論文

Discovery Regularities by Materials Databases
物質・材料データベースに基づく規則性の発見

Ying CHEN*　Shuichi IWATA*　Yasunori KANETA*
and
Pierre VILLARS†

陳 迎　岩田 修一　金田 保則　ピエル ビラス

With the spectacular development of information technology nowadays, the data mining approach based on well organized materials database with large amount of high quality data is becoming a powerful tool in materials design. PAULING FILE is a comprehensive materials database for all non-organic solid state materials, covering crystallographic data, phase diagrams, diffraction patterns and physical properties data published within the last 100 years. The newly released PAULING FILE, BINARIES EDITION contains about 28,000 crystal structure entries, 28,000 diffraction entries, 42,000 property data and 8,000 constitution entries and 8,000 images of phase diagram. Systematic searching and analyzing within this huge amount of data from various aspects have revealed the regularities and correlations implicit in materials which directly provide hints on candidate materials in preliminary stage. The regularities discovered in binary systems among substance structure, property and the atomic properties of its constitution are presented.

近年のめざましい情報技術開発に伴い，大量で高品質のデータを備えた物質・材料データベースに基づくデータ・知識発掘型のアプローチが，物質・材料設計に対する強力な手法となってきている．「ポーリングファイル」は，無機物質・材料についての包括的な物質・材料データベースであり，ここ 100 年間に発表された結晶構造データ，状態図，回折パターン，物性値データを格納している．現在公開されているの「ポーリングファイル二元系版」は，結晶構造に対して約 28,000 件，回折パターン 28,000 件，物性値 42,000 件，状態図データ・画像各 8,000 件を含んでいる．この大量のデータを様々な観点から系統的に検索・解析することにより，物質・材料群に内在する規則性や相関が示され，さらに物質・材料設計における初期段階としての物質・材料候補のヒントを得ることができる．ここでは，この二元系物質の物質構造，物性値，物質を構成する元素特性に対し，発見された規則性について述べる．

Keywords: Materials database, materials design, regularity, mapping
物質・材料データベース，材料設計，規則性，マッピング

1　Introduction

One of the most challenging tasks in material science is designing materials with certain atomic constitutions, which achieves some distinct properties. The procedure of the design is generally composed of several stages such as specification design to integrate needs, function design to specialize properties, structure design to achieve certain properties and process design to realize special structures from raw materials. Among many factors that determine the properties of materials, the intrinsic relationship between the structure, the properties of the material and its atomic constitution plays a key role on the physical basis. With the spectacular development of information technology and computational tech-

* School of Engineering, The University of Tokyo, Japan
Email:ying@q.t.u-tokyo.ac.jp
† Materials Phases Data System, Switzerland

niques, accompanying the progress of various experimental techniques, people have been dreaming of designing materials with specific properties from all possible atomic combinations by using powerful computational physics approaches and intelligent experimental techniques like combinatorial synthesis. However, the huge amount of possibilities to be treated is still a big problem in practice. To increase the efficiency and accelerate the speed of finding target atomic configuration from huge number of candidates, a data-driven approach is proposed based on the well-organized large-scale materials database. It is highly probable that the regularities and correlations discovered by data mining will provide a good guidance in the preliminary stage.

The PAULING FILE[1] is a comprehensive materials database for all non-organic solid state materials, covering crystallographic data, phase diagrams, diffraction patterns and physical properties data published within the last 100 years. The newly released PAULING FILE, Binaries Edition contains about 28,000 crystal structure entries, 28,000 diffraction entries, 42,000 property data and 8,000 constitution entries and 8,000 images of phase diagram, with the dynamic link among them, as well a design platform as tool for data mining and materials design. Based on a systematical analysis of these data, some important regularities in binary systems have been extracted, which could serve as an initial reference in setting strategy of theoretical calculations and selecting experiments.

2 Regularities in binary systems

By considering the huge bibliographic literatures in materials science, people might have an impression that scientists have investigated a large part of materials known. However, the statistics shows that the experimental knowledge available is only substantial for binary (approximately 70% of the binary systems has been fully or partially investigated). However, less than 5% of the ternary and less than 0.5% of the quaternary systems have been studied. Therefore, this is the main reason why we start to find knowledge in the binary systems. The other reason is that our previous study[2] has showed the results obtained on binaries could be extended to ternary, quaternary and multinary systems.

The substantial meaning of “regularities” can be interpreted as finding answers to the following four questions:

1) If a compound is formed or not when two elements meet?
2) Which compositions are realized in a binary forming system?
3) Which structures are formed for a compound?
4) What the physical property is realized for a compound in certain structure?

Mapping approach is utilized for answering these questions by mining the data in the PAULING FILE. The basic idea is to relate a kind of “compound property” to one or several parameters of the constituting elements and to implement “mapping” by taking these elemental parameters as the axes. The aim of mapping is to group data with certain property into clearly defined domains, so as to reveal the patterns of various dependence of the property on the coordinates. There are two key points in mapping: how to select atomic properties as the coordinates of the map and how to define the domains according to the purpose. Two con-

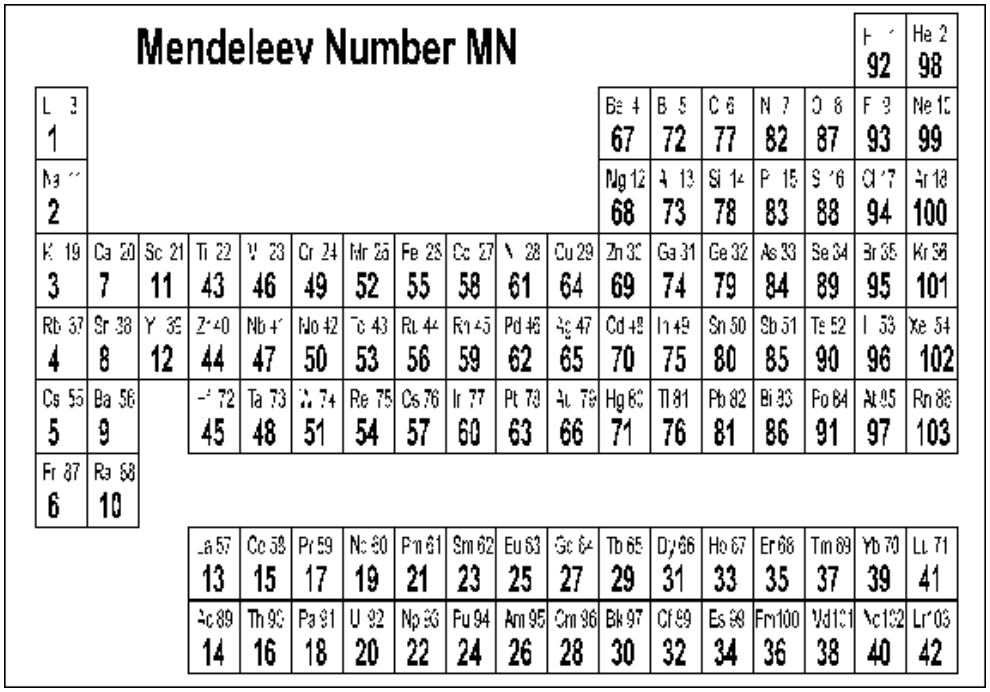

Fig. 1 Mendeleev numbers (MN) used in the present investigation.

cepts are introduced to achieve a successful separation of large amount of known phases into different, non-overlapping domains.

Elemental parameters: Based on the previous work[2], from 56 elemental properties such as atomic number, density, atomic radius, electro-negativity, etc., Mendeleev number (MN)[3], is selected as the coordinates of the structure map. Mendeleev number is an ordering number attributed to each chemical element in the Periodic Table basically along the direction of the "group". The values of Mendeleev number for all elements are showed in Fig. 1.

Atomic environment type (AET): In order to reduce the large number of possibilities between elemental properties and conventional structure types, which are based on space group theory with 2,300 structure types for binary system, the atomic environment type (AET)[4] that is less in detail but still contains enough information to distinguish main difference of structures is introduced as a generalized structural classification. The AET of a structure is defined by the coordinate polyhedral surrounding certain atom in a structure, and named in the number of coordinates atoms. The total *AETs* for binary compounds is about 150, and only 31 are the most populous ones (see Fig. 2).

The results of this kind of semi-empirical analysis on binary data are summarized briefly in the following. One can refer to related papers for the detailed description[2][5].

2.1 Compounds formation

A very fundamental question in metal chemistry is, which binary systems form compounds and which do not? A chemical system is defined as a compound *forming*

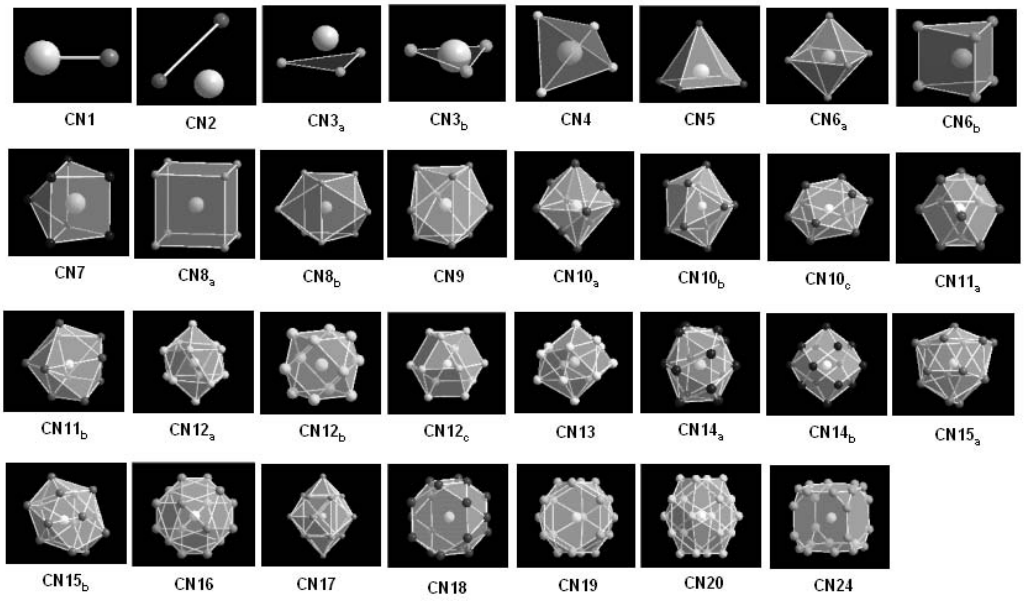

Fig. 2 The 31 most populous Atomic Environment Types (AETs) for binary compounds. The AET is named in the coordinate number (CN) of atoms to forming the coordinate polyhedron.

system if, at atmospheric pressure, it contains at least one phase that is separated from the terminal solid solutions of the constituting elements by a two-phase region. A system that does not fulfill this criterion is a *non-forming system*. Fig. 3 is a two dimensional formation map with $MN(A)$ and $MN(B)$ as the coordinates. The gray and white squares stand for the systems having at least one set of crystal data and have no crystal data, respectively. The cross symbols stand for compound *non-forming sysytems*. It has been shown that 30% of the systems with information cannot form compounds, while 70% of these form, and the domains for *forming* and *non-forming* are clearly defined. What is very remarkable is that there are almost no compounds in the upper-left hand corner of the plot, which implies an important fact that between two elements with $MN \leqq 50$, there is a tendency to not form the compounds. This conclusion can be taken as one of main criteria to screen target material systems from large amount of candidates in the materials design effort. While looking at the details of the *non-forming substances*, it is found that 66% systems whose constituent element A and B are iso-structural can not form compound, whereas, 90% non-iso-structural systems can form compound, which leads to the conclusion that elements have a tendency to keep the structure of their pure form as long as possible.

2.2 Composition of compounds

In order to investigate the compositions to be realized in a compound forming system, we list phase diagrams of binary systems between Mg and all other elements in the Periodic Table, as showed in Fig. 4, as an

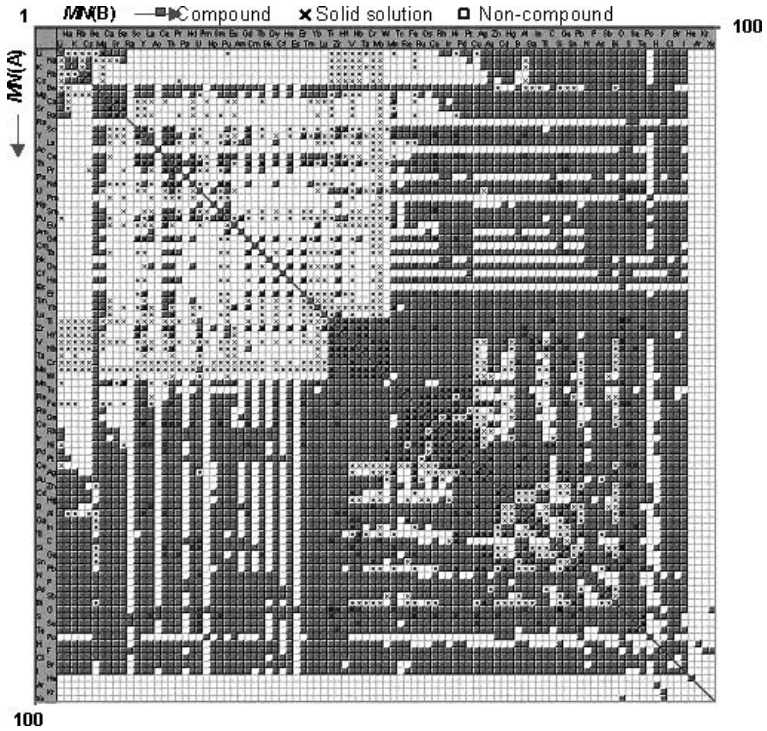

Fig. 3 Two-dimensional compound formation map with $MN(A)$ vs. $MN(B)$ of binary systems. The gray and white squares stand for the systems having at least one set of crystal data and have no crystal data respectively, and the cross symbols stand for *non-forming* substances.

example. This plot shows clearly the trend of Mg-X binary systems with respect to the location of element X in the Periodic Table. The obvious characters of phase diagrams between Mg and s-, p- and d- elements can be observed. It is noticeable that almost no compounds are formed between Mg and the transition metals in the left-middle part of the Periodic Table. Further observation shows that all possible compositions and *AETs* seem having selection with respect to an element X. For elements in the right hand side of the Periodic Table, only having concentration $\leqq 50\%$, and the CN12 and CN14 *AETs* dominate. Whereas, for left side transition metal elements X, almost all range of concentrations are possible and the structures are in complicated situation; with p-elements, only CN4 and CN6 appear in a limited concentration range.

Analogous analysis for other fixed elements also show obvious features with respect to the fixed elements, and one can understand these patterns are closely related to the interaction types between atoms.

2.3 Structures of compounds

A systematical analysis of the structures of compounds has been carried out. The $MN(A)$ vs. $MN(B)$ maps achieved relatively good separation of the different chemical systems into distinct AET and *non-forming* domains.

Further attempts has been made to define the stability criteria for each AET. Sev-

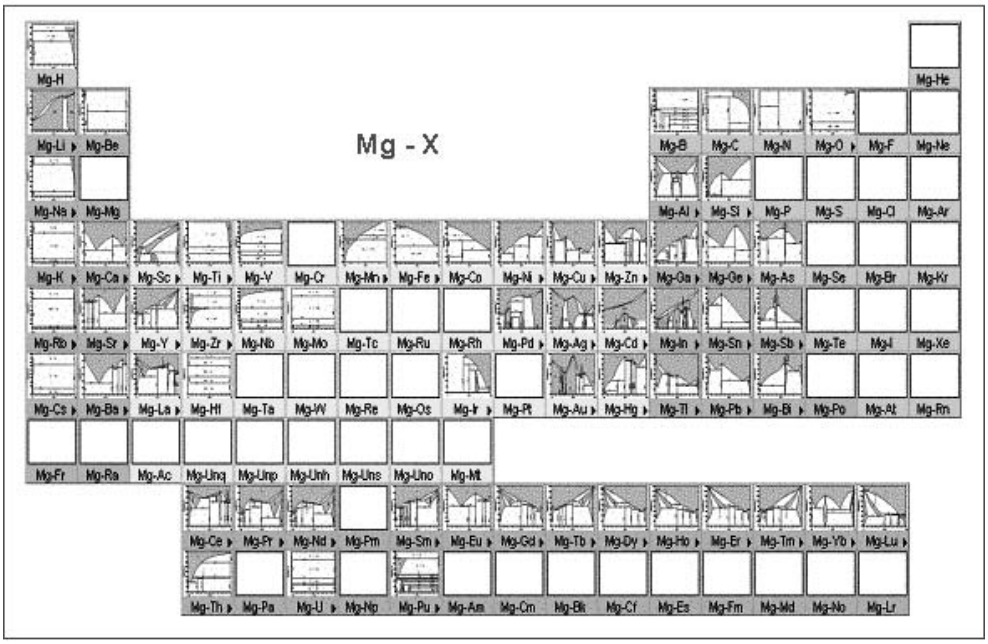

Fig. 4　Phase diagrams of Mg-X binary systems.

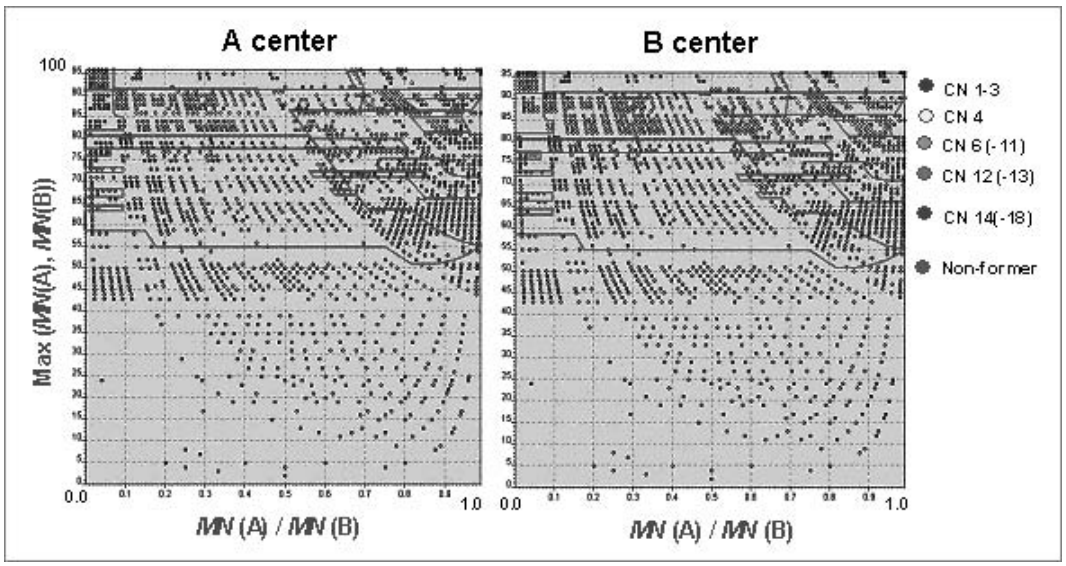

Fig. 5　AET stability map MN_{max} vs. MN_{min}/MN_{max} for AB compounds in case of A, B element as the center of AET, respectively.

eral simple mathematical operators are applied to correlate elemental parameters of the constituting elements with compound parameters. It is found that MN_{max} vs. MN_{max}/MN_{min} are optimal coordinates for 1 : 1 binary systems, by which the AET

stability diagrams can be constructed with rather high accuracy. Fig. 5 is the structure map for 1 : 1 binary compounds. Since a binary compound is consists of two kinds of atom, atom A and atom B, we need to investigate the atomic environment types of atom A as center and B as center respectively. The separation of the different categories of chemical systems into distinct stability domains is very good, and the formers domain can be sub-divided into different AET domains, clearly moving with increasing max(MN) from non-formers to AET with CN15 towards to AET with CN1. The ordering according to the maximum Mendeleev number MN_{max} (y-axis) has a very strong separating effect, the ratio MN_{min}/MN_{max} (x-axis) is, however, necessary to achieve a good separationd.

Meanwhile, this map provides the conditions of stability of AET. This would be very helpful in prediction of the structure for a not yet investigated system by calculating the appropriate parameters and see if the conditions for the formation of a particular AET are fulfilled.

2.4 Physical properties

Mapping approach can be also applied to the physical properties. Fig. 6 shows the distribution of binary systems according to their values of molar enthalpy of formation with $MN(A)$ vs. $MN(B)$. Different symbols mark different range of values of enthalpy of formation. It is noticed that the systems with high formation enthalpy are allocated in the area near the edges of the map, which implies that those systems with one element in large MN number have rather high formation enthalpy. Such kind of mapping is helpful directly when searching new candidates for expected values of property in the preliminary stage. However, this application requires large amount of highly reliable properties data, which should be an essential demand to the materials database nowadays.

3 Combination of theoretical calculation in database approach

Through a thorough search conducted on all available data in several types of discovery spaces, one will get a global picture on the materials world, as to where we are? It should be very helpful to get hints for designing new materials. However, we should realize that working alone on the data sometimes might have problems to supply the satisfactory solution. The most strategic way in the research on materials design is to combine the discovery tool with calculations based on cleverly constructed physical models.

3.1 Theoretical structure map

Since the materials data available are far from filling the discovery space, there should be some uncertainty in determining the boundary of different mapping domains. The computational data would provide a complement to the experimental data in the mapping space. The systematic calculations on carefully selected physical model will be possible to create a “theoretical structure map”. A preliminary study has been done on AB intermetallic compounds[7].

3.2 Further screen and prediction of hypothesis

It is predictable that the number of suggested systems with a pre-defined physical

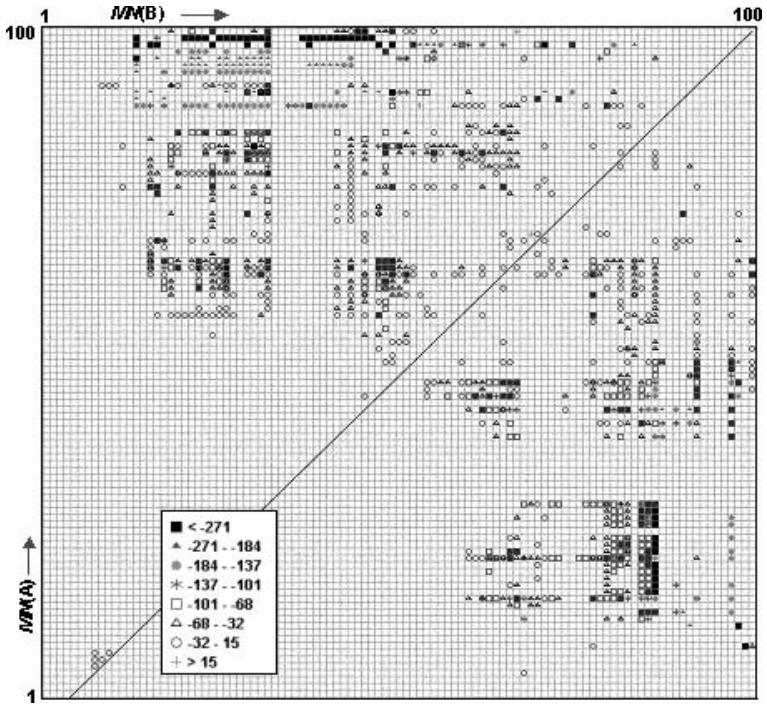

Fig. 6 Map of binary systems classified by values of Molar Enthapy of Formation (KJ/Mol) with $MN(A)$ vs. $MN(B)$. Different symbols mark different range of values of enthalpy of formation.

property after the filtering from conditions in the discovery tool is still large, the more acceptable list of candidates will be further subjected to screening by various computational approaches. This would involve testing the structure stability by calculation of formation energy, and predicting target physical properties on those hypothetical systems or hypothetical structures.

3.3 Physical interpretation of regularities

The calculations on the proper models will help to provide the insight of some successful regularities and correlations discovered by semi-empirical database approaches. For example, the first principles calculation of phase stability could reveal the intrinsic reason of non- formation of compounds, and the possibility of compound formation although two elements are assigned to non-formation by empirical approaches. Also, a series of first principles calculations on phase diagram revealed the systematical trend of structure stability in certain group of elements[7–9]. Meanwhile, the calculation of stability on the atomic environment type (AET) would provide/verify the key factors governing the relation between structure of materials and atomic properties of composition elements.

4 Summary

Preliminary studies have been carried by taking advantage of the large-scale materials database to discover the regularities and the correlations in the binary systems. This

practice further provides experience on the following aspects: how to find design solutions by dealing with data and models; how to prepare working hypothesis effectively to converge to a solution, and furthermore, how to design the experiments. By combining with theoretical prediction method, deeper hidden patterns would be revealed, and it will lead to the practicable procedures of materials design.

In spite of the non-equilibrium property of many new materials such as nano materials, the PAULING FILE, which is a collection of data in equilibrium state also provides valuable information on the potential candidate materials indirectly by using its "Discovery Space" to monitor some regularities as the starting references to explore new materials with high possibilities. The final design solutions will be the emergence of the integrating approach after examining a broad base of candidates intelligently.

References

[1] P. Villars, M. Berndt, K. Brandenburg, K. Cenzual, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz and S. Iwata: PAULING FILE CD-ROM, Binaries Edition, MPDS, 2002. http://www.paulingfile.ch/

[2] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.-H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen and S. Iwata: J. Alloys Comp., 317–318, pp.26–38, 2001.

[3] D.G. Pettifor: Solid State Commun., 51, pp.31–34, 1984.

[4] P. Villars, F. Hulliger, J. Less-common Met., 132, pp.289–315, 1987.

[5] P. Villars, K. Cenzual, J. Daams, Y. Chen and S. Iwata, accepted by J. Alloys Comp., 2003.

[6] Y. Chen, S. Iwata, J. Liu, P. Villars and J. Rodgers, Modelling Simul: Mater. Sci., Eng. 4, pp.335–348, 1996.

[7] Y. Chen, T. Atago and T. Mohri: J. Phys.,Condens. Matter, 14, pp.1903–1913, 2002.

[8] T. Mohri, Y. Chen: Materials Transactions, 43, pp.2104–2109, 2002.

[9] Y. Chen, S. Iwata and T. Mohri: CALPHAD, 26, pp.583–589, 2002.

(Received May 8, 2003)
(Accepted June 6, 2003)