

The introduction of structure types into the Inorganic Crystal Structure Database ICSD

Rudolf Allmann^{a*} and Roland Hinek^b^aIm Grund 5, D-35043 Marburg, Germany, and ^bEltviller Strasse 20, D-65399 Kiedrich, Germany. Correspondence e-mail: rudolfallmann@t-online.de

Both the approach used and the progress made in the assignment of structure types to the crystal structures contained in the ICSD database are reported. Extending earlier work, an hierarchical set of criteria for the separation of isopointal structures into isoconfigurational structure types is used. It is shown how these criteria, which include the space group (number), Wyckoff sequence and Pearson symbol, c/a ratio, β ranges, ANX formulae and, in certain cases, the necessary elements and forbidden elements, may be used to uniquely identify the representative structure types of the compounds contained in the ICSD database.

© 2007 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

In 2005, FIZ Karlsruhe (Fachinformationszentrum Karlsruhe) began to introduce structure types into the Inorganic Crystal Structure Database ICSD (Bergerhoff *et al.*, 1983; ICSD, 2007). Technically, this was done by introducing new standard remarks (labels) into the database, called TYP and STP, which can be assigned to any entry – and subsequently be searched for. Each subset of entries, belonging to a given TYP label, is represented by one arbitrarily chosen member of this subset in order to serve as the prototype. This representative entry is in addition labelled by a STP remark.

Since the existing theoretical approaches to the definition of structure types (Parthé & Gelato, 1984, 1985; Burzlaff & Malinovsky, 1997; Bergerhoff *et al.*, 1999) already pointed at the difficulties one encounters when trying to determine structure types automatically, methods needed to be developed in order to be able to assign crystal structures contained in the ICSD to their corresponding structure types. This problem is also discussed in the first volume of structure types by Villars & Cenzual (2004).

The report of the IUCr Commission on Crystallographic Nomenclature entitled *Nomenclature of Inorganic Structure Types* (Lima-de-Faria *et al.*, 1990) provides some useful definitions of the different kinds of structure types. The two most important of them – *isopointal* and *isoconfigurational* structures – proved to be sufficient to serve as theoretical concepts in guiding our practical work with the ICSD database. According to Lima-de-Faria *et al.*, two structures should be described as *isopointal* if:

- (i) they have the same space-group type or belong to a pair of enantiomorphic space-group types; and
- (ii) the atomic positions, occupied either fully or partially at random, are the same in the two structures, *i.e.* the complete

sequence of the occupied Wyckoff positions (including the number of times each Wyckoff position is occupied) is the same for the two structures when the structural data have been standardized.

Note that, for not uniquely standardized structures, the Wyckoff sequence depends on the chosen cell origin, *e.g.* for those spinels crystallizing in space group $Ia\bar{3}d$ (with two standard settings with the origin chosen at $\bar{3}$ or $\bar{4}$), a shift of the origin (or alternatively of all atomic positions) by $\frac{111}{222}$ will change the Wyckoff sequence from ' $e d a$ ' into ' $e c b$ ', respectively.

Lima-de-Faria *et al.* define '*isoconfigurational structures*' as a subgroup of the *isopointal* structures, *viz* two structures are defined as *isoconfigurational* (configurationally isotypic) if:

- (i) they are isopointal; and
- (ii) for all corresponding Wyckoff positions, both the crystallographic point configurations (crystallographic orbits) and their geometrical interrelationships are similar.

Unfortunately, definition (ii) is not an explicit and constructive definition since the exact meaning of 'similar geometric interrelationships' is not specified (Parthé & Gelato, 1984, 1985; Burzlaff & Malinovsky, 1997; Bergerhoff *et al.*, 1999), and thus novel methods that combine different criteria needed to be introduced. According to Lima-de-Faria *et al.* (1990), we use an *a priori* definition of geometric criteria for distinguishing structure families.

The main body of this paper consists of three parts. In the next section, we will discuss the search criteria introduced in order to determine all those entries in the ICSD which belong to a given structure type. The third section covers some typical examples and the final section is devoted to a discussion and a brief outlook.

2. Search criteria (structure descriptors)

2.1. General description

In the ICSD, two crystal structures are regarded as isostructural if they are isoconfigurational. Note that for zeolite crystal structures only the framework atoms (Baerlocher *et al.*, 2001) are taken into account in the determination of isoconfigurational structures.¹ Such types will get the ending ‘-frame’.

In detail, our approach for the determination of isoconfigurational structures consists of the following two steps.

1. Determination of isopointal structure types characterized by space group, Wyckoff sequence and Pearson symbol. As we use the data ‘as-published’, all non-standard settings are considered separately, *i.e.* all space-group settings and all equivalent Wyckoff sequences that are used by the authors are taken into consideration.

2. Subdivision of isopointal [characterized by definition (i)] structures into different structure types by additional ‘structural descriptors’.

These are the fundamental steps in the determination of structure types in the ICSD database. At the beginning of our work, we focused on the introduction of structure types with high symmetry (cubic, tetragonal). It soon became evident that for this approach one must be able to manage a large amount of data in a well defined, systematic, reproducible and fast way as nowadays is provided by the use of up-to-date relational database techniques. Especially using the powerful ‘structured query language’ (SQL) as a workhorse – embedded within the relational database management system *MySQL*, which in fact stores the complete ICSD data – turned out to be essential (Reese *et al.*, 2002).

For the purpose of classification, *i.e.* the subdivision of isopointal structures, we had to consider further criteria (the structure descriptors) that define a structure type uniquely. It was indispensable to develop an easy-to-use database application tool with integrated *MySQL* database connectivity and full data access. Fig. 1 shows this tool providing a fast and highly automated process.

1. Recording all of the search criteria by their (alpha)numerical values and persistent storing into a suitable table. The grid in Fig. 1 shows the structure of this table.

2. Allowing an automated robust search of entries over the whole database – by generating the search conditions using the criteria stored in step 1 – and a subsequent comparison of the crystal structures found in the resulting search subsets.

3. Searching for intersections due to overlaps in search conditions (defined in step 1) automatically by running an appropriate SQL routine and subsequently resolving the found overlaps of structure types by fine-tuning of criteria.

4. Ultimately assigning structure types, *i.e.* labelling all entries of the whole database that match the criteria for all defined structure types (in step 1) with TYP or STP remarks,

¹ If the usual criteria that take into account all atoms (including guest atoms) had been taken into account, one would get for instance more than a hundred different faujasite structure types (as in Villars & Cenzual, 2004), whereas there is only one faujasite framework type.

Table 1

Progress in introducing structure types in the ICSD.

Release	STP†	TYP‡	(%)§	Total entries (100%)
2005-01	107	15874	18.4	86308
2005-02	109	16872	18.9	89384
2006-01	802	32970	37.0	89064¶
2006-02	1347	40170	42.9	93720
2007-01	1600	50717	52.1	97376
2007-2	2485	59291	59.1	100243

† Number of entries labelled with a STP remark, *i.e.* number of distinct structure types introduced. ‡ Number of entries labelled with a TYP remark, *i.e.* number of entries assigned to any structure type § Percentage of the number of entries labelled with a TYP remark referring to the total amount of entries ¶ In 2006-01, about 3000 duplicates were removed from the ICSD.

respectively. This indeed takes less than half an hour for about 100000 entries and 2485 distinct structure types, owing to the highly performing SQL engine of the database. In release 2007-2, about 59% of all the entries could uniquely be assigned to a structure type.

While our work continues on introducing further structure types, these four steps serve as the actual work flow for the production of each release again (twice a year). The progress over the past years in introducing structure types into the ICSD is visualized in Table 1.

For the large majority of entries it proved to be sufficient to use the following criteria.

For the definition of *isopointal structure types*:

- (i) equivalent space groups (or space-group number);
- (ii) equivalent Wyckoff sequences;
- (iii) the Pearson symbol.

For the subdivision into individual *isoconfigurational structure types*, the criteria:

- (iv) crystallographic composition type (ANX formula);
- (v) range of *c/a* ratios;
- (vi) beta range;
- (vii) necessary elements (combined by ‘and’ or ‘or’);
- (viii) forbidden elements (also combined by ‘and’ or ‘or’);
- (ix) atomic coordinates (by manual inspection, in a few cases only).

Which of the criteria (iv)–(vii) are actually used in order to define a special structure type is determined by a semiautomatic, and often iterative, trial-and-error procedure until the chosen descriptors for a given structure type suffice to obtain all representatives and only these representatives. Exactly this attempt of uniquely assigning all the representatives in the ICSD for a given structure type means a lot of hard pragmatic (iterative) work and indeed makes the difference between our approach and approaches that mainly rely on the definition of structure types only. The criteria (vii) and (viii) take into consideration the crystal chemistry: some elements occur in all representatives of a given type (*e.g.* O in oxide structures or F, Cl, Br, I in halides), whereas in intermetallics O is a ‘forbidden’ element.

When the assignment of structure types is completed, the user of the ICSD can ask for all representatives of a structure

type without bothering with all the different settings of space groups and cell origins because this is already done. In our effort to introduce structure types, we tried to cope with all the different settings, but some unusual settings may have been overlooked. Users of the ICSD who find a missing representative of an already introduced structure type are requested to inform FIZ Karlsruhe or the first author. Many of the remaining structures represent their own singular structure type (about 1/3 of all structures in the ICSD) and will not be registered as a structure type.

In a few cases only, these criteria do not suffice for a clear separation and then, as the ultimate and time-consuming step, the representatives of such a structure type must be set by hand, e.g. by checking the atomic coordinates. Fields 'Include' and 'Exclude' are used for this purpose.

2.2. Structure descriptors

In order to clarify the meaning and usage of the different structure descriptors, these criteria will be described in more detail in this section.

2.2.1. Space-group symbol. Using the space-group number, about 700 settings of space groups are immediately accessible in the ICSD [e.g. apart from the standard setting of space group number 14, i.e. $P121/c1$ (with 3566 structures), the following settings are also found: $P121/a1$ (804), $P121/n1$ (2389), $P1121/a$ (68), $P1121/b$ (115), $P1121/n$ (71), $P21/b11$ (38), $P21/n11$ (14), $B121/a1$ (2), $B121/c1$ (7), $B121/d1$ (6), $A21/d11$ (1) and $C1121/d$ (3)].

2.2.2. Pearson symbol. In addition to the original definition of Lima-de-Faria *et al.* (1990), the Pearson symbol (Bravais type plus the number of atoms per standard cell) is used as a structure descriptor. In contrast to the Wyckoff sequence, the Pearson symbol can (and should) be defined in such a way that it is independent of any cell transformation: just one unique symbol per Bravais type. Therefore, the symbols A , B , C and, in the monoclinic system, also I were unified to one symbol S for mono-side centred. The 14 symbols now used in the ICSD are: aP , mP , mS , oP , oS , oI , oF , tP , tI , hP , hR , cP , cI , cF . The number of atoms per unit cell is that of the standard setting, which for the rhombohedral structures is the primitive cell as used in *Pearson's Handbook*, even though in most cases the

The screenshot shows the 'FRM_StructureTypes' application window. At the top, there is a search bar with 'fes' entered. Below the search bar are buttons for 'Browse', 'Edit', and 'select All (9)'. There are also dropdown menus for 'CrystalSys', 'Status', and 'Paket', all set to 'Alle'. The main part of the window is a table with columns: CRY, STRUCT_TYPE, SGR, WYCK, PEAR, ANX, CA_MIN, CA_MAX, PROTO, BETA_MIN, BETA_MAX, SGR_NUM, ELEMENTS, NORLEM, ST, and PAKET. The table lists various structure types, with 'FeS2(cP12)' selected. Below the table, there are several text boxes: 'TypeDesc' containing 'Isopointal to C02(cP12); S-S-dumbbells: S-S=2.18', 'AET' containing 'Fe: octahedron: S6; S: tetrahedron: SFe3', and 'PearsonNot, Aliasnames' containing 'pyrite,SB:C2'. There are also fields for 'Excluded ColNr' and 'Included ColNr' with lists of numbers. At the bottom, there are buttons for 'Print entries list', 'Filter H+.D+', 'Filter H-.D-', and 'Visualize'.

CRY	STRUCT_TYPE	SGR	WYCK	PEAR	ANX	CA_MIN	CA_MAX	PROTO	BETA_MIN	BETA_MAX	SGR_NUM	ELEMENTS	NORLEM	ST	PAKET
CU	Cu	FM3-M	a	cF4				53755						A	First12
CU	W	IM3-M	a	cI2				53800						A	40_Cub
CU	C02(cP12)	PA3-	c a	cP12				22398				OR C,N	Th	A	40_Cub
CU	FeS2(cP12)	PA3-	c a	cP12				53935					F	A	40_Cub
CU	FeS2(cP12)	PA3-	c b	cP12				53935					F	A	40_Cub
CU	PdF2(cP12)	PA3-	c a	cP12				100567				OR F,0		A	40_Cub
CU	PdF2(cP12)	PA3-	c b	cP12				100567				OR F,0		A	40_Cub
TE	In	I4/MMM	a	tI2		1,258	1,783	53777						A	56_Tet
TE	Pa	I4/MMM	a	tI2		0,75	0,90	24622						A	Tet_TeI2

Figure 1

Part of the search list for structure types in the ICSD. Because atomic coordinates are not introduced in the search criteria, the representatives of the PdF₂- and CO₂-type cannot be distinguished from the FeS₂(pyrite)-type automatically and must be set by hand.

hexagonal setting is used in the ICSD (a change to the threefold hexagonal cell is currently under discussion).

The Pearson symbol has one additional advantage that it allows one to distinguish between fully occupied structures and those defect ones that have some positions only partially occupied. It suffers the drawback, however, that, for ammonium compounds that are isotypical to the corresponding potassium compound, the numbers of atoms per unit cell are different and thus the Pearson symbol changes for the ammonium compound.²

2.2.3. Wyckoff sequence. The Wyckoff sequences in the ICSD are not complete with respect to the H atoms in the crystal structures. The Wyckoff letters of the H atoms are systematically omitted since in earlier structure determinations H atoms were rarely located and their Wyckoff sites are quite frequently unknown.

The Wyckoff sequence also changes if the axes of the unit cell are interchanged (e.g. in *Pmmm* twofold axes run along **a**, **b** and **c** and the 12 sites $2i, 2j, \dots, 2t$ can be transformed into each other by cell shifts of $\frac{1}{2}$ in any direction or by interchanging the axes).

This manifold of equivalent Wyckoff sequences could have been reduced by standardizing all structures in the ICSD using a program such as *STRUCTURE TIDY* (Gelato & Parthé, 1987), but then relationships to similar structures in different space groups may have been lost. For example, monoclinic space-group settings like *P121/n1* or *I12/a1* are transformed to *P121/c1* and *C12/c1*, respectively, even when the monoclinic angles become greater than 120° and the directly discernible similarities of the reported structure to orthorhombic structures is lost. Further, two similar structures that have corresponding atoms with coordinates that are slightly above and below zero, respectively, are transformed to completely different standardized structures.³ Nevertheless, the inclusion of standardized data into the ICSD is currently under debate.

Finally, we would like to mention that *STRUCTURE TIDY* has been used for the determination of a standardized setting for a few *prototypical* structures. (*Prototype*: one arbitrarily chosen 'representative' entry of all entries belonging to the same structure type, see below. The prototype entry also contains a survey of the atomic environments.)

As already mentioned, the most complicated part of our approach is the separation of the isopointal structures into their individual isoconfigurational structure types. Identifying the isoconfigurational structures also requires the analysis of **axial ratios** (*c/a* ratios) which can result in transition of one structure type into another. One simple example in *I4/mmm* ($2a$ in 000) may illustrate this. For $c/a = 1$, one gets the cubic body-centred W-type, but for $c/a = (2)^{1/2} = 1.41$, one gets the cubic close-packed Cu-type (non-standard setting: *F4/mmm*

with $c/a' = 1$). Therefore, for the tetragonal representatives of the W- and the Cu-type, respectively, the borderline between the two types should be set at $c/a = (1.41)^{1/2} = 1.19$, i.e. the acceptable *c/a* ratio for a given special type should not deviate more than $\pm 20\%$ from the ideal value, an even sharper criterion would only allow deviations of $\pm 10\%$. The finally chosen ranges for *c/a* as well as for the angle β depend on the ranges found in the existing set of representatives.

In very exceptional cases, an examination of the *atomic positions* may be required too. For example, the isopointal structures with space group $P\bar{a}3$ ('*c a*' and non-standard '*c b*') have only one free parameter: the x value of position $8c: xxx$. For the pyrite family, dumbbells along the threefold axis exist for $x > 0.355$. For $0.32 < x < 0.355$ (PdF₂-type), the distances to the six other atoms on $8c$ become shorter than that along the threefold axis, i.e. there are no dumbbells any more. For very small values ($x \sim 0.11$), the atoms on $8c$ approach the atom on $4a$ and linear molecules C–A–C are formed (CO₂-type).^{4,5}

Among the used search criteria, there is also a field for the collection code (COL) of the *prototype* of a structure type. As mentioned above, the prototype of a structure type is an arbitrarily chosen representative of this structure type, mostly one of the early published structures. On request and with good reasons, the chosen prototype and with it the used name of the structure type can easily be changed. Structures belonging to the approximately 1600 prototypes that are currently identified can be searched for both in the program *FindIt* and the web version of the ICSD database, the details of the search procedure are described in *Appendix A*.

A final criterion that must be fulfilled before a new structure type is introduced into the ICSD is that it must represent the structures of at least three different compounds with the same given structure (sometimes only two representatives). Thus, for an estimated third of all structures in the ICSD no isotypic structures exist until now and therefore are not assigned to a structure type apart from self-assignment. With release 2007-01, about 52% of all the 97 000 structures in the ICSD had been classified into about 1600 structure types. The progress in introducing structure types in the ICSD is summarized in Table 1.

Statistics of the 1600 distinct structure types present in 2007-01:

- four structure types have more than 1000 representatives (Al₂MgO₄, CaTiO₃, GdFeO₃, and NaCl),
- 13 structure types have 500–999 representatives: (AuCu₃, CeAl₃Ga₂, CsCl, Cu, Cu₂Mg, K₂MgF₄, Mg₂SiO₄, MgSrSi, NaCrS₂, NdAlO₃, PbCl₂, PbClF and YbBa₂Cu₃O_{6+x}(*orth*),
- 16 structure types have 250–499 representatives,
- 53 structure types have 100–249 representatives,
- 95 structure types have 50–99 representatives,
- 155 structure types have 25–49 representatives.

² To avoid this problem, Bayliss *et al.* proposed a Pearson symbol that disregards the H atoms (Bayliss, 2000). Villars & Cenzual (2004) use this definition too.

³ In addition, a practical difficulty with the program *STRUCTURE TIDY*, which works well in about 90% of all cases, is that it does not transform the anisotropic displacement parameters automatically during the cell transformations in the standardization procedure.

⁴ Most structure types show a variation in the acceptable atomic coordinates and the ranges we finally accepted are somewhat arbitrary, but based on thorough analysis of the crystal structure.

⁵ Obviously, in some cases the subdivision of an isopointal type into its isoconfigurational structure types may not be complete and users are asked to inform FIZ if they find such incomplete separations.

The first 33 most frequent structure types contribute to about 1/3 of all assigned representatives, the first 336 structure types to about 3/4.

3. Examples

In some cases, all isopointal structures belong to one single isoconfigurational structure type only, e.g. the spinels in space group $Fd\bar{3}m$. The space-group number (227), the Pearson symbol (cF56) and the two equivalent Wyckoff sequences (eda and ecb) are sufficient to find all 1890 spinel structures present in the ICSD (1878 for ' eda ' and 12 for ' ecb ', TYP = 'Al2MgO4', COL number of the prototype: 56116).

As a supplementary criterion, the c/a range was used to separate ThSi_2 (Th surrounded by 12 Si) and $\text{TiO}_2(\text{tI12})$ (Ti octahedrally surrounded by six O atoms) in space group $I4_1/amd$ (No. 141), Pearson symbol tI12 and Wyckoff sequence ' ea ' or ' eb '. For $c/a = 2.11$ – 2.99 , one gets 15 (ea) + 12 (eb) = 27 representatives for $\text{TiO}_2(\text{tI12})$ and for the ThSi_2 -type with $c/a = 2.99$ – 4.14 , one obtains $75(ea) + 2(eb) = 77$ representatives.

Chemical information was needed to separate CoW_2B_2 and K_2PtS_2 in $Immm$, Pearson symbol oI10 and Wyckoff sequence ' hfa '. With necessary elements B or Si (and $c/a = 0.4$ – 0.5), one gets seven representatives of the intermetallic CoW_2B_2 -type and with elements O or S or Se six representatives of the K_2PtS_2 -type are obtained.

Finally, we would like to mention that isopointal structure types in space groups with a small number of Wyckoff letters tend to split into many structure types. E.g., for the isopointal structures that are characterized by space-group symbol $P12_1/c1$ (No. 14), Pearson symbol mP12 and Wyckoff sequence $e3$, 145 structures are found. Until now, 98 of them could be assigned to the following six structure types:

CeAsS (4 structures) with c/a ratio = 4–4.8;

CoSb_2 (18) with β range = 111 – 120° , pnictide element (except nitrogen) necessary;

CuP_2 (2) with $\beta = 110$ – 115° , Cu or Ag necessary;

$\text{ZrO}_2(\text{mP})$ (68) with $c/a = 0.92$ – 1.1 , $\beta = 90$ – 104° , O or S necessary;

NdAs_2 (2) with $\beta = 105$ – 107° , P or As or Sb necessary, O forbidden;

CaPSi (4) setting $P121/n1$ only, $\beta = 106$ – 110° .

For the last example, the 145 isopointal structures were at first standardized by *STRUCTURE TIDY*. The clustering of the standardization parameters (γ and CG) as well as the β angles around certain values were taken as an indication for a structure type. Then those criteria were chosen that clearly could separate all these clusters.

4. Discussion and outlook

For intermetallic phases, Villars & Calvert (1991, 1997) have compiled extensive lists of structure types and their representatives. A first compendium of structure types including ionic structures too (TYPIX) was published by Parthé *et al.* (1993) with more than 3600 critically evaluated data sets. In

2003, Villars & Cenzual started their voluminous compendium in book form.

Bergerhoff *et al.* (1999) have introduced a simple procedure to compare pairs of isoconfigurational structures. The product of the differences of standardized atomic coordinates and the ratios of lattice constants results in the so-called Δ value. The smaller Δ , the better the two structures coincide. Application of this procedure for the determination of structure types would require standardized structures and is a complementary tool to the methods introduced in this publication. Alternatively, one could think of classifying the crystal structures according the group–subgroup relationship, and families of structures types could be identified systematically (Megaw, 1973; Bärnighausen, 1980; *International Tables for Crystallography*, 2004). In such a scheme, the structure type with the highest symmetry would be the aristotype and by deleting some symmetry elements one would arrive at the hettotypes.⁶ Clearly the systematic treatments of these topics are a very interesting piece of work for the future, but are beyond the scope of the current work.

For the zeolite-type structures, the group–subgroup relations in the form of Bärnighausen trees have already been published by Baur & Fischer (2000–2006).

APPENDIX A

Searching for structure types in the PC version of *FindIt* and the Web version of the ICSD

A1. FindIt

To search, for example, for all the entries of the ' NaCl '-type, one has to proceed as follows.

1. Within the Reference tab click onto the 'STD Remarks/description' radio button.
2. Select 'TYP' from the pop-up menu on the right.
3. Switch to the 'Free text' radio button.
4. Enter the name of the type (here ' NaCl ' without quotes) into the upcoming 'Additional remarks' field on the right. Enter ' NaCl ' but not ' ClNa ' or ' halite ' because the inclusion of additional or alias names for structure types, which can be searched for, is planned for the future and will certainly require more time in production and development.
5. Click 'Search' button.

After the search, the name for a structure type appears in the comments sections of the extended output for a given crystal structure right after the keyword 'Structure type'.

In order to search e.g. for all the prototype structures present in the database do the following.

1. Within the Reference tab click onto the 'STD Remarks/description' radio button.
2. Select 'STP' from the pop-up menu on the right.
6. Click 'Search' button.

For *FindIt* 2007-01, this results in 1600 prototype entries, for which additional information about the structure type is given,

⁶ During our work, we noticed that no accepted name exists for the set of all members of a given structure type. In some cases, the aristotype (e.g. CaTiO_3 for the perovskite family) is mentioned.

e.g. the atomic environments (AE) of each atom (except for H atoms). A prototype entry is marked by the keyword 'Structure type prototype' within the comments section of the output.

A2. Web version

In the web version, all 'NaCl'-type entries can be searched using the field 'ANX/Pearson/S.Type' by entering 'T = NaCl' (without quotes).

We thank David Brown (McMaster University, Canada) for proofreading the first draft and polishing the English. Alexander Hannemann (FIZ Karlsruhe, Germany) gave us many helpful comments during the preparation of this manuscript.

References

- Baerlocher, Ch., Meier, W. M. & Olson, D. H. (2001). *Atlas of Zeolite Framework Types*. Amsterdam: Elsevier.
- Bärnighausen, A. (1980). *Commun. Math. Chem.* **9**, 139–175.
- Baur, W. H. & Fischer, R. X. (2000–2006). *Landolt-Börnstein, Numerical Data and Functional Relationships in Science & Technology, Group IV: Physical Chemistry*, Vol. 14, *Microporous & other Framework Materials with Zeolite-Type Structures*. Subvol. B: *Zeolite-Type Crystal Structures & their Chemistry. Zeolite Structure Codes ABW to CZP* (2000). Subvol. C: *Zeolite Structure Codes DAC to LOV* (2002). Subvol. D: *Zeolite Structure Codes LTA to RHO* (2006). Subvol. E will contain the remaining codes (in the press). Berlin: Springer-Verlag.
- Bayliss, P. (2000). *Powder Diffr.* **15**, 217.
- Bergerhoff, G., Berndt, M., Brandenburg, K. & Degen, T. (1999). *Acta Cryst.* **B55**, 147–156.
- Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. (1983). *J. Chem. Inf. Comput. Sci.* **23**, 66–69.
- Burzlaff, H. & Malinovsky, Y. (1997). *Acta Cryst.* **A53**, 217–224.
- Gelato, L. M. & Parthé, E. (1987). *J. Appl. Cryst.* **20**, 139–143.
- ICSD (2007). ICSD is available at FIZ Karlsruhe at <http://www.fiz-karlsruhe.de/icsd.html> or <http://icsdweb.fiz-karlsruhe.de>. Currently the ICSD database contains about 100 000 entries of fully determined inorganic crystal structures.
- International Tables for Crystallography* (2004). Vol. A1, edited by H. Wondratschek & U. Müller. Dordrecht: Kluwer Academic Publishers.
- Lima-de-Faria, J., Hellner, E., Liebau, F., Makovicky, E. & Parthé, E. (1990). *Acta Cryst.* **A46**, 1–11.
- Megaw, H. (1973). *Crystal Structures: a Working Approach*. Philadelphia: W. B. Saunders.
- Parthé, E. & Gelato, L. M. (1984). *Acta Cryst.* **A40**, 169–183.
- Parthé, E. & Gelato, L. M. (1985). *Acta Cryst.* **A41**, 142–151.
- Parthé, E., Gelato, L. M., Chabot, B., Penzo, M., Cenzual, K. & Gladyshevskii, R. (1993, 1994). *Gmelin Handbook of Inorganic & Organometallic Chemistry, TYPIX – Standardized Data & Crystal Chemical Characterization of Inorganic Structure Types* (updated as the computer program TYPIX1995).
- Reese, G., Yarger, R. J., King, T. & Williams, H. E. (2002). *Managing and Using MySQL*, 2nd ed. Sebastopol, CA: O'Reilly.
- Villars, P. & Calvert, L. D. (1991). *Pearson's Handbook of Crystallographic Data for Intermetallic Phases* (4 volumes). Materials Park, OH: ASM International.
- Villars, P. & Calvert, L. D. (1997). *Pearson's Handbook*, Desk Edition (2 volumes). Materials Park, OH: ASM International.
- Villars, P. & Cenzual, K. (2004). *Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, Group III: Condensed Matter*, Vol. 43, *Crystal Structures of Inorganic Compounds*. Subvol. A: *Structure Types*, Part 1 (Space groups 230–219). Berlin: Springer Verlag.