ELSEVIER

Journal of Alloys and Compounds 279 (1998) 1–7

Journal of ALLOYS AND COMPOUNDS

The Linus Pauling file (LPF) and its application to materials design

P. Villars[a],*, N. Onodera[b], S. Iwata[c]

[a]MPDS, Materials Phases Data System, CH-6354 Vitznau, Switzerland
[b]JST, Japan Science and Technology Corporation, 5-3, Yonbancho, Tokyo 102, Japan
[c]RACE, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Tokyo 113, Japan

Abstract

The LPF is a database which consists of a data part, as well as a smart software part. The data part covers all non-organic (e.g., alloys, intermetallics, ceramics, minerals, etc.) ordered solid state materials (systems) and consists of structure, diffraction, constitution, intrinsic property and bibliographic information. In materials science, the crystal structure of a compound holds the key-position, and is governed by five elemental factors: size, atomic-number, electrochemical, valence-electron and angular valence-orbital. The relations between these factors and the crystal structure are very complex, but the many regularities found to date prove that these relations are not only of a qualitative but, as well, a quantitative nature with an acceptable accuracy. The authors are completely convinced that, in the future, if one wishes to find, efficiently and systematically, novel multinary materials, the process will involve linking large databases such as the LPF (information) in a clever way with already known principles (knowledge), and thus creating an information-knowledge system, preferably via international collaboration.

Keywords: Regularities; Materials design; Materials structure prediction; Materials database

1. Introduction

Data creation for the LPF is a long-term project co-funded by the Japan Science and Technology (JST) Corporation and MPDS in collaboration with the Research of Artifacts Center for Engineering (RACE). Commencing in 1996, the LPF is currently a comprehensive data organization and development activity but within the next 12 years is projected to become a yearly update project. The LPF will be available world-wide through on-line (via the INTERNET) as well as off-line (via CD-ROM and/or printed hardcopy) products. The LPF project goal is to have four groups of materials data (structure, diffraction, constitution, intrinsic property) as numerical, factual and image data under one computing environment.

With the help of an international, highly experienced group of scientists, the LPF will include experimentally determined and calculated data published from about 100 000 relevant publications which will contain at least 200 000 structure, diffraction and intrinsic property data entries as well as about 35 OOO constitution data entries (image) covering the world literature from 1900 to the present. In addition, value-added information will be given by the section-editors, such as, e.g. critical comments, calculated powder patterns (LAZY PULVERIX and other), fully standardized structure data (STRUCTURE TIDY) etc., as well as going through sophisticated computer-aided checking procedures. Besides newly created data entries taken from journal articles, JST Crystal Structure (JST [1]), parts of CRYSTMET (MPDS formerly CIST/NRCC [2]), Pearson's Handbook [3], Structure Type Atlas [4], and Ternary Alloy Phase Diagram Handbook [5] (all published by ASM International) will be incorporated after being fully adjusted to the new LPF database specifications.

In materials science, the crystal structure of a compound holds the key-position, and is governed by five factors: size factor, atomic-number factor, electrochemical factor, valence-electron factor and angular valence-orbital factor. The relations between these factors and the crystal structure is very complex, but the many regularities found to date prove that these relations are not only of a qualitative, but a quantitative nature with an acceptable accuracy to establish compound-formation, atomic environment, and structure map principles.

The fundamentals of the constitution of a material system are determined by the crystal structure of its compounds and its phase diagram. Knowing these fun-

---

*Corresponding author.

0925-8388/98/$19.00 © 1998 Published by Elsevier Science S.A. All rights reserved.
PII: S0925-8388(98)00605-7

damentals enable scientists to solve many problems in materials science, and therefore it is important to have easy access to the experimentally determined data, especially the experimental work to determine such information is very time- and cost-intensive.

Looking at the research activities in the last 20 years, with respect to the number of systems investigated, grouped according to binary, ternary, and quaternary systems, a shift in the ongoing research from binary to ternary and quaternary systems is clearly seen. The research field of multinary systems is a huge field for potential novel materials with optimized known physical properties (e.g. tP68 BFe14Nd2-type permanent magnets and high temperature superconductors) as well as 'new' physical properties, as the recent past has proven with the quasicrystals.

The trend in moving from binaries to ternary and quaternary systems presents many additional 'difficult-to-handle' problems involving the number of experimental variables, i.e. the chemical elements, their possible combinations, their concentrations, the temperature, and the pressure. Of enormous consequence, from the practical point of view, is the variety of combinations possible involving the chemical elements, and the combinatorial increase due to a much larger and more available concentration range. Unfortunately, the number of available experimentally determined data is inversely proportional to these opportunities. We have a relatively robust database for binaries, but a sparse database for ternaries/quaternaries, and almost no data for higher systems. In the binary systems, taking 100 chemical elements into account, there exist 4950 binary systems, a large number, but still, with united, coordinated international research efforts, all these systems could be experimentally investigated by the end of this century.

Without good guidelines, it is a hopeless situation to search systematically for novel materials with an adequate success rate in multinary systems. Therefore, the only practicable way to go is, with the help of the robust binary database, to find regularities, such as laws, rules, principles, factors, tendencies, and patterns valid within the binaries, and then to extend these to the ternaries and quaternaries in such a way that the binaries and multinaries can be treated together; otherwise the regularities would be based on too few data sets and therefore would not be trustworthy. In addition, the regularities should show an accuracy clearly above randomness to be of practical use. To increase the efficiency in the successful search for 'new' compounds, the main efforts should be directed towards creating an internationally accessible information-knowledge system incorporating all data-bases (experimentally determined facts) as well as generally valid principles and the 'highest-quality' regularities (knowledge), and thus creating a workbench in materials science having suitable interoperability with other systems of different scopes, functions, etc.

The combination of the experience and intuition of the experimentalist together with easy access to experimental determined data in LPF or other major database forms, as well as access to the envisioned information-knowledge system, will very much help to coordinate world research activities. Furthermore, the availability of data would reduce the number of duplicative efforts as well as increase the probability of investigating the most promising systems first, and avoid investigating new material systems in a random statistical sequence.

2. Strategy for discovery of factors governing crystal structure

Of interest to experimentalists are factors governing crystal structures which can be used to reduce or classify material systems by their atomic property expressions (APEs) associated with their constituent chemical elements. Regularities of real practical value are those which systematize a large group of data with an accuracy in the range of at least 95%. For the not yet investigated systems, one goes in exactly the reverse direction, assuming the validity of the regularity found for a well-defined group of data. The larger the materials database considered, the more trustworthy is a prediction based on such regularities. To make predictions, one starts from the chemical elements, their concentrations and atomic properties (APs), and calculates the APEs for the system of interest. Noteworthy is the fact that APs are only of practical interest where they are known with adequate accuracy. From the APEs, in the context of a learned regularity, one can predict how new systems of interest will behave, i.e. in terms of their APs, with an accuracy of at least 95%.

In principle, it would be sufficient to use as input only the atomic numbers (ANs) and the compositions of the compounds of the systems under consideration. In 1956, Slater [6] once made a comment as follows:

> *I don't understand why you materials scientists are so busy in working out experimentally the constitution* [*crystal structure and phase diagram*] *of multinary systems*. *We know the structure of the atoms* [*needing only AN*], *we have the laws of quantum mechanics, and we have electronic calculation machines*, *which can solve the pertinent equation rather quickly*!

Some 35 years later Chelikowsky [7] wrote, in an excellent review, the following:

> *Although the interactions in compounds are well understood, it is not an easy task to evaluate the total energy of solids*, *even at absolute zero*. *As the energy of an isolated atom is in the order of about* 10^6 *eV, but the cohesive energy only in the order of about* 1–10 *eV/atom, one must have a method that is accurate to one part in* 10^6, *or better*.

The other fact that greatly complicates evaluating the cohesive energy by theoretical methods is the number of particles involved. Given that a macroscopic solid may contain 10^{23} nuclei and electrons, it is impossible to determine the total energy of the crystal structure without some approximations. Within the last 15 years, two advances have made it possible to predict the cohesive energy of solids by numerical solutions of the quantum-mechanical equations of motion, i.e. solving the Schrodinger equation via the invention of high-speed supercomputers and the device of one-electron potentials, which greatly simplifies many-body interactions.

The accuracy of these computations is usually not at the same level as experiments. Nonetheless, it is now possible, for chemical elements and simple compounds, to predict whether a given crystal structure is most stable at 0 K and 1 atm. We have to stress that the crystal structure, as expressed using structure types or prototypes, which from Pearson's Desk Edition by Villars [8] we know of at least 4000 different ones exist, have to be given as input for first-principles calculations for each potential compound. Assuming that a potential compound crystallizes in one of the 4000 known crystal structures and knowing its nominal stoichiometry, AB, AB_2, etc., may still, in some cases, require a few hundred first-principles calculations for each potential compound. Even with supercomputers, this is not yet computationally tractable. But the larger problem is the calculation accuracy of the cohesive energies for those few hundred atoms, whose differences would need be in the range of one part in 10^9 to determine the most stable crystal structure.

The complexity of the above-mentioned problem shows that one cannot expect that within the next decade the constitution (crystal structure and phase diagram) of multinary systems will be calculated from first principles. Meanwhile it is therefore sensible to adopt semi-empirical approaches, based on experimentally determined data, to search for reliable regularities. At the moment it is impossible to start only from the atomic number (AN), and thus, it is obvious that one attempt to determine which other APs of the chemical elements are needed to describe the compound formation behavior and predict other unknown APs.

Villars [9] conducted a survey of 53 different APs as a function of the AN (182 variables in all when the different methods of determination are taken into account), and it was found that there were only five main factors which are defined below. The results are best appreciated when depicted as AP versus AN plots, as shown in Fig. 1, for the pseudopotential radii R (after Zunger [10]) for s, p. and d elements. The f elements have been left out because in most cases only incomplete data are available. In all but a few cases (19 out of 182) very regular symmetric patterns were obtained in such plots; this means a regular behavior, both within a group with increasing quantum number (QN) as well as along a period with increasing AN. In these diagrams the chemical elements are ordered by increasing quantum number. This special 'AN' scale was used instead of a linearly increasing AN scale because it gives units of comparable size for distance between atoms within any group as well as along any period. The scale chosen for comparison does not affect the number of factors (groups), of course; only the appearance of the pattern changes. It should be stressed that the equivalence of APs belonging to the same group is of a qualitative, not a quantitative nature. Adherence to one or another of the five factors is very obvious. In Fig. 2 two APs belonging to the same factor plotted against each other, the cohesive energy versus melting point T of the chemical elements are shown. At a first approximation, a linear dependence may be seen to exist.The five factors (groups) are:

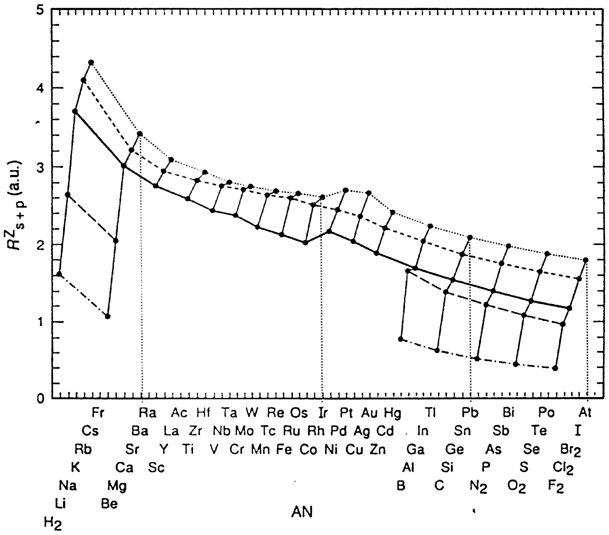

Fig. 1. An atomic property, AP, versus a function of atomic number, AN, plot of the pseudopotential radii after Zunger, R^Z_{s+p}. In such diagrams the chemical elements are ordered by increasing group number and each group is ordered by increasing quantum number. This special atomic-number (AN) scale was used instead of a linearly increasing atomic number scale because it gives units of comparable size for distance between the atoms within any group as well as along any period. The scale chosen for the comparison, of course, does not affect the number of factors (groups), only the appearance of the pattern changes.

- Size factor
- Atomic-number factor
- Cohesive-energy factor
- Electrochemical factor
- Valence-electron factor

3. Atomic environment approach

Given the traditional crystal structure definition based on the space-group theory, there exist at least 4000 different crystal structure types or prototypes; this number increases at a rate of about 4% per year. From the practical point of view, this high number makes it nearly impossible to find rigorously valid, general regularities within crystal struc-

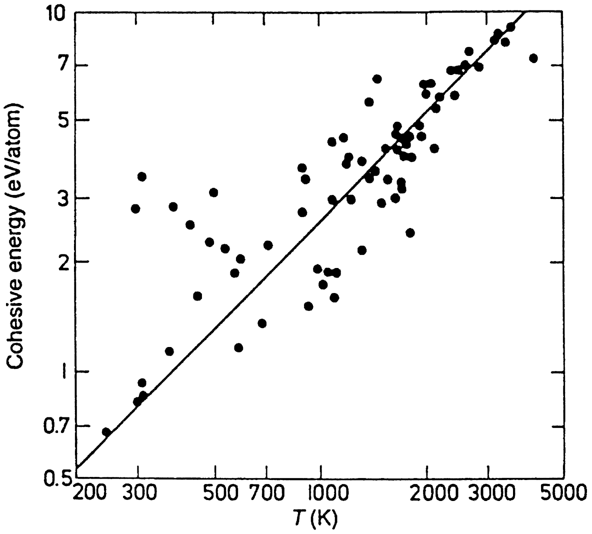

Fig. 2. Plot for the chemical elements of the relationship of two atomic properties, APs, belonging to the cohesive-energy factor (group), i.e. cohesive energy and melting point *T*.

tures. The atomic-environment approach described herein drastically reduces the number of distinctly different crystal structures by focusing on gross crystal structure features.

The lattice symmetry and space group of compounds are important for certain physical properties such as piezoelectricity and ferroelectricity, but for crystal chemistry they are less important in that minor changes in the position of the atoms in a crystal structure can often reduce its symmetry, therefore leading to a new mathematical description and a 'new' crystal structure. The resulting deformation of the atomic environment (AE) (for the cases where those atoms form a convex polyhedron, this is also called the coordination polyhedron), however, is often negligible, so that the AE does not change significantly.

The aim of the AE concept is to define the AB as clearly as possible, so that we will ultimately be able to group these environments into distinct atomic environment types (AETs). With the help of the maximum-gap rule after Brunner and Schwarzenbach [11], we define the AEs. All interatomic distances (d) between a central atom and its neighbors are plotted in a histogram as shown in Fig. 3(a). The height of the bars is proportional to the number of neighbors (n), and it is convenient to express all distances (d) relative to the shortest distance (d_{min}). In most cases a clear gap is revealed, as can be seen in Fig. 3(b) for *cF*4 Cu crystal structure. For cases where no clear gap is revealed, clearly defined rules are given in Daams and Villars [12]. It is important to stress that the method for defining the AE is not important; as long as one finds all the 'idealized', distinctly different AETs realized in compounds. It is also important to realize that the AETs are not isolated building units, as some crystal structures consist of interpenetrating AETs.

Daams and Villars [13–15] investigated all cubic, rhombohedral, hexagonal and tetragonal intermetallic compounds and found that 14 AETs given in Fig. 4 represent over 80% of all positions. In Fig. 5 their frequency within cubic crystal structures is shown. The tetrahedron, octahedron, cube, icosahedron, cubooctahedron, and rhombic dodecahedron are strongly preferred within the 14 AETs. Preliminary results in the remaining crystal systems indicate that all 4000 crystal structures can be described with the addition of less than 20 frequently occurring AETs, and 40–60 rarely occurring AETs. This enables us to reduce

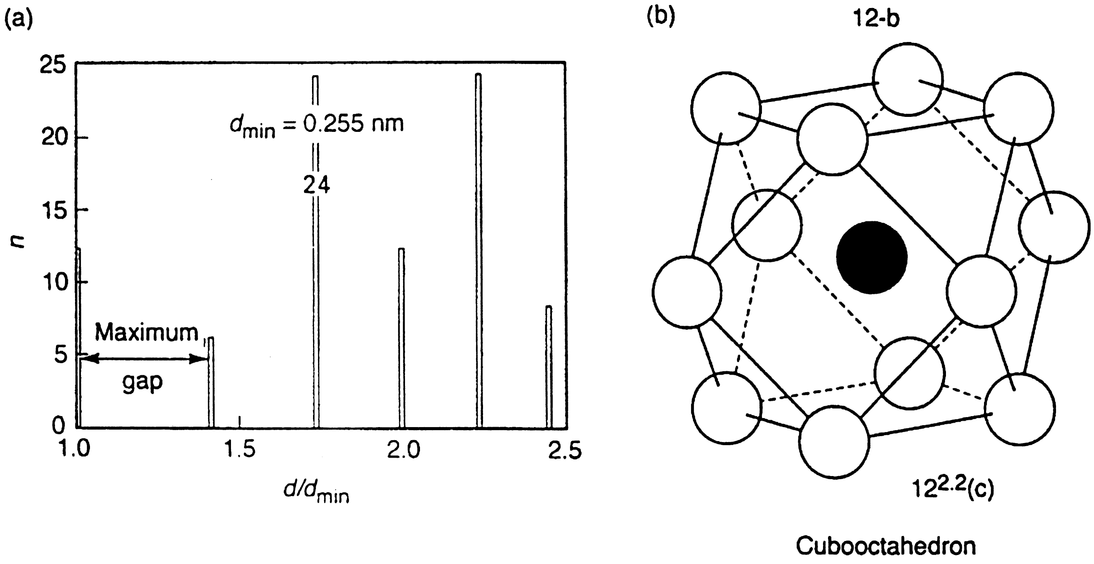

Fig. 3. (a) A typical example of a next-neighbor histogram and (b) the atomic environment, AE, for a cubooctahedron (12-b). The example is constructed with the atoms before the maximum gap in this histogram for the *cF*4 Cu crystal structure. $12^{2.2}$(c) is the polyhedron code, based on the number of triangles, squares, pentagons, hexagons, etc., that join each other in the different vertices. The code gives the number of equivalent vertices (with the number of faces in the above-mentioned sequence, separated by a period) as an exponent.

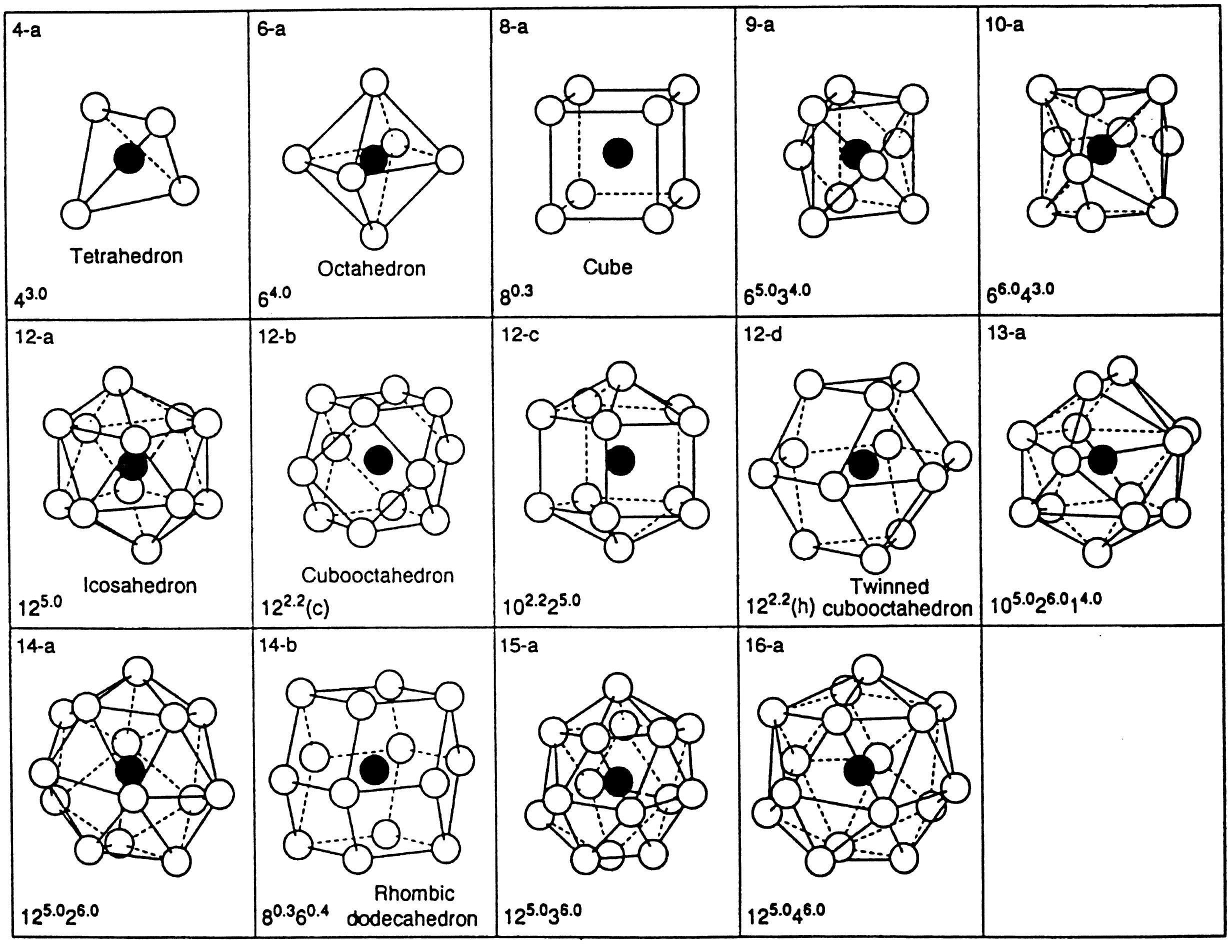

Fig. 4. The 14 most frequently occurring atomic environment types (AETs) with their codens and labels. The label is based on the coordination number and a small letter to differentiate AETs with identical coordination number. The code is based on the number of triangles, squares, pentagons, and hexagons that join each other in the different vertices. The code gives the number of equivalent vertices (with the number of faces of each type, in the above-mentioned sequence, separated by a period) as an exponent.

the 4000 crystal structures to about 400 coordination types, i.e. crystal structures having the same number and kinds of AETs. The results of this investigation strongly indicate that, in compounds, the short-range order is very dominant, and the long-range order is of much less importance.

4. Conclusions

In a comprehensive review [16] we have tried to summarize the most outstanding publications on regularities and gave special preference to the ones which included in their investigation large groups of well-defined data sets, in the range of hundreds to thousands of data sets. With the help of the regularities found semi-empirically within the experimentally established compounds, four factors are shown to govern crystal structure from the initially suggested five factors found with the help of the AP versus AN plots. When we consider the strong tendency of the atoms in crystal structures to realize one of these 14 different preferred AETs, we were not able to associate this tendency with any one of our five factors. Pettifor [17] has mentioned in a review about structure maps an additional factor, which he named the ‘Angular-valence-orbital factor’. The AB approach proves the importance of this factor on a very comprehensive scale.

Finally, we end up with the following five factors governing crystal structures:

- Size factor
- Atomic-number factor
- Electrochemical factor
- Valence-electron factor
- Angular-valence-orbital factor

The ‘cohesive-energy’ appears not to be a factor governing the crystal structure, but it does govern compound formation, particularly the inability of a compound to form. The relations between these factors and the crystal structures are very complex, but the many regularities found prove that these relations are not only of a qualitative nature, but already in many cases exhibit a quantitative nature with an

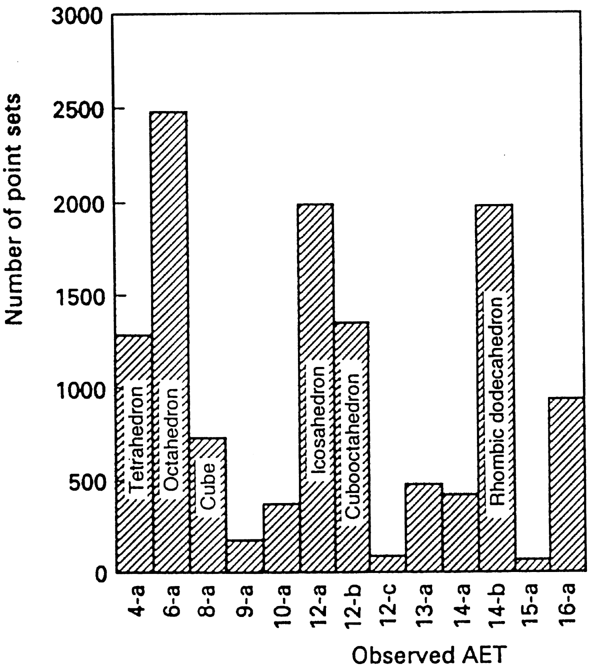

Fig. 5. A frequency plot of the 14 most frequently occurring AET's found in 5086 cubic intermetallic compounds.

acceptable accuracy. Forced by the almost unlimited number of available chemical element combinations in multinary systems, the authors are completely convinced that, if in the future one wishes to find efficiently and systematically, novel ternary (multinary) materials, it will only be possible by linking large databases like, e.g. LPF (information) with known regularities like, e.g. AET approach, which we highlighted in this publication as one example (knowledge), and thus creating an information-knowledge system in a clever way for advances in materials science. This information-knowledge system should be continually extended by including newly found 'highest-quality' regularities published in the world literature. The benefits will be:

1. Refinement of the already known regularities with exact definition of their validity and boundary conditions
2. Opportunity to search in a systematic way, with a high probability of success, for regularities until now missed
3. Verification of all regularities against the databases and derivation of an 'accuracy level' for each regularity
4. Reduction of the number of unwanted duplications of experimental work
5. Provision of a highly developed tool for the experimentalist working in materials science. This tool, together with the intuition and experience of the scientists, should reduce the 'unlimited' large number of potential systems to a limited, though large, number of systems. The result would be the synthesis of all potential novel materials with desired physical properties with highest probability of success
6. The interaction of the results of first-principles calculations will be very stimulating for the improvement of such an information-knowledge system and vice versa
7. Enormous economic benefits to those wishing to maintain their competitiveness in the area. This approach, combined with effective materials processing techniques, should show new results in materials science and engineering.

Savitskii [18] was one of the first to stress the need for information-knowledge systems in materials science. He was also the first [19] to develop, in that context, a database on the physical properties of ternary inorganic phases. Kiselyova [20] uses that database together with cybernetic-statistical methods to develop an information-knowledge system. Zhou et al. [21] developed such a system for retrieval and prediction of binary intermetallic compounds.

The need for an international information-knowledge system has been strongly evident for the last 10 years as most research activities have moved from binaries to multinaries, opening an unlimited 'reservoir' of potential novel materials. The main efforts should emphasize creating an internationally accessible information-knowledge system as a workbench with interoperability to all other databases of experimental facts. This workbench will need to be designed with internet type browser 'plug-ins' for interaction with other knowledge-based systems. These plug-ins will need to support various:

- first-principles modelling and simulation,
- access to the 'highest-quality' regularities (knowledge),
- data-mining/knowledge-discovery methods, and
- on-line use real world environments (laboratories, production lines, maintenance etc.) so as to make the system an international 'work' bench.

We would like to see the LPF positioned as one of the major information sources in such an international information-knowledge system, giving future materials scientists a powerful prediction tool. In such an open human network, the eyes of many may guide us toward an evolution, or possibly, a revolution in materials design.

References

[1] JST Crystal Data 1997, K. Osaki (Ed.), Japan Science and Technology Corporation, 5-3 Yonbancho, Tokyo 102. Japan.

[2] CRYSTMET 1997, P. Villars (Ed.), MPDS Materials Phases Data System, CH-6354 Vitznau, Switzerland.

[3] P. Villars, L.D. Calvert, Pearson's Handbook of Crystallographic Data for Intermetallic Phases, vols. 1–4, ASM International, Materials Park, OH, 1991.

[4] J.L.C. Daams, P. Villars, J.H.N. van Vucht, Atlas of Crystal Structures for Intermetallic Phases, vols. 1–4, ASM International, Materials Park, OH, 1991.

[5] P. Villars, A. Prince, H. Okamoto, Handbook of Ternary Alloy Phase Diagrams, vols. 1–10, ASM International, Materials Park, OH, 1995.
[6] J.C. Slater, Theory of Alloy Phases, American Society for Metals, Cleveland, OH, 1956.
[7] J.R. Chelikowsky, Encyclopedia of Applied Physics, vol. 3, VCH Publishers, Weinheim, Germany, 1991.
[8] P. Villars, Pearson's Desk Edition, Vols. 1 and 2, ASM International, Materials Park, OH, 1997.
[9] P. Villars, J. Less-Common Met. 92 (1983) 215.
[10] A. Zunger, in: M. O'Keeffe, A. Navrotsky (Eds.), Structure and Bonding in Crystals, Academic Press, New York, 1981.
[11] G.O. Brunner, D. Schwarzenbach, Z. Kristallogr. 133 (1971) 127.
[12] J.L.C. Daams, P. Villars, J.H.N. van Vucht, J. Alloys Comp. 182 (1992) 1.
[13] J.L.C. Daams, P. Villars, J. Alloys Comp. 197 (1993) 243.
[14] J.L.C. Daams, P. Villars, J. Alloys Comp. 215 (1994) 1.
[15] J.L.C. Daams, P. Villars, J. Alloys Comp. 1997, in press.
[16] P. Villars, in: J.H. Westbrook, R.L. Fleischer (Eds.), Intermetallic Compounds, John Wiley and Sons, New York, 1995, p. 227.
[17] D.G. Pettifor, Mater. Sci. Technol. 4 (1988) 675.
[18] E.M. Savitskii, in: O. Henkel (Ed.), Intermetallische Phasen, VEB Deutscher Verlag fUr Grundstoffindustrie, Leipzig, 1976.
[19] E.M. Savitskii, N.N. Kiselyova, B.N. Pischik, N.V. Kravchenko, M.S. Golikova, Sov. Phys. Dokl. 29(11) (1984) 971.
[20] N.N. Kiselyova, J. Alloys Comp. 197 (1993) 159.
[21] Z. Ban, J. Shiming, S. Jun, C. Nianyi, Acta Metall. Sin. 2B(6) (1989) 428.