting from the titanium silicates and inserted into the *PCOD*. Containing now more than 10.000 entries. It seems quite probable that several of these hypothetical structures will be confirmed in a near future. However, without the availability of an easy to use tool allowing to compare fast a given structure to a list of possibilities, the recognition will not occur. Generalizing the concept of CS may provide a possible solution to that problem.

Other series produced by *GRINSP*

B_2O_3 – Very few crystalline polymorphs are known with such a formulation, characterized by $[BO_3]$ triangles sharing exclusively corners. Thousands of hypothetical models were proposed by *GRINSP*, even limiting to R < 0.006, certainly too much, but maybe an indication for the B_2O_3 amorphous preference.

V_2O_5 – More than 500 hypothetical models built up from $[VO_5]$ square pyramids sharing exclusively corners are waiting for some analyzes before to be added into the *PCOD*.

Aluminophosphates, aluminofluorides – These studies combine two elements adopting a same polyhedra but different radii. The AlPO family is as large as the zeolites one, sharing many common structures. However, the $[PO_4]$ and $[AlO_4]$ tetrahedra dimension differences induce the occurrence of specific structures. Compounds such as calcium or sodium aluminofluorides were examined, with the combination of octahedra having two different sizes. In that case, the number of acceptable models is

quite much more reduced, nevertheless allowing to recognize known 6-connected frameworks with formula $[Ca_4Al_7F_{33}]^{4-}$ and $[NaAl_6F_{21}]^{2-}$ corresponding respectively to the really existing compounds $Na_4Ca_4Al_7F_{33}$ and $Rb_2NaAl_6F_{21}$.

Borosilicates, boroaluminates – All hypothetical compounds produced by *GRINSP* with exclusive corner-sharing of $[BO_3]$ triangles and $[SiO_4]$ tetrahedra are automatically electrically neutral. There is only one structure of this kind in ICSD, though thousands are predicted, too much again, even limiting to R < 0.006. Boroaluminates combining triangles with octahedra are a bit less numerous.

Other series are currently examined, such as vanadium phosphates with corner-sharing $[PO_4]$ tetrahedra and $[VO_5]$ square pyramids, or octahedra and square pyramids producing some oxygen lacunary perovskites, etc. On Fig. 3 are depicted a few examples of these *GRINSP* predictions.
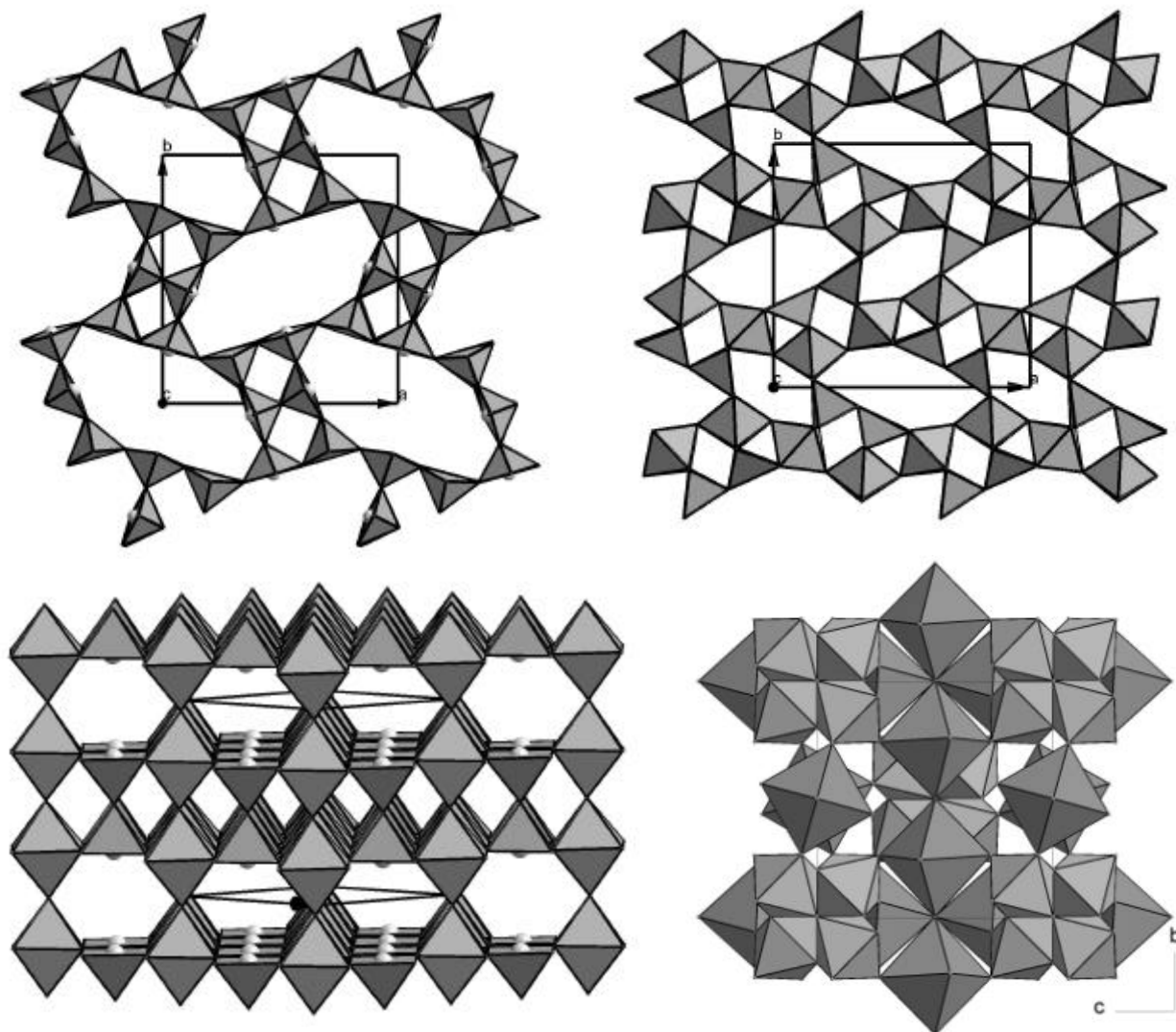
Fig. 3. Some *GRINSP* predictions for different combinations of polyhedra connected by corners: square pyramids (up left), square pyramids and octahedra (up right), which may lead to oxygen lacunary perovskites (down left), two different sizes of octahedra for an hypothetical $[Ca_3Al_4F_{21}]^{3-}$ framework (down right).

Conclusion

Accurate prediction of both structures and properties appears as one of the main scientific challenge of the XXIth century, finding some equilibrium between too much and not enough, by submitting the numerous structure-candidates built up from crystallography rules to validation by quantum

mechanics, waiting for even faster computers which would allow direct prediction by quantum chemistry. In spite of its limitations, *GRINSP* appears as an efficient generator of hypothetical crystal structures. Improvements would consist in the consideration of other linkage modes than only by corner-sharing (edge and face-sharing) and in the increase of the complexity to combinations of 3 different polyhedra, so as to be able to explore the large domain of the mixed octahedra-pentahedra-tetrahedra framework silicates, for instance. The usefulness of the *PCOD* will be maximal when powder patterns will be calculated and inserted into search-match tools for identification. All this appear to be attainable with automatization (locating holes, filling them with appropriate ions for electrical neutrality, optimizing according to bond valence rules), but will require considerable efforts in code developing. Then, some ill-crystallized compounds, with unindexable powder paterns, may well be finally characterized, more or less. Identification will be equivalent to a structure determination, when crystal structure prediction will attain total efficiency, in some future, the present paper being only a small step along of this long route. A modified version (GRINSPM, M for Metal) is now devoted to the prediction of structure candidates for intermetallic compounds.

References

[1] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos: Rev. Mod. Phys. 64, (1992) 1045-1097.

[2] Information on http://www.hypotheticalzeolites.net

[3] J.D. Gale: J. Chem. Soc., Faraday Trans. 93 (1997) 629-637.

[4] S.M. Woodley, in: Application of Evolutionary Computation in Chemistry, edited by R. L. Johnston, Structure and bonding series, Springer-Verlag 110 (2004) 95-132.

[5] J.C. Schön and M. Jansen: Z. Kristallogr. 216 (2001) 307-325.

[6] J.C. Schön and M. Jansen: Z. Kristallogr. 216 (2001) 361-383.

[7] M.W. Lufaso and P.M. Woodward: Acta Crystallogr. B57 (2001) 725-738.

[8] C. Mellot-Draznieks, J.M. Newsam, A.M. Gorman, C.M. Freeman and G. Férey: Angew. Chem. Int. Ed. 39 (2000) 2270-2275.

[9] C. Mellot-Draznieks, S. Girard, G. Férey, J.C. Schön, Z. Cancarevic and M. Jansen: Chem. Eur. J. 8 (2002) 4103-4113.

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

[11] A. Le Bail: J. Appl. Cryst. 38 (2005) 389-395.

[12] R. Hundt, J.C. Schön and M. Jansen: J. Appl. Cryst. 39 (2006) 6-16.

[13] M.M.J. Treacy, private communication.

[14] A. Le Bail and F. Calvayrac: J. Solid State Chem. 179 (2006) 3159-3166.

[15] A. Le Bail, J.L. Fourquet and U. Bentrup: J. Solid State Chem. 100 (1992) 151-159.

[16] Information on http://www.crystallography.net/pcod

[17] A. Le Bail: Z. Kristallogr. (2007) in press.