Data-driven atomic environment prediction for binaries using the Mendeleev number Part 1. Composition AB

P. Villars [a,*], K. Cenzual [b], J. Daams [a], Y. Chen [c], S. Iwata [c]

[a] *Materials Phases Data System, Schwanden 400, Vitznau CH-6354, Switzerland*
[b] *Department of Inorganic, Analytical and Applied Chemistry, Geneva University, Geneva, Switzerland*
[c] *QUEST, The University of Tokyo, Tokyo, Japan*

Abstract

The atomic environment types (AETs) (coordination polyhedra) realized by each chemical element in binary compounds at the equi-atomic composition were analyzed based on a comprehensive set of literature data. The Mendeleev number (*MN*) (ordering number listing the chemical elements column by column through the periodic system) was successfully used to classify the chemical systems. An atomic environment type map, using as coordinates the maximum Mendeleev number versus the ratio between the minimum and the maximum Mendeleev number, sub-divided the chemical systems where different atomic environment types occur in distinct stability domains. The same maps also showed a clear separation between chemical systems where intermediate compounds form and those where no compounds form. These maps make it possible to predict the existence of compound that have not yet been investigated with a particular atomic environment.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Data mining; Atomic environment; Mendeleev number

1. Introduction

The large amount of experimental data collected over the years contains information and correlations that can be used for a rational semi-empirical approach to materials design. The ideal situation is to be able to relate any kind of "compound property" to one or several parameters characteristic of the constituent elements. In a previous study [1], concerning the formation of binary compounds, a very clear separation between chemical systems where compounds form and those where no compounds form was achieved using maps based on combinations of the Mendeleev numbers (*MN*) of the constituent elements. Very encouraging was the fact that the results obtained on binaries could be extended to ternaries and quaternaries. This is an extremely important point because materials design is nowadays more and more focused on multinaries, but the experimental knowledge available is only substantial for binaries (approximately 70% of all binary systems have been studied, but less than 5% of the ternary and less than 0.5% of the quaternary systems).

After having defined the conditions for the chemical systems where binary compounds form, it seemed of interest to further study the structural features of these compounds. We have focused on the most frequently occurring stoichiometry for binary compounds, AB, and for each compound we have considered the atomic environment type (AET), or co-ordinations, of the atoms. The major motivation of the study was to find answers to the following questions: Do particular chemical elements strongly prefer certain AETs? How complex is the distribution of A and B atoms in mixed AETs? Do there exist 'stability criteria' for the different AETs?

1.1. Elemental parameters

The elemental parameters proposed so far for the construction of different compound property maps can be subdivided into different groups: atomic number factor, group number factor, quantum number factor, Mendeleev number factor, cohesion-energy factor (enthalpy of formation, etc.), electrochemical factor (electronegativity, etc.), size factor (Zunger pseudo-potential, covalent radius, atomic volume, etc.), atomic environment (AE) factor. The first four groups are completely independent of the nature of other atoms.

* Corresponding author.
E-mail address: villars@mpds.ch (P. Villars).

Mendeleev Number *MN*

																H 1 92	He 2 98
Li 3 1												Be 4 67	B 5 72	C 6 77	N 7 82	O 8 87	F 9 93	Ne 10 99
Na 11 2												Mg 12 68	Al 13 73	Si 14 78	P 15 83	S 16 88	Cl 17 94	Ar 18 100
K 19 3	Ca 20 7	Sc 21 11	Ti 22 43	V 23 46	Cr 24 49	Mn 25 52	Fe 26 55	Co 27 58	Ni 28 61	Cu 29 64	Zn 30 69	Ga 31 74	Ge 32 79	As 33 84	Se 34 89	Br 35 95	Kr 36 101
Rb 37 4	Sr 38 8	Y 39 12	Zr 40 44	Nb 41 47	Mo 42 50	Tc 43 53	Ru 44 56	Rh 45 59	Pd 46 62	Ag 47 65	Cd 48 70	In 49 75	Sn 50 80	Sb 51 85	Te 52 90	I 53 96	Xe 54 102
Cs 55 5	Ba 56 9		Hf 72 45	Ta 73 48	W 74 51	Re 75 54	Os 76 57	Ir 77 60	Pt 78 63	Au 79 66	Hg 80 71	Tl 81 76	Pb 82 81	Bi 83 86	Po 84 91	At 85 97	Rn 86 103
Fr 87 6	Ra 88 10																

La 57 13	Ce 58 15	Pr 59 17	Nd 60 19	Pm 61 21	Sm 62 23	Eu 63 25	Gd 64 27	Tb 65 29	Dy 66 31	Ho 67 33	Er 68 35	Tm 69 37	Yb 70 39	Lu 71 41
Ac 89 14	Th 90 16	Pa 91 18	U 92 20	Np 93 22	Pu 94 24	Am 95 26	Cm 96 28	Bk 97 30	Cf 98 32	Es 99 34	Fm100 36	Md101 38	No102 40	Lr103 42

Fig. 1. Mendeleev numbers (*MN*) used here, in a periodic system representation.

The Mendeleev number is an ordering number attributed to each chemical element in the periodic system with the aim to order them so that chemical elements with a similar behavior follow each other. The fundamental ordering number of the chemical elements, the atomic number, does not fulfill this criterion. The string of increasing atomic numbers runs through the periodic system period by period. The first officially termed Mendeleev number was introduced by Pettifor in 1984 [2], based on a phenomenological optimization of the separation of binary AB compounds into different structure types. Basically, the string of any variant of Mendeleev numbers runs through the periodic system group by group. The particular optimization used by Pettifor had the effect to displace Eu and Yb just after Ca, Pb after Ga, etc. We use a Mendeleev number that is more directly derived from the periodic system, and where the string of increasing Mendeleev numbers runs through the groups, from up to down, one after the other from left to right. With respect to the most common presentations of the periodic systems, H has been placed on top of the halogen group, and Be, Mg on top of group 12 (Zn, Cd, Hg), see Fig. 1. These exceptions are strongly supported by the present study.

1.2. Data

This investigation is based on information contained in the Pauling File, Binaries Edition [3]. This phase-oriented database contains, under the same computer frame, phase diagrams, and crystal structure data, including information about the atomic environments. The latter are searchable parameters and can easily be visualized on the screen. Additional information on chemical systems where no compounds are formed was taken from Okamoto [4].

A chemical system is defined as a *former* if, at atmospheric pressure, it contains at least one phase that is separated from the terminal solid solutions of the constituent elements by a two-phase region. A system that does not fulfill this criterion is called a *non-former*. Non-former systems are characterized by enthalpies of formation that are either positive or

close to zero, in agreement with the theory of Miedema [5]. Two-phase mixtures of the constituent elements (simple eutectic, simple peritectic, and complete insolubility), as well as complete solid solutions are non-formers. Systems with terminal or complete solid solutions where ordered phase areas are observed are also considered as non-formers, because the enthalpy of formation associated with the ordering of the atoms at the positions of the elemental crystal structure is very close to zero. We will also distinguish between chemical systems where a phase, which can be a line compound or have a significant homogeneity region, is formed at the equi-atomic composition—AB formers, and former systems where a two-phase region is observed at room temperature at the equi-atomic composition—non-AB formers.

Two different datasets were prepared. For dataset 1, non-formers were distinguished in agreement with the definition given above. For dataset 2, less emphasis was given to the thermodynamical aspect, i.e. the atomic environments observed in any single-phase sample of composition AB formed at ambient pressure, including complete solid solutions, terminal solid solutions extending beyond 50 at.% and phases undergoing eutectoid decomposition before reaching room temperature, were taken into consideration. The pseudo-binary systems A–A, i.e. the chemical elements, were also included in the second dataset.

The information used in the present investigation is summarized in Table 1. For two thirds of the chemical systems formed by the chemical elements considered here (all except the rare gases, Ra, and some actinides), information about what happens at the equi-atomic composition is available. Based on such a high percentage it should be possible to make sound predictions for the remaining 30% of systems that have not yet been investigated, if a strong pattern can be identified.

1.3. Atomic environment

The atomic environment, or coordination, of the atoms, i.e. the number of nearest neighbors and the geometrical arrangement formed by these, is an important characteristic of crystal structures. A classification scheme of intermetallic

Table 1
Overview of the data used in the present study

Chemical systems	Dataset 1	Dataset 2
Not yet investigated	1480 (34.5)	1481 (33.9)
No compounds formed (non-formers)	760 (17.8)	540 (12.3)
No AB compounds formed (non-AB formers)	839 (19.6)	794 (18.2)
AB compound reported but no crystal data (AB formers?)	73 (1.8)	79 (1.8)
AB compound with crystal data (AB formers)	1126 (26.3)	1384 (31.7)
Chemical elements		93 (2.1)
Total	4278	4371

Values in parenthesis are percentages.

structure types based on the coordination of the smallest atoms in the structure, where the coordination numbers (CN) range from 2 to 12, was proposed by Kripyakevich [6] (see also [7]).

In order to determine the atomic environment, as described in [8] and the references therein, complete crystal structure data must in principle be available, i.e. space group, unit-cell dimensions, coordinates of the point-sets (atom sites) and their occupation. For the majority of the different structure types adopted by inorganic compounds, however, the atomic environments are similar for isotypic compounds, so that the atom coordinates of the type-defining compounds can be used as a first approximation. A large number of equi-atomic compounds crystallize with structure types where all atom positions have fixed coordinates.

The atomic environment is defined using the method of Brunner and Schwarzenbach [9], where all interatomic distances between an atom and its neighbors are plotted in a next-neighbor histogram. In most cases a clear maximum gap is revealed. All atoms to the left of this maximum gap belong to the atomic environment of the central atom. This rule is called the maximum gap rule and the atomic environment type, also called coordination polyhedron, is constructed with the atoms to the left of the maximum gap. In those cases where this rule leads to AETs with more than the central atom enclosed in the AET or to AETs with atoms on the faces of the coordination polyhedron, we use the maximum-convex-volume rule. This rule is defined as the maximum volume around only one central atom enclosed by convex faces, with all the atoms lying at the intersections of at least three faces. This rule was also used in those cases where no clear maximum gap was detectable. In those cases where two (or more) equal, or practically equal, maximum gaps were observed, we kept the number of different AETs

in a structure type as small as possible. For combinations of metals with p-elements, on the right hand-side of the Zintl line, or in combinations with hydrogen, we sometimes obtained irregular AETs.

The atomic environment type is not a simple compound property in the sense that different atomic environments are generally observed in the same compound. Experience shows, however, that the number of atomic environments realized by one chemical element in the same structure is generally low. For each binary compound we have thus considered two atomic environments, one for each kind of element, leaving the exceptions in a category-labeled 'complex'.

2. Results

2.1. Observed atomic environment types

The coordination of each crystallographic point-set within all binary AB compounds was examined. In total, the AETs of the two elements in 1126 binary phases AB for dataset 1 and 1384 for dataset 1, were investigated.

The results show that 31 different AETs are realized in AB compounds, but coordination polyhedra such as tetrahedron, trigonal prism, octahedron, cube, tricapped trigonal prism, cuboctahedron, rhombic dodecahedron, heptacapped pentagonal prism, largely predominate. The 15 most populous AETs are shown in Fig. 2 and their occurrences are given in Table 2. It was further observed that

- The elements prefer to realize one single kind of AET, even in structures having several different crystallographic sites.
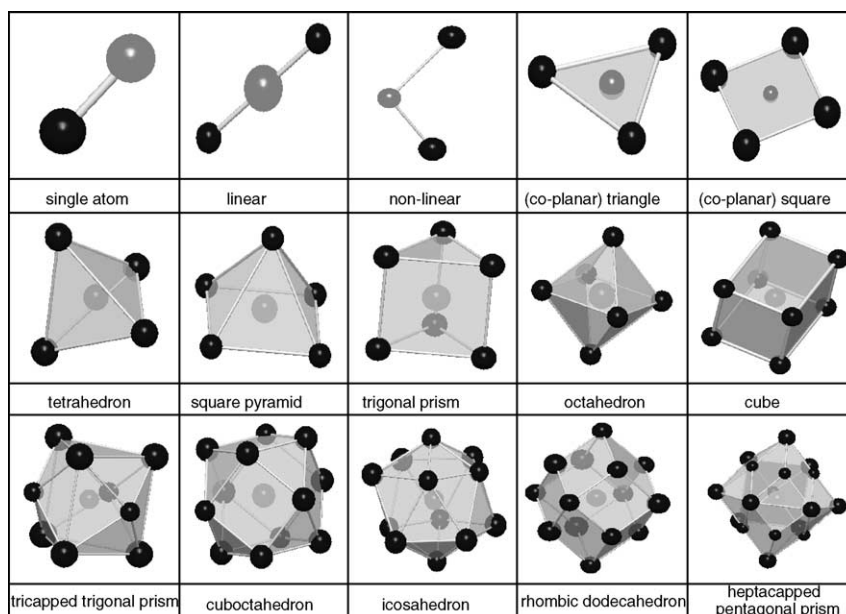

Fig. 2. The 15 most populous AETs for AB compounds.

Table 2
Particular atomic environment types observed in AB compounds, their coordination numbers (CN) and occurrences in dataset 1 (basic definition of non-formers) and dataset 2 (including simple chemical elements and solid solutions)

Atomic environment type	CN	Dataset 1	Dataset 2
Single atom	1	27	36
Linear	2	9	11
Non-linear	3	20	25
(Co-planar) triangle	3	61	70
(Co-planar) square	4	12	12
Tetrahedron	4	103	108
Square pyramid	5	8	10
Trigonal prism	6	120	121
Octahedron	6	628	637
Cube	8	226	227
Tricapped trigonal prism	9	194	194
(Anti)cuboctahedron	12	104	525
Icosahedron	12	16	18
Rhombic dodecahedron	14	326	445
Heptacapped pentagonal prism	17	168	169
Others		268	225

- The distribution of A and B atoms within a mixed coordination polyhedron is in general limited to the highest symmetrical distribution, although there exist in principle many ways to distribute, e.g. eight B and six A atoms at the vertices of a rhombic dodecahedron.
- The maximum number of coordinating atoms B observed in the atomic environment of an atom A is nine, although the coordination numbers vary from 1 to 18.
- For the lower coordination numbers (≤ 6), all coordinating atoms are in general of the other kind.

For graphical presentations, the less frequently observed AETs were sub-divided into four categories: 7-vertex polyhedra, 9–11-vertex polyhedra, 13–16-vertex polyhedra and complex. In the latter group one finds also pseudo-ternary phases like AuSe or TlS, where the cations are present in two different valence states, each one with a well-defined atomic environment.

2.2. Atomic environment type map MN_A versus MN_B

Fig. 3a shows the distribution of the AETs observed for binary AB phases in dataset 1, including information about non-AB former and non-former systems. The chemical element occupying the center of the polyhedron, hereafter referred to as A, is given on the y-axis and the other element in the binary compound, B, on the x-axis. It should, however, be noted that both A and B atoms can be part of the atomic environment. On both axes the chemical elements are ordered according to the Mendeleev number defined above. Each chemical system is thus present twice in the table, the AETs of the chemical element with the highest Mendeleev number being indicated in the fields situated below the diagonal corresponding to the simple chemical elements. The diagonal is left empty for dataset 1.

It should be noted that an AET map is different from the better-known structure maps, where each AB compound is characterized by its structure type. On one side several closely related structure types, like TlI and FeB, have similar atomic environments and are thus grouped together. On the other side, some commonly occurring structure types change their atomic environments substantially when the cell parameter ratios change. For the simple cubic CsCl type the histogram used for the determination of the maximum gap is identical for all representatives. Depending on the combination of elements, different AETs have been considered for this type: a cube for both elements in ionic compounds such as CsCl itself, a rhombic dodecahedra for both elements in intermetallic compounds like AuZn, a rhombic dodecahedron for the larger element and a cube for the smaller element in rare-earth transition element compounds.

For ionocovalent compounds, A atoms have been considered to belong to the AET of A only when linked to the central atom by a covalent bond. The number of homonuclear bonds per central atom is indicated by a point pattern in Fig. 3a. The density of the pattern increases when going from 1 (e.g. KO) to three homonuclear bonds (e.g. NaSi) per atom.

Some of the information included in the AET map MN_A versus MN_B in Fig. 3a is by definition symmetric with respect to the diagonal. This is the case for the codes for non-formers, non-AB formers and AB formers with no structure data. The atomic environments of the elements A and B do not have to be identical, however, the overall picture presents a strong symmetry with respect to the diagonal, due to the high occurrence of simple structure types like CuAu, β′-AuCd, NaCl, ZnS sphalerite, CsCl (see restrictions above) where the two atom sites are interchangeable. This situation is particular for the equi-atomic composition, where, among the 200 known AB structure types, most are single environment types, meaning that the AETs of both elements are the same. Simple exceptions are structure types like for example NiAs.

A relatively good separation of the different chemical systems into distinct AET and non-former domains, is achieved using this simple classification. No attempt was made to optimize the Mendeleev number by modifying the order of the rows and columns in the table. It is, however, worth noting that when the Mendeleev number is replaced by an elemental parameter from one of the other elemental parameter groups (e.g. atomic number, cohesion-energy factor, electrochemical factor or size factor) the resulting AET maps show a lower degree of local ordering.

Well-defined regions of identical colors are observed also for dataset 2 where solid solutions and the chemical elements are included (see Fig. 3b). In particular, the categories of (anti)cuboctahedra (close packed structures) and rhombic dodecahedra (W/CsCl type) are substantially increased.

2.3. Atomic environment type map MN_{max} versus MN_{max}/MN_{min}

The pattern observed in the AET map MN_A versus MN_B encouraged us to search for more strongly defined correlations by combining elemental parameters. Our aim was to try to define stability conditions for each AET, e.g. identify the region of binary systems where

the chemical element Zr forms equi-atomic compounds where each Zr atom is surrounded by six atoms forming an octahedron. If this is achieved, it is possible to quantify the conditions for the formation of a particular AET. As the next step it is easy to calculate the appropriate parameters for a non-investigated system and see if the conditions for the formation of a particular AET are fulfilled.
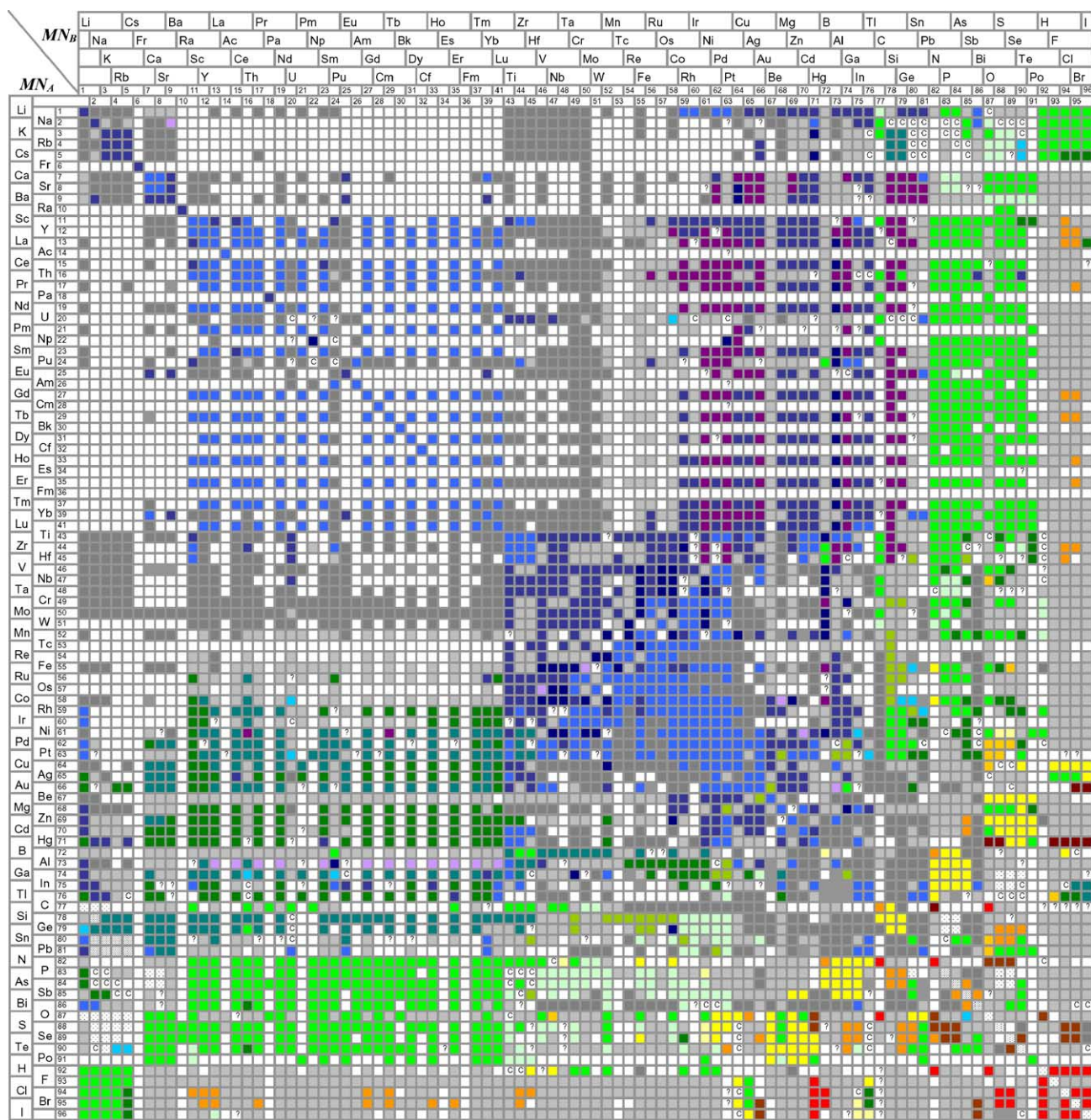

Fig. 3. Atomic environment type (AET) maps showing Mendeleev numbers MN_A vs. MN_B for AB compounds for (a) dataset 1 (AETs being limited to thermodynamically defined AB former systems) and (b) dataset 2 (AETs including simple chemical elements and extended terminal solid solutions). Element A is defined as the central atom of the atomic environment, so that AETs centered by the same element in different compounds are found along the same row. Dotted patterns indicate the existence of covalent homonuclear bonds.
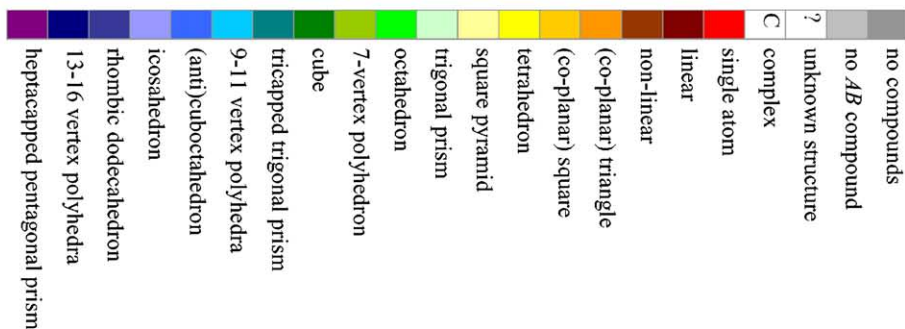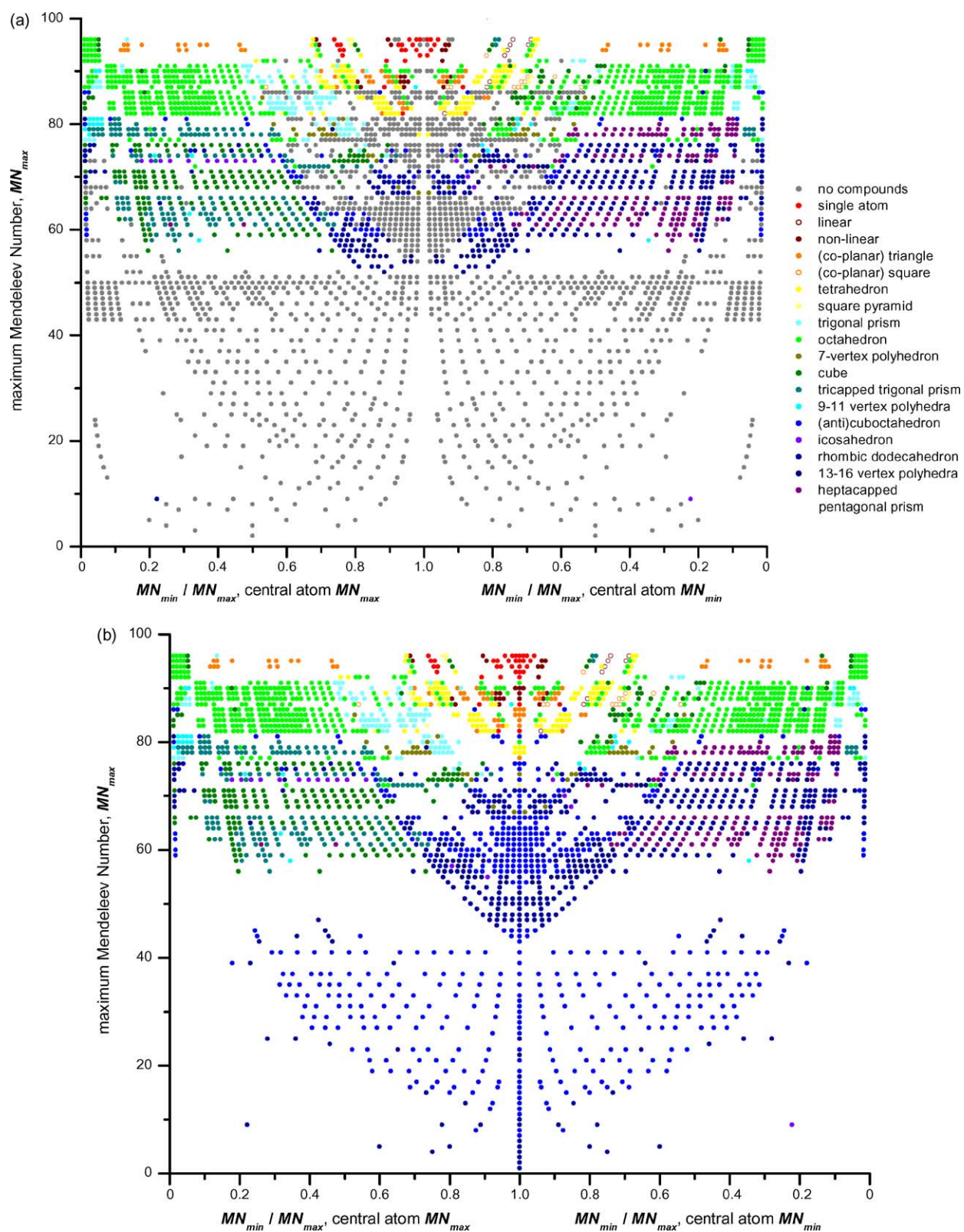
(b)

Fig. 3. (*Continued*).

Fig. 4. Atomic environment type (AET) stability maps showing Mendeleev numbers MN_{max} vs. MN_{min}/MN_{max} for AB compounds for (a) dataset 1 (AETs being limited to thermodynamically defined AB former systems) and (b) dataset 2 (AETs including simple chemical elements and extended terminal solid solutions). AETs of the elements with the highest Mendeleev number in the AB compounds are given on the left-hand side of $MN_{min}/MN_{max} = 1$, AETs of the elements with the lowest Mendeleev number on the right hand-side.

In order to find such correlations we combined elemental parameters of the constituent elements (EP_A and EP_B) by applying simple mathematical operators ($+$, $-$, $\times$, $/$, min, max). The following expressions were used, where the elemental parameters can be the atomic numbers, group numbers, size factors, etc.:

- sum $= EP_A + EP_B$
- difference $= |EP_A - EP_B|$
- product $= EP_A \times EP_B$
- ratio $= EP_A/EP_B$, with $EP_A < EP_B$
- maximum $= \max (EP_A, EP_B)$
- minimum $= \min (EP_A, EP_B)$

As noted above, the atomic environment type represents a particular case since for each compound we have defined not one but two compound properties (the atomic environment of each of the chemical elements). For the tests each dataset was thus divided into two sub-sets: the AET of the element with highest Mendeleev number in the compound, AET (MN_{max}), and the AET of the element with lowest Mendeleev number, AET (MN_{min}). All possible 2- and 3-dimensional combinations, e.g. sum $= EP_A + EP_B$ versus product $= EP_A - EP_B$, were investigated for several elemental parameters and the correlations evaluated by the closest neighbor (domain) method. The best separation into domains was achieved for the AET map MN_{max} versus MN_{min}/MN_{max}. The two sub-sets of dataset 1 are presented together in Fig. 4a with a pseudo-mirror plane at $MN_{min} = MN_{max}$, i.e. the vertical line corresponding to the chemical elements.

The separation of the different categories of chemical systems into distinct stability domains is very good and several observations can be made. First of all, the non-formers are clearly separated from the AB formers. The majority of the non-formers correspond to combinations of elements with low Mendeleev numbers, i.e. metals situated on the left-hand side of the periodic system. The ordering according to the maximum Mendeleev number (*y*-axis) has a very strong separating effect, which is also the case for different AETs. From the top to the bottom of the graph the binary compounds are sub-divided into classes such as iodides, bromides, chlorides, sulfides, oxides, etc., each one represented by one row in the graph. With increasing values of the Mendeleev number one goes from non-formers to AB formers, and within the AB formers the coordination numbers decrease from 15 to 1. The ratio MN_{min}/MN_{max} (*x*-axis) is, however, necessary to achieve a good separation. Compounds formed by elements with similar Mendeleev numbers are situated close to the vertical line at $x = 1$. For each compound class (iodides, sulfides, etc.) the compounds where the other element has a lower Mendeleev number are equally distributed along the line section 1–0–1, realizing a maximum spread.

In the corresponding graph made for dataset 2 (Fig. 4b), about one third of the non-formers have been replaced by

AET codes, mainly located next to the pseudo-mirror plane at $MN_{min}/MN_{max} = 1$. It is satisfactory to see that the additional AETs, including those of the simple chemical elements, fit well into the general pattern.

The chemical systems where compounds form, but not at the equi-atomic compositions (non-AB formers), are not shown in Fig. 4. They are not grouped together in clearly defined domains, nevertheless some tendencies are observed: In particular, very few non-AB formers are located in the non-former domains. Very few non-AB formers are also located in the domains where $MN_{min}/MN_{max} < 0.5$ and octahedra are observed. For certain MN_{max} values (e.g. 56, 58, 67, 72 and 93, corresponding to Ru, Co, Be, B and F, respectively) a high density of non-AB formers is noted. Non-AB formers are also common at the boundary between the non-formers and AB formers.

3. Conclusions

We have found answers to the questions that motivated this study of the atomic environment types observed in binary equi-atomic compounds:

- The elements prefer a limited number of AETs (tetrahedron, trigonal prism, octahedron, cube, tricapped trigonal prism, cuboctahedron, rhombic dodecahedron, and heptacapped pentagonal prism).
- Each element strongly prefers one kind of AET within the same phase.
- A highly symmetric distribution of atoms of different kinds is observed for mixed coordination polyhedra.

Stability domains are clearly distinguished in AET maps based on a combination of the Mendeleev numbers of the constituent elements MN_{max} versus MN_{min}/MN_{max}. It is now possible to predict the existence of equi-atomic compounds systems in systems that have not yet been studied and to assign a probable atomic environment, or to search for equi-atomic compounds with a particular atomic environment.

The present study proves that the analysis of critically evaluated datasets can lead to the discovery of so far "hidden" patterns, which can then be used to develop tools (rules) for semi-empirical materials design. In the future we will study the effect of the concentration on the AET stability criteria, by investigating AB_2, AB_3, A_2B_3, A_3B_5, etc., compounds. The approach will then be extended to ternaries and quaternaries.

References

[1] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.-H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen, S. Iwata, J. Alloys Compd. 317–318 (2001) 26–38.
[2] D.G. Pettifor, Solid State Commun. 51 (1984) 31.

[3] P. Villars, M. Berndt, K. Brandenburg, K. Cenzual, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz, S. Iwata, PAULING FILE CDROM, Binaries Edition, ASM International, Materials Park, OH, 2002.

[4] H. Okamoto, Binary Alloy Phase Diagrams, Desk ed., ASM International, Materials Park, OH, 2001.

[5] F.R. de Boer, R. Boom, W.C.M. Mattens, A.R. Miedema, A.K. Niessen, in: F.R. de Boer, D.G. Pettifor (Eds.), Cohesion in Metals, North-Holland, Amsterdam, 1988.

[6] P.I. Kripyakevich, Structure Types of Intermetallic Compounds, Nauka, Moscow, 1977 (in Russian).

[7] E.I. Gladyshevskii, O. Bodak, in: J.H. Westbrook, R.L. Fleischer (Eds.), Intermetallic Compounds, vol. 1, Wiley, New York, 1994, pp. 403–418.

[8] J.L.C. Daams, in: J.H. Westbrook, R.L. Fleischer (Eds.), vol. 1, Intermetallic Compounds, Wiley, New York, 1994, pp. 363–383.

[9] G.O. Brunner, D. Schwarzenbach, Z. Kristallogr. 133 (1971) 127–133.