Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 14 October 2002
Accepted 14 February 2003

Variations in abundance of known crystalline compounds as a function of lattice constants

Benjamin Constant[a] and Paul J. Shlichta[b]*

[a]EF International School of English, Olympia, WA, USA, and [b]Crystal Research, 9112 Tucker Road SE, Olympia, WA 98513, USA. Correspondence e-mail: query@crystal-research.com

Preliminary plots of the abundance of inorganic compounds as a function of their crystallographic parameters disclosed sharp peaks. A plot of cubic compounds as a function of the unit-cell length yielded numerous peaks which appeared to correspond to sets of structure types. For hexagonal compounds as a function of *c*/*a*, there were two peaks at about 1.0 and 1.6. The abundance of tetragonal compounds as a function of *c*/*a* yielded numerous peaks which did not obviously correspond to any specific structure types.

One tends to assume that the abundance of known inorganic crystalline compounds, as exemplified by such databases as NIST's Crystal Data (NIST, 1997) or Gmelin's Inorganic Crystal Structure Database (ICSD, 2002), is a more or less monotonic function of their unit-cell lattice constants. However, during a series of exhaustive searches for lattice matches to various proteins, superconductors and semiconductors (Shlichta, 1997; Shlichta & Dowty, 2003), it was noted that, for a given mismatch tolerance, the number of matches would vary with the choice of target compound (*i.e.* with different unit-cell dimensions) by an order of magnitude or more. We therefore decided to undertake a preliminary census of inorganic compound abundance as a function of lattice parameters.

For our survey, we chose Volume 2 (*Inorganic Compounds*) of the 1973 3rd edition of *Crystal Data Determinative Tables* (Donnay & Ondik, 1973). This book version has only about 24 000 entries, as compared to the 150 000 'inorganic' entries in the latest Release J CD-ROM version (NIST, 1997) published in 1997; however, the later edition gives rise to several problems which hindered its use. The book version was amenable to hand counting, with exclusion, by inspection, of duplicate entries. In contrast, use of the CD-ROM version would have required the development of special software, for which the exclusion of duplicate entries would have been a difficult problem. Moreover, the CD-ROM version uses four categories – organic, inorganic, alloys and minerals – with many entries classified under two or more categories, making the avoidance of duplication particularly difficult.

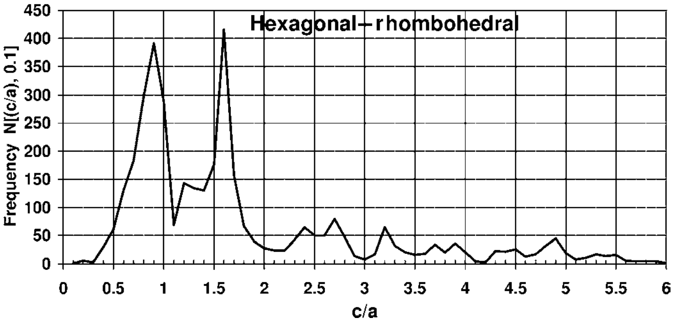

Figure 1
Frequency plot for hexagonal–rhombohedral compounds.

We therefore chose to use the 3rd edition to make three preliminary sample counts: (1) hexagonal and rhombohedral compounds as a function of *c*/*a*; (2) tetragonal compounds as a function of *c*/*a*; and (3) cubic compounds as a function of *a* (in Å). In each case, the 'frequency' (number of compounds within each 0.1 or 0.05 range of the parameter) was entered as a function of the minimum value of the range. Thus, $N[P, \Delta P]$, the number of different compounds between the parameter values of P and $(P + \Delta P)$, was plotted as a function of P as shown in Figs. 1–3.

The hexagonal–rhombohedral case (Fig. 1) behaved more or less as we had expected, with a broad peak at about $c/a = 0.9$ and a second at about 1.6. But the tetragonal and cubic plots (Figs. 2 and 3) had so

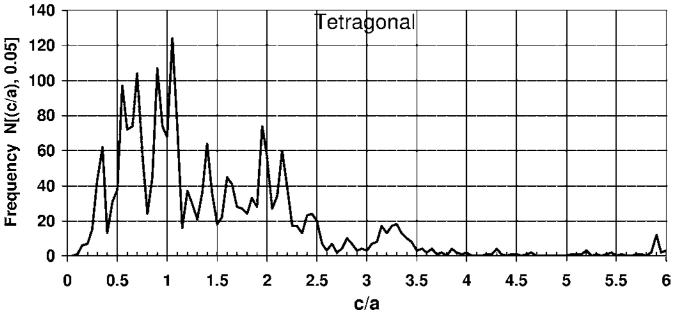

Figure 2
Frequency plot for tetragonal compounds.

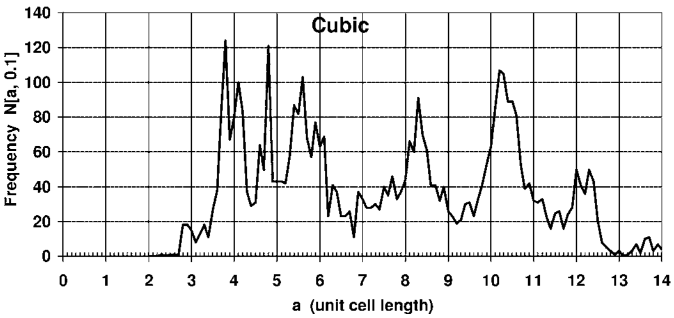

Figure 3
Frequency plot for cubic compounds.

Table 1
Prominent structure types at cubic frequency peaks.

a peak (Å)	Frequency	Prominent structure or composition types
3.8	125	CsCl (*B*2) alloys, perovskites (*E*2_1), tungstates
4.1	100	*B*2 and $AuCu_3$ (*L*1_2) alloys, hexaborides (*D*2_1), NaCl (*B*1) oxides, perovskites (*E*2_1)
4.8	120	NaCl (*B*1) oxides and nitrides, $AuCu_3$ (*L*1_2) alloys
5.4	88	Fluorite (*C*1) types, sphalerite (*B*3) types, pyrite (*C*2) types *etc.*
5.6	105	Fluorite (*C*1) types, sphalerite (*B*3) types, pyrite (*C*2) types *etc.*
5.9	75	Fluorite (*C*1) and NaCl (*B*1) alloys and compounds, $AuCu_3$ (*L*2_1) alloys
6.1	68	NaCl (*B*1) alloys and sulfides
8.1	65	Spinels (*H*1_1) and perovskite (*E*2_1) oxides
8.3	90	Spinels (*H*1_1) and perovskite (*E*2_1) oxides
10.2	108	Spinels (*H*1_1), pyrochlores *etc.*
10.3	105	Spinels (*H*1_1), pyrochlores *etc.*
10.5	90	Pyrochlores *etc.*
12.0	50	Garnet (*S*1_4) silicates, aluminates *etc.*
12.3	50	Garnet (*S*1_4) silicates, aluminates, gallates *etc.*

many sharp peaks that the former had to be replotted at $\Delta(c/a) = 0.05$ intervals to resolve them.

The peaks of the cubic plot correlate to some degree with specific prominent structure types, as shown in Table 1. It should be noted, however, that, since each frequency peak is composed of several structure types, the peak for any one structure type might be at a different *a* value.

One is inclined to assume that the hexagonal–rhombohedral peak at $c/a \simeq 1.6$ includes hexagonal-close-packed alloys and compounds based on hexagonal-close-packed anion packing and that the $c/a \simeq 0.9$ peak corresponds to hexagonal-layer sandwich structures such as AlB_2. However, we were unable to verify this since most of the entries did not include structure types.

We found no structure-type correlations to the tetragonal peaks except that the $c/a = 0.90$ peak included many chromates and vanadates. Therefore, we do not as yet have a plausible explanation for the surprising irregularity of this frequency plot.

It should be noted that the inorganic entries in even the newest CD-ROM versions of *Crystal Data* (*i.e.* about 150 000) represent only a tiny fraction of the billions of inorganic compounds that are estimated to exist (Shlichta, 2004). Therefore, one might argue that the results reported here have doubtful relevance to the abundance of inorganic compounds in general but only to those which, because of their simplicity and/or stability, were easy to prepare and/or happened to have been of interest to a particular investigator. (For example, the peak at the extreme right of Fig. 2 is due entirely to work on the rare-earth tritellurides by one laboratory.) But even so, such data may have predictive significance. It therefore appears worthwhile to continue such 'census taking', using the most recent databases and a wider variety of plots. For example, three-dimensional plots of frequency *versus* absolute values of *a* and *c* for tetragonal and hexagonal–rhombohedral compounds might determine whether the peaks correspond to modes of packing of the more common anions.

References

Donnay, J. D. H. & Ondik, H. M. (1973). *Crystal Data Determinative Tables*, 3rd ed., Vol. 2, *Inorganic Compounds.* US Department of Commerce, National Bureau of Standards and Joint Committee on Powder Diffraction Standards, Washington, DC, USA.

ICSD (2002). Gmelin Inorganic Crystal Structure Database website at http://barns.ill.Fr/dif/icsd/index.html.

NIST (1997). National Institute of Standards and Technology website at http://www.nist.gov/srd/nist3.htm.

Shlichta, P. J. (1997). *J. Cryst. Growth*, **174**, 480–486.

Shlichta, P. J. (2004). *Progress in Crystal Growth and Characterization of Materials.* In preparation.

Shlichta, P. J. & Dowty, E. (2003). *Thin Solid Films.* Submitted.