

SCIENTIFIC REPORTS



OPEN

An informatics guided classification of miscible and immiscible binary alloy systems

R. F. Zhang¹, X. F. Kong¹, H. T. Wang², S. H. Zhang¹, D. Legut³, S. H. Sheng¹, S. Srinivasan⁴, K. Rajan⁵ & T. C. Germann⁶

The classification of miscible and immiscible systems of binary alloys plays a critical role in the design of multicomponent alloys. By mining data from hundreds of experimental phase diagrams, and thousands of thermodynamic data sets from experiments and high-throughput first-principles (HTFP) calculations, we have obtained a comprehensive classification of alloying behavior for 813 binary alloy systems consisting of transition and lanthanide metals. Among several physics-based descriptors, the slightly modified Pettifor chemical scale provides a unique two-dimensional map that divides the miscible and immiscible systems into distinctly clustered regions. Based on an artificial neural network algorithm and elemental similarity, the miscibility of the unknown systems is further predicted and a complete miscibility map is thus obtained. Impressively, the classification by the miscibility map yields a robust validation on the capability of the well-known Miedema's theory (95% agreement) and shows good agreement with the HTFP method (90% agreement). Our results demonstrate that a state-of-the-art physics-guided data mining can provide an efficient pathway for knowledge discovery in the next generation of materials design.

The purpose of the Materials Genome Initiative (MGI) is to accelerate the discovery of novel materials by means of modern computational techniques and data mining methods^{1–3}. Successful examples reported so far include solar water splitters, solar photovoltaics, topological insulators, scintillators, clean energy materials, piezoelectrics, thermoelectrics, catalysts, hydrogen storage materials, and Li-ion batteries^{2,3}. With the emergence of MGI, an exciting trend has recently appeared, utilizing collaborative informatics technology, e.g. data collection and data-mining, to extract meaningful information and patterns from massive data for the fast design and development of novel materials⁴. The core task is to efficiently leverage various open-source data with state-of-the-art data-mining algorithms. In the fields of physical metallurgy, vast new data have been recently collected and reported for the mixing properties of binary alloy systems. This provides a unique opportunity to construct a visual miscibility map for binary alloy systems by mining these large numbers of data sources.

The miscibility of a binary alloy system, which indicates the tendency of two elements to form either a homogeneous solid solution or an intermetallic compound, plays an important role in the development of high-performance metallic materials such as diffusion barrier materials, bulk metallic glasses, damage-resistant nanomaterials, microelectronics, or superalloys^{5,6}. From a thermodynamical point of view, an immiscible system corresponds to a positive (i.e., endothermic) formation enthalpy^{7–9} (e.g., Cu-Nb systems^{10,11}). At the other extreme, a binary alloy system may readily form intermetallic compounds, or a homogeneous or clustered solid solution, typically with a solubility of about 10%^{5,12–14}. For such miscible cases, the alloying process is exothermic, with a negative formation enthalpy (e.g., Cu-Ti system)¹⁵. Several heuristic rules have been proposed to qualitatively predict the alloying behavior or reactivity of two metals, of which the most important ones are: the electronegativity rule¹⁶, the Hume-Rothery rule¹⁷ and the Darken and Gurry maps^{18,19}. A very successful

¹School of Materials Science and Engineering, and International Research Institute for Multidisciplinary Science, Beihang University, Beijing, 100191, P. R. China. ²CAS Key Laboratory of Nuclear Materials and Safety Assessment, Institute of Metal Research, Chinese Academy of Sciences, Shenyang, 110016, P.R. China. ³IT4Innovations Center & Nanotechnology Centre, VSB-Technical University of Ostrava, CZ-70833, Ostrava, Czech Republic. ⁴Plant Sciences Institute, Iowa State University, 2031 Roy J. Carver Co-Lab, Ames, IA, 50011, USA. ⁵Department of Materials Design and Innovation, University at Buffalo-State University of New York, 311 Bell Hall, Buffalo, NY, 14260, USA. ⁶Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. Correspondence and requests for materials should be addressed to R.F.Z. (email: zrf@buaa.edu.cn)

semi-quantitative theory has been proposed by Miedema *et al.*¹² in the late 1970's to distinguish the negative (miscible) and positive (immiscible) systems. This model makes it possible to predict the formation enthalpies of binary alloy systems in a semi-quantitative manner²⁰. However, the model is sensitive to the choice of its parameters, which are determined empirically. The most critical problems with Miedema's model are: a) it needs more reliable experimental data to validate, improve and extend it^{12, 13, 21–23}; b) as will be shown later, a large scatter of the results occurs when the formation enthalpy of the system is close to zero, i.e., for weakly alloying systems; and c) several outliers need be sorted out before it can be reliably used to guide the development of novel alloys.

The validation of such empirical and semi-quantitative rules depends generally on the availability of reliable and sufficient thermodynamic data. The recent progress in the development of the experimental techniques, thermodynamic analysis and modern first-principles methods provides an opportunity to improve the database of alloying behavior. For example, Kleppa *et al.*^{24–35} published experimental values of formation enthalpies with high accuracy, leading to updates in the Landolt-Boernstein database⁷, the handbook of binary alloy phase diagrams compiled by Okamoto⁹, and the alloy phase diagram database organized by ASM community⁸. In addition, modern high-throughput first-principles (HTFP) calculations^{2, 36–40} provide a unique way to check and compare the database from different sources. All of these data sources provide not only a pathway for a global data mining of alloying behavior of binary alloy systems, but an opportunity to validate the widely used Miedema theory and compare with HTFP method^{21–23}.

In this article we first present a global collection and survey of the alloying behavior of binary systems based on the most recent phase diagrams, thermodynamic data and HTFP calculations. Second, we propose a data mining scheme to classify miscible and immiscible systems based on a general consideration of thermodynamic rules. Note that such data mining is not a trivial task; one may obtain a misleading conclusion if the data interpretation is not accompanied by a proper statistical analysis. It should be also noted that the requirement of massive data provides a foundation of the reliability of the present scheme for data mining, which however limits the present method to be used when sufficient data are unavailable for statistical analysis. Third, a series of two-dimensional miscibility maps are constructed based on several physically relevant descriptors to distinguish the immiscible systems from miscible ones by pattern recognition. Fourth, the neural network algorithm and elemental similarity are used to predict the alloying behavior of unknown systems, thus providing a complete miscibility map. Finally, a thorough validation of the classical Miedema theory and a comprehensive comparison with HTFP method is performed by statistical analysis.

Results

Classification of miscible and immiscible alloy systems. In all, 44 metallic elements comprise the 27 transition metals (TM), including Zn, Cd and Hg, and 17 lanthanide (rare earth (RE)) metals, including Sc and Y. Because of the scarcity of experimental and HTFP data for binary alloy systems consisting of binary lanthanide metals, i.e. RE-RE, we shall not discuss this group except for three binary systems of Sc-Y, Sc-La and La-Y. Consequently, we investigate a total of $C_{27}^2 = 351 + 459 + 3 = 813$ binary alloy systems in the present study. The classification of different solid solutions (with positive and negative formation enthalpy) is not an easy task since there is no direct unified experimental data to distinguish them. However, the formation enthalpy provides an energetic criterion for this purpose. After a global survey, we found 35 binary alloy systems where a disagreement appeared between phase diagram analysis and HTFP calculations. These are listed in Table 1 as “outliers”. As regards this group, our final choice of miscibility is based on an evaluation of the available experimental information, the similarity of chemical elements, and a global justification by comparison of different data sources. Remarks are added in the last column: SSL: low solid solubility; SSH: high solid solubility; COM: formation of compounds; COM1: only one compound reported; IMM: immiscibility gap exists.

Special binary systems deserve particular attention. For the Ag-Hg, Au-Hg, Co-Cr, Cr-Mn, Cr-Tc, Cr-Zn, Hg-Mn, Hg-Ni, Hg-Zn, Mn-Y, Mo-Tc, Re-W, and Tc-W systems, the formation of stable intermetallic compounds have been reported in the published phase diagrams. Thus, we believe that these systems should be identified as miscible systems with a negative formation enthalpy, although the HTFP calculations provide opposite results. For Ag-Mn, Au-Cr, Fe-Os, Fe-Ru, Mn-W, Os-Pt, and Ta-Ti, the HTFP calculations are accepted because there is no information on phase diagrams or thermodynamic data available. For other binary alloy systems listed in Table 1, our judgment is based on the elemental similarity, as will be discussed later.

Selection of physical descriptors. In order to enable pattern recognition, several physical descriptors, i.e. semi-empirical parameters and scales, are considered for a systematic description (or prediction) of the alloying behavior of the different metals. The physical descriptors are based on the similarity of atomic, physical and chemical properties of the elements, such as the atomic number (AN), Pauling electronegativity (PEN)^{16, 41, 42}, Teatum metallic radii of elements with coordinate number CN12 (Rij)^{43–45}, Martynov-Batsanov electronegativity (MBEN)^{35, 46}, Zunger's pseudopotential radii sum (Rsp)^{35, 47}, Miedema's work function (Φ)^{5, 14}, Miedema's electron density at the Wigner-Seitz cell boundary (Nws)^{5, 14}, Miedema's molar volume (V_m)^{5, 14}, and Mendeleev number or Pettifor chemical scale (Pcs)⁴⁸. The choice of these parameters or scales is based on the success of these parameters in the description of different alloying behaviors.

The first choice is the atomic number, which uniquely identifies a chemical element. The second criterion is based on the consideration of the diagrams developed by Darken and Gurry for solid solution predictions¹⁹. The two coordinates represent the Pauling electronegativity and the atomic size of the elements. Electronegativity is a chemical property that describes the tendency of an atom or a functional group to attract electrons (or electron density). Pauling originally proposed the concept of electronegativity as an explanation of the fact that the covalent bond between two different atoms (A-B) is stronger than that would be expected by taking the average of the strengths of the A-A and B-B bonds. Allred^{16, 41, 42} updated Pauling's original values by taking into account the available thermodynamic data. These “revised Pauling” values of the electronegativity are most often used.

Alloy systems	Alloying behavior			Remarks ^a
	HTEFP	Phase diagram	Present choice	
Ag-Hg	+	–	–	COM
Au-Cr	+	–	+	COM1
Au-Hg	+	–	–	COM
Cd-Cu	+	–	–	COM
Cd-Nb	+	–	+	COM
Co-Cr	+	–	–	COM
Co-Ir	+	–	–	SSH
Co-Os	+	–	–	SSH
Co-Rh	+	–	–	SSH
Co-Ru	+	–	–	SSH
Cr-Fe	+	–	–	COM1
Cr-Mn	+	–	–	COM
Cr-Re	+	–	–	SSH
Cr-Ru	+	–	–	SSH
Cr-Tc	+	–	–	COM
Cr-Zn	+	–	–	COM
Cu-Hg	+	–	–	COM1
Cu-Nb	–	+	+	IMM
Fe-Os	+	–	+	SSH
Fe-Ru	+	–	+	SSH
Hf-V	+	–	–	COM
Hg-Mn	+	–	–	COM
Hg-Ni	+	–	–	COM
Hg-Zn	+	–	–	COM
Mn-W	–	+	–	IMM
Mn-Y	+	–	–	COM
Mo-Sc	–	+	+	SSL
Mo-Tc	+	–	–	COM
Ni-Re	–	+	–	SSL
Os-Pt	+	–	+	SSH
Re-W	+	–	–	COM
Sc-Y	+	–	+	SSH
Tc-W	+	–	–	COM
Ti-Zr	+	–	+	SSH
V-Zr	+	–	–	COM1

Table 1. The binary alloy systems disagreed between phase diagram information and high throughput first principles (HTEFP) calculations. The present choice is based on the estimation on the reliability of available information of phase diagram and thermodynamic data and elemental similarity. Plus symbol represents the immiscible systems, and minus symbol indicates the miscible or compound-formation systems. *Note: SSL: low solid solubility; SSH: high solid solubility; COM: formation of multiple compounds; COM1: only one compound reported; IMM: immiscibility gap.

The atomic radius of a chemical element is usually the mean or typical distance from the nucleus to the boundary of the surrounding cloud of electrons. Since the boundary is not well-defined, there are various non-equivalent definitions of atomic radii. Four widely used definitions are the Van der Waals radius, ionic radius, covalent radius and metallic radius. Among these definitions, the metallic radius proposed by Teatum^{43–45} is probably the most useful and widely accepted for metallic alloys, and will be adopted in the present work.

The third group of parameters come from the work by Villars, who analyzed the behavior of alloy systems on the properties of the component elements³⁵. He used a systematic elimination procedure to find atomic properties which could be used to distinguish the crystal structures of intermetallic compounds. 182 sets of tabulated physical properties and calculated atomic properties were considered. The best separation was obtained using three dimensional maps with the following variables as coordinates: (1) the mean number of valence electrons of the two elements, (2) the electronegativity difference proposed by Martynov-Batsanov in 1980^{35,46}, and (3) the difference of Zunger's pseudopotential radii sum^{35,47}. In view of the success of Villars' structure map, the two critical parameters of Martynov-Batsanov electronegativity and Zunger's pseudopotential radii sum are adopted in the present study.

The fourth group of parameters is derived from the famous Miedema's model because of its success in the description of the energy effects during alloying. Positive (immiscible) and negative (miscible) systems can be

separated by two parameters of the constituent chemical elements: the chemical potential for electronic charge (electronegativity), Φ and the electron density at the boundary of the Wigner-Seitz atomic cell, N_{ws} . Miedema's model became very popular due to the scarcity of experimental data, and the model provided a semi-quantitative evaluation. By assigning the two parameters as two coordinates it was possible to build a map in which a clear separation was observed between all binary alloys with positive heats of formation and those with negative values^{5,12–14}. Additional justification will be provided later, in section 3.5. A quantum-mechanical interpretation of Miedema's parameters has been proposed by Chelikowsky and Phillips⁴⁹.

The last criterion is based on Pettifor's work on the chemical scale (or Mendeleev number, M)⁴⁸. This has been set up by ordering the elements along a single axis so that the Mendeleev-type features of the periodic table are preserved. The Mendeleev numbers start with the least electronegative elements and end with the most electronegative ones. By using them, an excellent separation of similar structures is achieved for numerous A_mB_n phases with a given stoichiometry within M_A - M_B maps. Although the chemical scale is entirely phenomenological and has no a priori significance other than that it orders the elements relative to each other, Pettifor has shown that the corresponding two-dimensional structure maps achieve excellent structural separation. It will be shown in the following sections that the progressive order of chemical scale provides an excellent pattern clustering in the miscibility map of binary alloy systems.

Table S1 (in the Supplemental Materials) lists the adopted values of Pauling electronegativity (Pen)^{16, 41, 42}, Teatum metallic radii of elements with coordinate number CN12 (R_{ij})^{43–45}, Martynov-Batsanov electronegativity (Men)^{35, 46}, Zunger's pseudopotential radii sum (R_{sp})^{35, 47}, Miedema's work function (Φ)^{5, 14}, Miedema's electron density at the Wigner-Seitz cell boundary (N_{ws})^{5, 14}, Miedema's molar volume (V_m)^{5, 14}, the Mendeleev number or the Pettifor chemical scale (P_c)⁴⁸, and atomic number (AN) for ordering the 44 elements (27 transition metals and 17 Lanthanides). We refer the reader to the original publications on these parameters for the detailed physical meaning and their derivations.

The miscibility map sorted by different descriptors. We use the descriptors (“parameters”) defined in the previous section to analyze the pattern as it appears in the two-dimensional miscibility maps. Figure S1 (in the Supplemental Materials) shows the miscibility map ordered by Atomic number, Pauling electronegativity, Teatum metallic atomic radii, Martynov-Batsanov electronegativity, and Zunger's pseudo radii sum, respectively. These physical descriptors belong to the first three groups of variables identified in the preceding section. It can be seen that the atomic number, Pauling electronegativity, Teatum metallic atomic radii, Martynov-Batsanov electronegativity, and Zunger's pseudo radii sum can distinguish (“cluster”) the immiscible systems distinctly. This indicates that these parameters provide elemental similarity and the underlying physics. However, the clustering patterns ordered by Pauling electronegativity, Teatum metallic atomic radii, Martynov-Batsanov electronegativity, and Zunger's pseudo radii, do not provide any better feature identification than that by atomic number. Interestingly, the descriptors belonging to the fourth and fifth group of variables are found to divide the immiscible systems within clusters much better than those in the former three criteria. Figure S2 (in the Supplemental Materials) shows the clustering feature of immiscible alloy systems ordered by Miedema's electronegativity, Miedema's electron density, Miedema's molar volume, and especially by Mendeleev number or the Pettifor chemical scale. Although Miedema's two parameters were shown to be a great success for predicting the formation enthalpies of intermetallic compounds, no single one provides a better ordering sequence than those according to criteria 1, 2 and 3. We further see that the Pettifor chemical scale shows the best separation between the miscible and immiscible systems and the distinct clustering of immiscible systems.

After an in-depth analysis of the pattern feature ordered by Mendeleev number (Pettifor chemical scale), we found that some outliers appear to break the continuous clustering feature for the lanthanide-based systems (see the two green points circled by blue squares in Fig. 1a). In order to avoid this breaking of continuity of the distribution of immiscible systems, we slightly modify the order of rare earth/lanthanide elements. Figure 1b shows the modified miscibility map ordered by the modified Mendeleev number or the Pettifor chemical scale with the new elemental ordering as: Yb < Eu < Lu < Tm < Er < Ho < Dy < Tb < Gd < Sm < Pm < Nd < Pr < Ce < La < Y < Sc < Zr < Hf < Ti < Nb < Ta < V < Mo < W < Cr < Tc < Re < Mn < Fe < Os < Ru < Co < Ir < Rh < Ni < Pt < Pd < Au < Ag < Cu < Hg < Cd < Zn. It is seen from Fig. 1b that the new elemental sequence clearly divides the immiscible systems into two major clustered regions. Such clustering can be understood via similarity and dissimilarity of the elements in physics and chemistry.

Prediction of unknowns via artificial neural network algorithm. We now apply the incomplete miscibility maps (Fig. 1b) and the radial basis function artificial neural network (RBF-ANN) algorithm as described in Methods section to predict the unknown classification of binary alloy systems in order to obtain a complete miscibility map. Figure 2a shows the filled miscibility map of binary alloy systems ordered by the slightly modified Mendeleev number or the Pettifor chemical scale based on the RBF-ANN analysis. The three combinations are binary transition metal alloys (TM + TM), a combination of transition metals and rare earth metals (TM + RE), and a binary combinations of rare earth metals (RE + RE). Two immiscible regions are located at TM + RE, and TM-TM combinations as seen in Fig. 2a.

It is seen that the prediction is mostly in agreement with the empirical prediction based on simple elemental similarity, with a few outliers for Tc-based systems. The agreement between the elemental similarity and the RBF-ANN analysis can be summarized as follows:

- (1) For Pm-TM binary alloy systems: based on the similarity between Pm, Sm and Nd, one may see that the transition metals Zr < Hf < Ti < Nb < Ta < V < Mo < W < Cr are immiscible with Pm, whereas Pm is miscible with Tc < Re < Mn < Fe < Os < Ru < Co < Ir < Rh < Ni < Pt < Pd < Au < Ag < Cu < Hg < Cd < Zn.

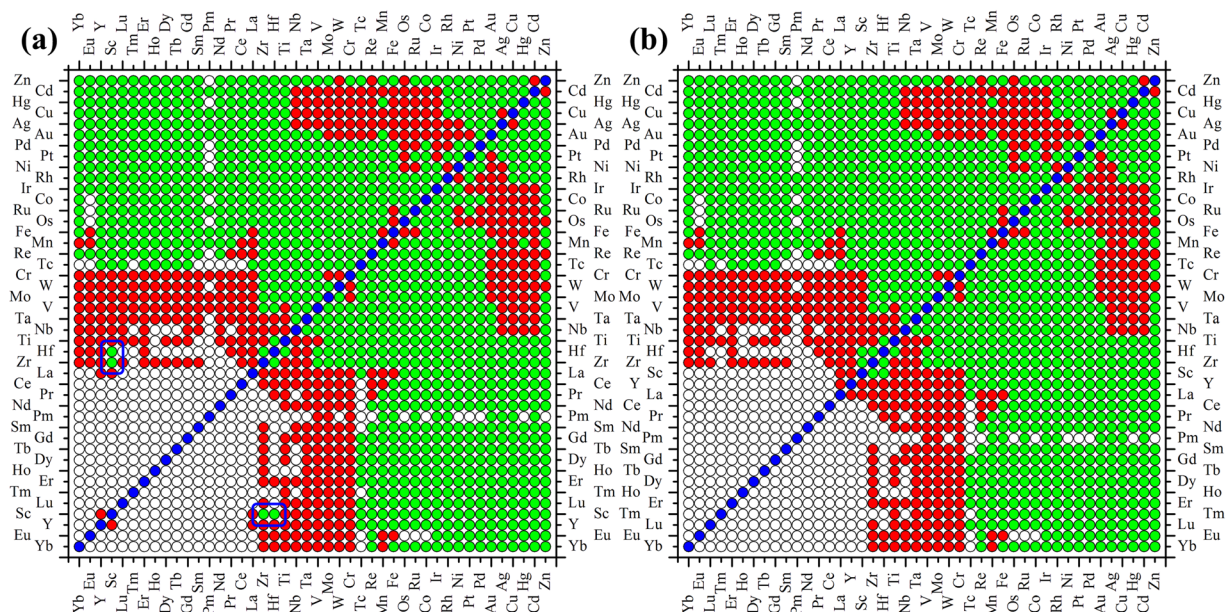


Figure 1. The incomplete miscibility maps of binary alloy systems ordered by (a) the original Pettifor chemical scale and (b) the slightly modified ordering the RE elements. The red and green symbols indicate immiscible and miscible systems, respectively. The white symbols represent the systems with the unavailable experimental information and are not considered in the present study, while the blue ones indicate the boundary of alloys consisting of dissimilar elements.

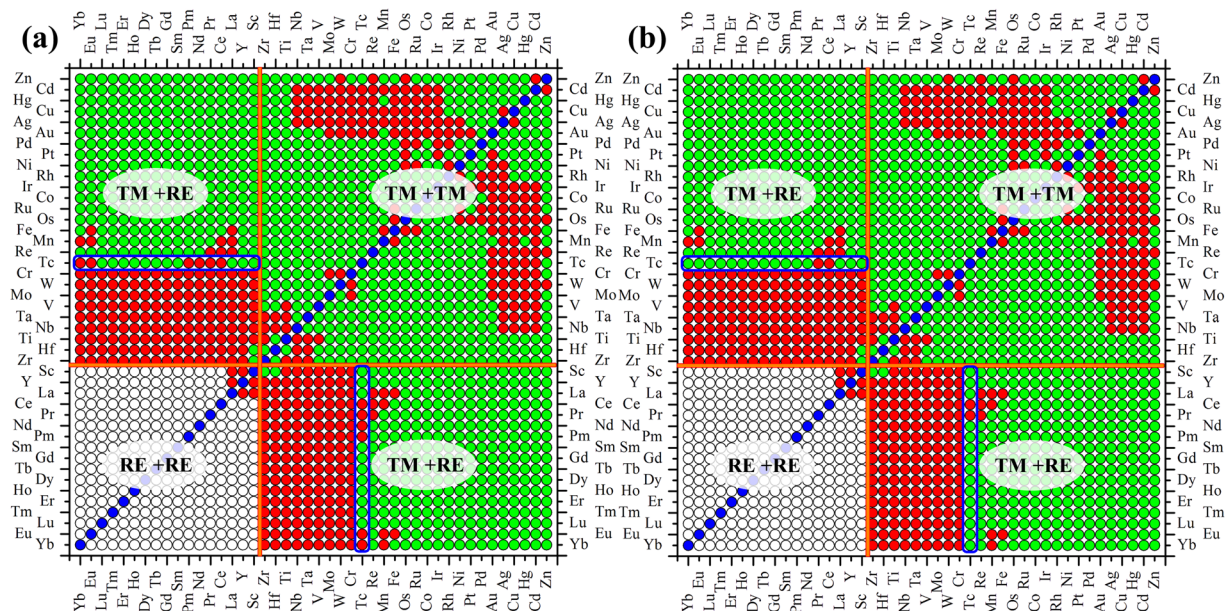


Figure 2. The complete miscibility map of binary alloy systems based on the slightly modified Pettifor chemical scale based on (a) the RBF-ANN analysis, (b) the RBF-ANN analysis plus elemental similarity between Re and Tc.

- (2) For binary alloy systems between elements consisting of $Zr < Hf < Ti < Nb < Ta < V < Mo < W < Cr$ and $Yb < Eu < Lu < Tm < Er < Ho < Dy < Tb < Gd < Sm < Pm < Nd < Pr < Ce$, the immiscibility is readily chosen for these classes of binary alloy systems based on the elemental similarity.
- (3) For $Eu-(Co, Ru, Os)$ binary alloy systems: based on the elemental similarity between $Rh(Ir)$ and $Ru(Os)$, a miscible alloying behavior is an appropriate choice for the three $Eu-(Co, Ru, Os)$ binary alloy systems.

Because the binary Tc-based alloy systems are located at the boundary of miscible and immiscible systems, one needs to further justify the miscibility of the related outliers based on the elemental similarity. Although our

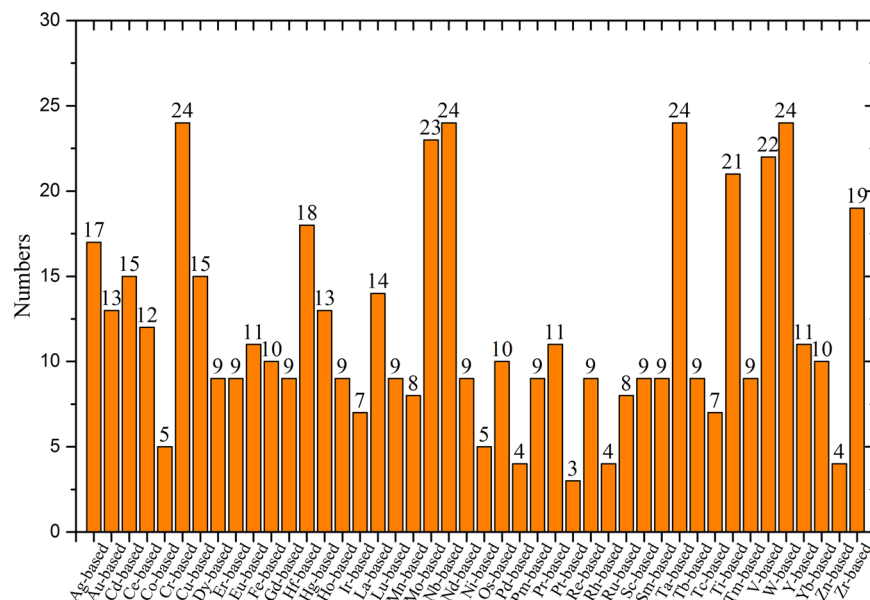


Figure 3. The numbers of immiscible binary alloy systems of different M-based binary systems (M is transition metals or rare earth metals). Significant feature is found that Cr, Nb, Ta, W, Mo -based systems possess the most number of immiscibility.

RBf-ANN analysis suggests that Tc-Eu, Tc-Yb, Tc-Pm and Tc-Nd systems are mostly immiscible, the elemental similarity between Tc and Re give an opposite conclusion, i.e. these systems should be miscible (negative formation enthalpy). Based on the similarity of valence electrons between Re and Tc, we would classify the four systems as miscible ones, and accordingly, Fig. 2b gives the complete miscibility map. Table S2 (in the Supplemental Materials) lists all of the 262 immiscible binary alloy systems identified from the 813 candidates. Figure 3 shows the numbers of the immiscible binary alloys formed from different transition or rare earth metals M. It is seen that Cr, Nb, Ta, W, Mo -based binaries have the largest number of immiscible elements.

Comparison with Miedema's theory and high throughput first principles methods. Miedema's theory is generally regarded as the most successful empirical model to describe the energy effects during alloying. The electronegativity difference and electronic density discontinuity, used as parameters, can be regarded as a 2-dimensional (2D) basis for alloy behavior since it can separate the positive and negative systems with a reasonable precision. As shown in Fig. S2 (in the Supplemental Materials), however, the single parameter of Miedema's model cannot separate unambiguously the immiscible systems into large clusters. However, assigning two coordinates to each transition element it was possible to separate all binary alloys with positive heats of formation from those with negative ones. The existence of intermetallic compounds in the binary phase diagram indicates a negative formation enthalpy. The same applies to the mutual solute solubility. A negative formation enthalpy is expected if one or more intermetallic compounds occur in the phase diagram, or if there is an appreciable solid solubility between the given elements. If neither condition is satisfied, a positive formation enthalpy is expected. Figure 4a shows the miscibility map of binary alloy systems ordered by the slightly modified Mendeleev number or the Pettifor chemical scale based on Miedema's theory. Compared with Fig. 2b (or Fig. S3 in the Supplemental Materials), it is clear that the majority of the clustering region matches to each other with only minor disagreements, validating the success of Miedema's theory. We recall the equation of Γ parameter for formation enthalpy in the form^{5,14}:

$$\Gamma = \frac{1}{(n_{ws}^{-1/3})_{av}} \{-P(\Delta\Phi^*)^2 + Q(\Delta n_{ws}^{1/3})^2\} \quad (1)$$

Accordingly, the value of $\Gamma = 0$ should separate the positive (immiscible) and negative (miscible) systems. A zero value of Γ means that

$$-P(\Delta\Phi^*)^2 + Q(\Delta n_{ws}^{1/3})^2 = 0 \quad (2)$$

Thus, a straight line in a plot between $\Delta\Phi^*$ versus $\Delta n_{ws}^{1/3}$ should separate regions with positive and negative formation enthalpy.

Figure 5 shows such plot for a large number of TM-TM and TM-RE alloys. Indeed the straight line $y = x/9.4$ separates two regions: one for binary alloy systems with negative formation enthalpy (green minus symbols) and one for systems with positive formation enthalpy (red plus symbols). The straight line gives the empirical value of the ratio $Q/P = 9.4$. As seen in Fig. 5a, Miedema's model provides a clear distinction between positive and negative systems when their formation enthalpy is sufficiently large. The agreement between the prediction and our

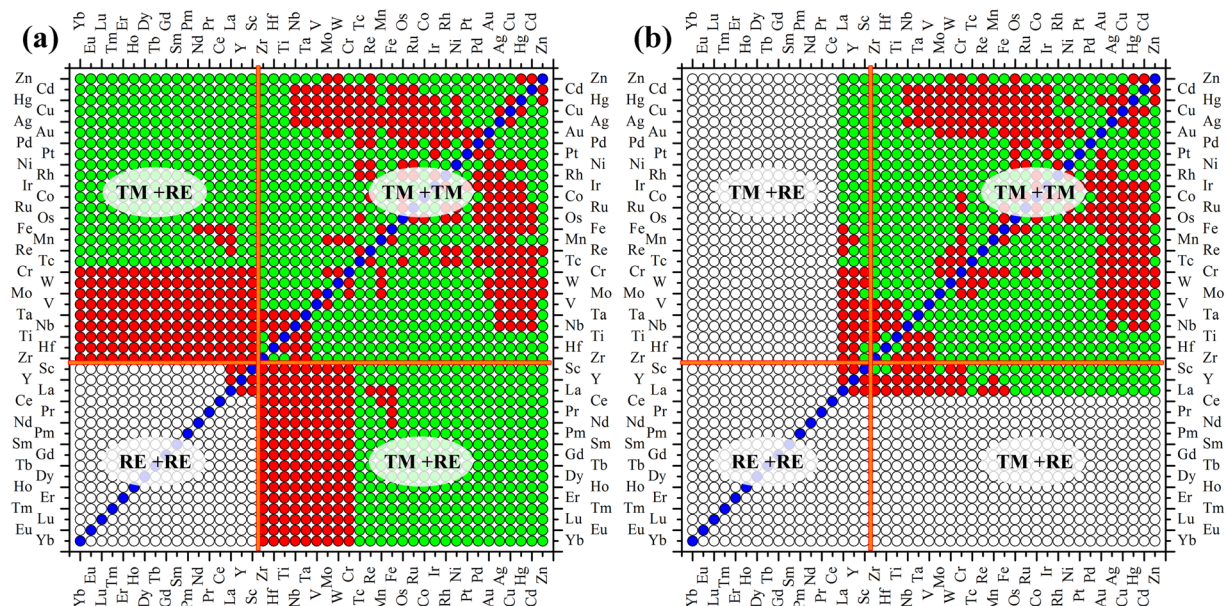


Figure 4. The miscibility map of binary alloy systems ordered by the slightly modified Pettifor chemical scale based on (a) Miedema's theory, and (b) HTFP calculations. Three different regions are divided as: TM+TM indicates binary transition metal alloy systems, TM+RE for the combinations of transition metals and rare earth metals, RE+RE for the combinations of binary rare earth metals.

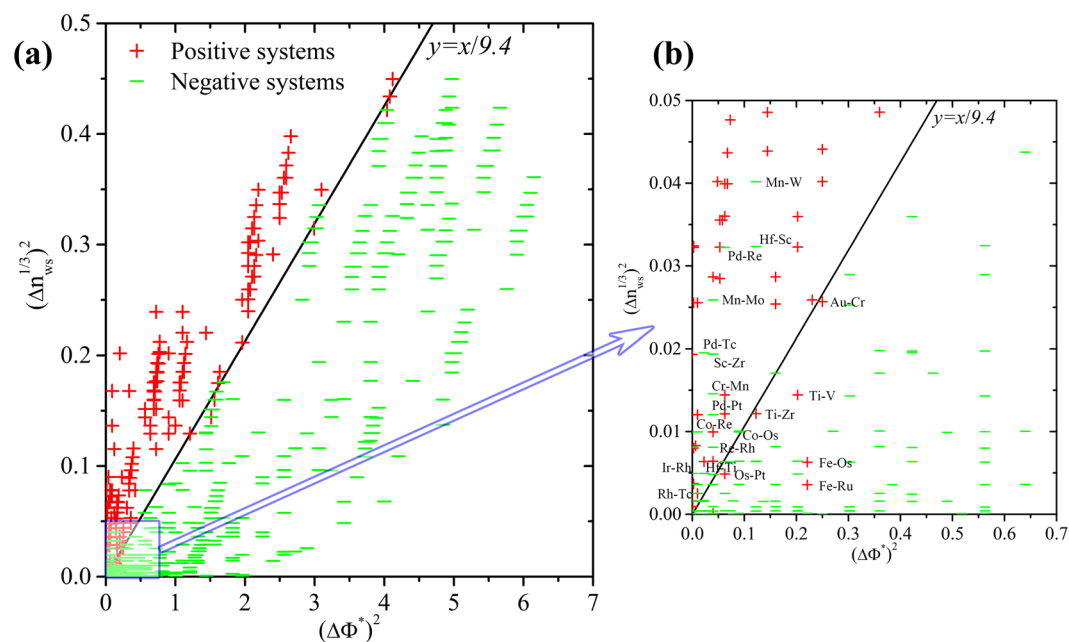


Figure 5. (a) The separation map of miscible and immiscible systems coordinated by the two Miedema's parameters as two-dimensional axes. Plus symbols '+' represents the immiscible systems, while minus symbols '-' stands for the miscible systems with negative formation enthalpy, i.e. no intermetallic compounds exist or solubilities are generally smaller 10%. (b) The zoomed-in region with $(\Delta\Phi^*)^2$ from 0 to 0.7 and $(\Delta n_{ws}^{1/3})^2$ from 0 to 0.05.

data mining results is beyond 94%, which corresponds to large values of $\Delta\Phi^*$ or $\Delta n_{ws}^{1/3}$. However, when we zoom in the region with small values of $(\Delta\Phi^*)^2$ from 0 to 0.7 and $(\Delta n_{ws}^{1/3})^2$ from 0 to 0.05, as seen in Fig. 5b, the separation between the "positive" and "negative" systems is not unambiguous. This illustrates the aforementioned weakness of Miedema's model when it is used for binary alloy systems with formation enthalpies close to zero. A further demonstration of the correlation between the miscibility and formation enthalpies by Miedema's theory

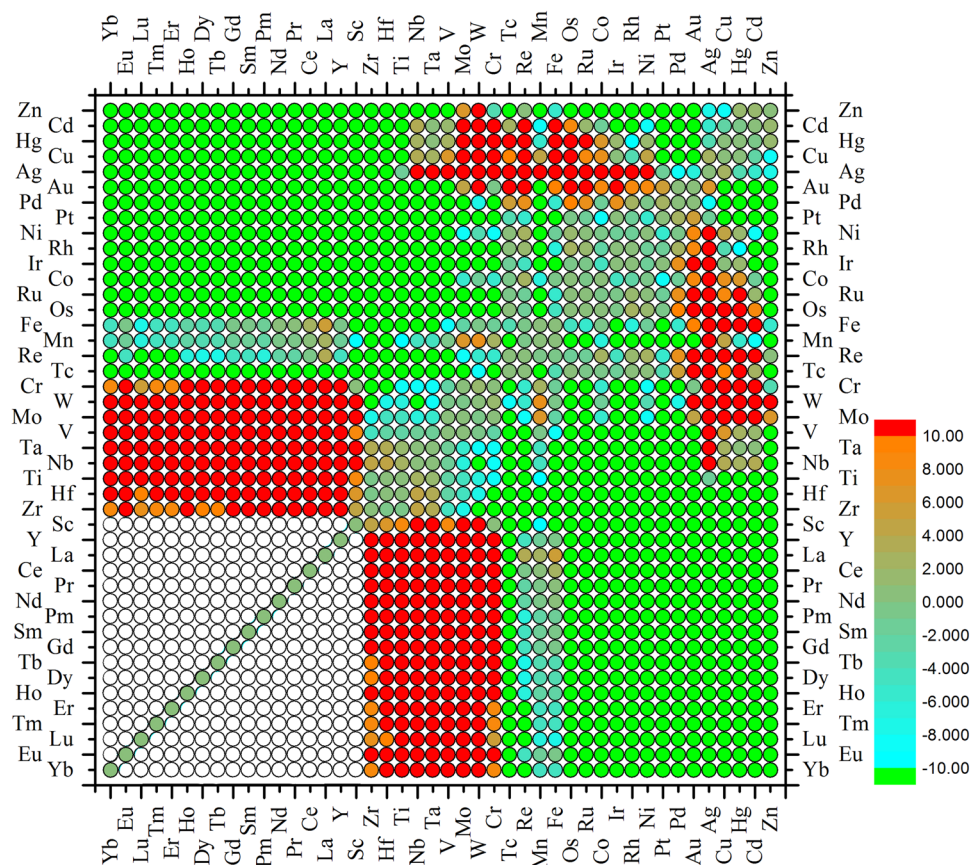


Figure 6. The distribution map of formation enthalpies of binary alloy systems with 1:1 stoichiometry calculated by Miedema's model to indicate the distribution regions of highly miscible and immiscible systems. Red and green colored circles correspond to the positive formation enthalpy larger than 10kJ/mol and negative ones lower than -10kJ/mol respectively. The white symbols represent the systems with the unavailable experimental information and are not considered in the present study. The elemental sequence is ordered by the slightly modified Pettifor chemical scale.

is shown in Fig. 6. It shows clearly that the distributions of highly positive ($>10\text{kJ/mol}$) and negative ($<-10\text{kJ/mol}$) binary alloy systems calculated by Miedema's model correspond to the regions of highly miscible and immiscible systems classified in the present studies (see Fig. 2b).

Table 2 lists the 46 immiscible binary systems that exhibit disagreement between the results of our data mining-generated map and Miedema's theory. It should be noted that our choice is based on the estimation on the reliability and validity of the available information and similarity of different elements. The inconsistency between our choice and Miedema's theory is about 5%. In view of the uncertainty of each method, these systems warrant further careful study, both experimentally and theoretically.

For comparison, we have also constructed the miscibility map based on the HTFP data shown in Fig. 4b. This choice of the miscibility properties is based on the results of HTFP data by Curtarolo *et al.*^{25, 32–34, 36–40}. One can see that there is some disagreement ($\sim 10\%$) with the results in Figs 2b and 4b. This implies that much more future work based on HTFP calculations is needed, and our present work shows the direction. For instance, further HTFP calculations based on a larger structure prototype library, e.g. ICSD or PCD, may help to validate the present results, although it is much computationally expensive for the TM-RE and RE-RE systems.

Discussion

The data mining presented in this paper (and in the prior cited studies⁴) provides a unique approach to discover novel materials, their properties and the supporting physical mechanisms. Our present study demonstrates a simple and illustrative example of obtaining the miscibility map of binary alloy systems from a large amount of raw data sources. It also provides a general pathway for one-, two- and three-dimensional spaces. The basic idea of data mining is not limited to the binary alloys, and therefore one may extend it for multicomponent alloy systems. To be noted that Miedema's theory is also applicable for ternary, quaternary and multicomponent systems via geometrical methods²⁰, while the HTFP method can also be readily applied for the ternary and even higher-order alloys when their structures are known or can be predicted. The success of Miedema's model and the HTFP method is clearly revealed, and some outliers are identified for further experimental and theoretical investigation. Because of the scarcity of reliable experimental and first-principles data for binary systems consisting of dissimilar lanthanides, it is of particular urgency to perform HTFP calculations on this group of systems.

Systems	Alloying behavior		Systems	Alloying behavior	
	Present	Miedema(kJ/mol)		Present	Miedema(kJ/mol)
Au-Cr	+	-0.2	Ir-Rh	-	+0.8
Au-Os	-	+21.8	Mn-Mo	-	+6.0
Cd-Co	+	-1.8	Mn-W	-	+7.6
Cd-Ir	+	-12.6	Mn-Yb	+	-4.2
Cd-Mn	+	-8.1	Mo-V	-	+0.0
Ce-Fe	-	+2.5	Mo-Zn	-	+6.3
Ce-Re	+	-0.1	Nb-Ta	-	+0.0
Co-Os	-	+0.1	Ni-Re	-	+2.8
Co-Re	-	+2.5	Ni-Tc	-	+0.7
Cr-Mn	-	+2.6	Os-Pt	+	-0.6
Cu-Hg	-	+0.4	Os-Rh	-	+2.5
Cu-Ni	-	+4.3	Os-Ru	-	+0.1
Eu-Fe	+	-1.0	Os-Tc	-	+0.2
Eu-Mn	+	-1.6	Os-Zn	+	-13.7
Fe-Nd	-	+0.7	Pd-Pt	-	+2.4
Fe-Os	+	-4.9	Pd-Re	-	+7.9
Fe-Pr	-	+0.7	Pr-Re	+	-2.5
Fe-Ru	+	-5.6	Re-Rh	-	+1.2
Hf-Sc	-	+6.2	Re-Tc	-	+0.2
Hf-Ti	-	+0.2	Rh-Ru	-	+1.5
Hg-Ni	-	+1.1	Sc-Zr	-	+4.9
Hg-Zn	-	+1.2	Ti-Zr	+	-0.3
Tc-Ce	+	-25.9	Tc-Pr	+	-28.2

Table 2. The binary alloy systems whose miscibility shows disagreement between present data mining and Miedema's theory. The present choice is based on the estimation on the reliability of the available information and elemental similarity. Plus symbol represents the immiscible systems, and minus symbol indicates the miscible or compound-formation systems.

Methods

Data collection and classification rule. A total of 946 combinations exist for binary alloy phase diagrams consisting of transition and lanthanide metals, of which about 800 are available in the Landolt-Boernstein Database⁷, in the “Desk handbook: Phase diagrams for binary alloys” compiled by Okamoto⁹, and in the ASM alloy phase diagrams database organized by the ASM community⁸. In addition, the experimental standard formation enthalpies determined by Kleppa *et al.*^{24–35} provide a further validation of the compound-formation systems. The recent computational binary alloy project, hosted in the *aflowlib.org* consortium repository⁴⁰, contains the formation enthalpies for hundreds of thousands of intermetallics comprising all of the transition metal systems. However, this repository does not include any alloying data between transition and lanthanide metals, and therefore a throughout data mining is much demanded for further design of novel alloys.

The binary solid phases consisting of elements A and B, may be classified into three groups: mechanical mixtures, solid solutions, and ordered intermetallic compounds. The mechanical mixture corresponds to binary systems with a positive formation enthalpy, while the intermetallic compound is favored for a negative formation enthalpy. The solid solution has three subtypes: clustered solid solution, disordered solid solution, or ordered solid solution. For the former two cases, the different kinds of atoms are energetically preferable to mix, while for the last case the different kinds of atoms prefer to segregate into different clusters. In comparison to the formation of intermetallic compounds with a negative formation enthalpy, the first two cases may be classified as poorly miscible systems, while the latter case can be classified as poorly immiscible systems since the formation enthalpy is close to zero²⁰.

With this classification rule of different solid phases and the five alloy databases at hand, a data mining scheme is proposed to separate the immiscible systems from those miscible ones in which the homogeneous disordered solid solution and intermetallic compounds may form. Figure 7a shows the proposed flowchart of the classification scheme that we shall use. Note that the miscible systems are those in which one or more intermetallic compounds, or one or more ordered solid phases, or a disordered solid solution exist with negative formation enthalpy. The immiscible systems correspond to those with positive formation enthalpy, in which case either a mechanical mixture or clustered solid solution may form at low temperature.

Artificial neural network algorithm for prediction. We use a radial basis function artificial neural network (RBF-ANN) algorithm to predict the immiscibility of an unknown system. The RBF-ANN has a three-layer structure: an input layer, a hidden layer and a linear output layer, as shown in Fig. 7b. In this work, the input and output of RBF-ANN are the map coordinates e.g. the “X- and Y-coordinate” in Fig. 7b and the miscible/immiscible classification, respectively. The input layer only transfers input data to the hidden layer. The hidden layer uses

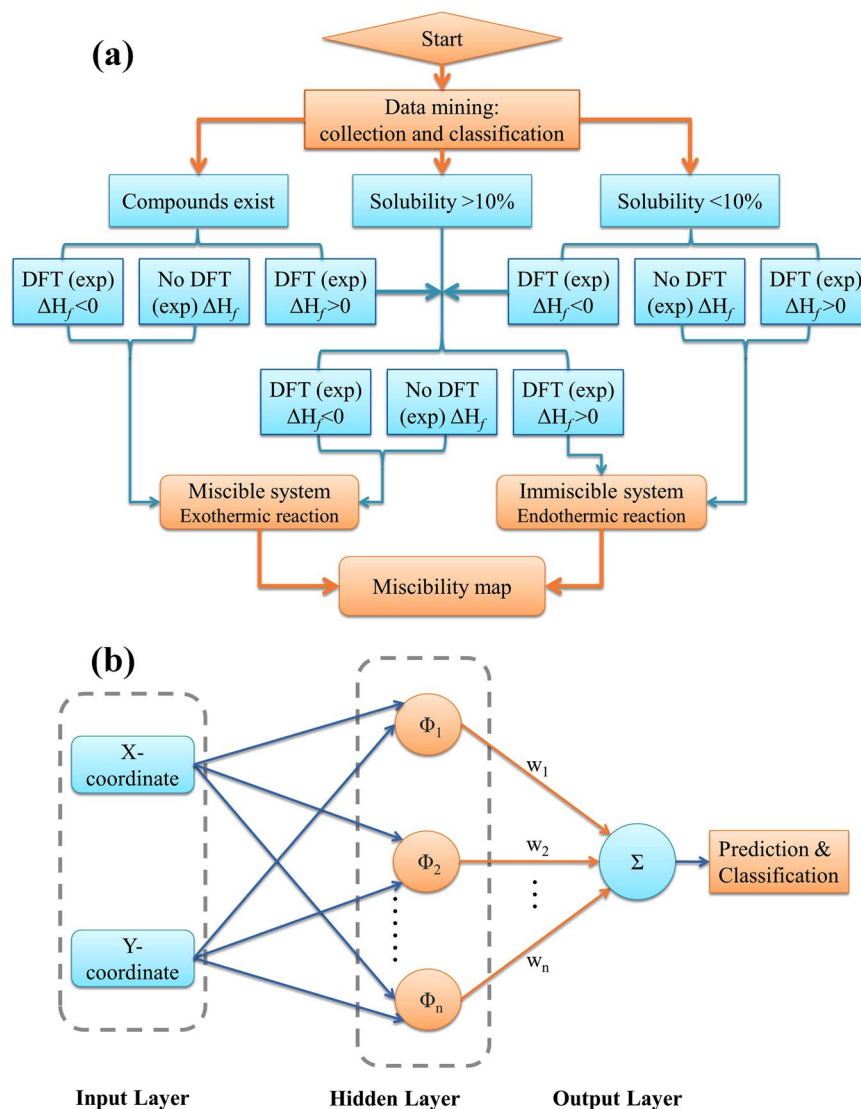


Figure 7. (a) A classification flowchart used in data mining to distinguish the miscible and immiscible systems based on the binary alloy phase diagrams, thermodynamic data and high-throughput first principles (HTFP) calculations. (b) A schematic diagram of radial basis function artificial neural network (RBF-ANN) for the prediction of the unknown systems.

Gaussian functions as the radial basis function Φ . It can produce a localized response to the input and introduce non-linearity into the network⁵⁰.

$$\Phi_i(X) = \exp\left[-\frac{\|X - c_i\|^2}{2\sigma_i^2}\right], \quad i = 1, 2, \dots, n \quad (3)$$

Where X is the input vector, c_i is the centroid of the basis function of the i^{th} hidden layer node, and σ_i is the variance of basis function.

Before the RBF-ANN is trained, all data are scaled to fit within the interval $[-1,1]$. During the training processes, the center vectors c_i and variances σ_i of basis functions in the hidden layer are determined using k-means clustering⁵¹. First the k samples are randomly chosen as the initial values of the center vectors c_i , and the other samples are classified to form k subclasses according to the Euclidean distance between samples and center vectors. The new center vectors (i.e. the mean values of subclass data) are continually calculated by an iterative technique until they do not change. After the center vectors are determined, the variances σ_i can be obtained by the maximum distance among the center vectors. Subsequently, the weights w_i between the hidden and output layers (see Fig. 7b) are modified by the least mean square algorithm. The training of the network is completed after the number of neurons in the hidden layer reaches the sample dataset size ($n = 770$). The mean square errors of training are 4.7×10^{-2} . We then use this network topological structure to predict the classification of unknown alloy systems.

References

- Kalil, T., Wadia, C. *Materials Genome Initiative for Global Competitiveness* (2011), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf (Date of access: 26/04/2017)
- Curtarolo, S. *et al.* The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
- Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
- Rajan, K. Data mining and inorganic crystallography. in *Data Mining In Crystallography* (ed. Hofmann, D. W. M.) 89–134 (Springer, (2009).
- Miedema, A. R., Niessen, A. K., De Boer, F. R., Boom, R. & Matten, W. C. M. *Cohesion in metals: transition metal alloys*. (North-Holland Publishing, (1989).
- Liu, B. X., Lai, W. S. & Zhang, Q. Irradiation induced amorphization in metallic multilayers and calculation of glass-forming ability from atomistic potential in the binary metal systems. *Mater. Sci. Eng. R-Rep.* **29**, 1–48 (2000).
- Effenberg, G. (ed.), Ilyenko, S. (ed.) *Landolt-Boernstein: Group IV Physical Chemistry* (Springer Berlin Heidelberg, 2008)
- Villars, P. (editor-in-chief), Okamoto, H. & Cenzual, K. (section editors). *ASM Alloy Phase Diagrams Database* (2016), <http://mio.asminternational.org/apd/> (Date of access: 26/04/2016).
- Okamoto, H. *Phase diagrams for binary alloys*. (ASM International, 2010).
- Zhang, R. F., Wang, J., Beyerlein, I. J., Misra, A. & Germann, T. C. Atomic-scale study of nucleation of dislocations from fcc-bcc interfaces. *Acta Mater.* **60**, 2855–2865 (2012).
- Zhang, R. F., Germann, T. C., Liu, X. Y., Wang, J. & Beyerlein, I. J. Layer size effect on the shock compression behavior of fcc-bcc nanolaminates. *Acta Mater.* **79**, 74–83 (2014).
- Miedema, A. R. & De Chatel, P. F. Proceedings of the Symposium on Theory of Alloy Phase Formation. In *Theory of alloy phase formation* (ed. Lawrence Herman Bennett) (New Orleans, 1979).
- Miedema, A. R., De Chatel, P. F. & De Boer, F. R. Cohesion in alloys-fundamentals of a semi-empirical model. *Physica B+C* **100**, 1–28 (1980).
- Bakker, H. Enthalpies in alloys: Miedema's semi-empirical model. *Trans. Tech. Publications*, 1–78 (1998).
- Brandl, C., Germann, T. C. & Misra, A. Structure and shear deformation of metallic crystalline–amorphous interfaces. *Acta Mater.* **61**, 3600–3611 (2013).
- Pauling, L. *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*. Vol. 18 (Cornell university press, 1960).
- Hume-Rothery, W., Mabbott, G. W. & Evans, K. M. C. The freezing points, melting points, and solid solubility limits of the alloys of silver, and copper with the elements of the B sub-groups. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **233**, 1–97 (1934).
- Waber, J. T., Gschneidner, K. Jr, Larson, A. C. & Prince, M. Y. Prediction of solid solubility in metallic alloys. *Trans. Metall. Soc. AIME* **227**, 717–723 (1963).
- Darken, L. S. & Gurry, F. W. In *Theory of alloy phase formation* (ed. Lawrence Herman Bennett) (New Orleans, 1979).
- Zhang, R. F., Zhang, S. H., He, Z. J., Jing, J. & Sheng, S. H. Miedema Calculator: A thermodynamic platform for predicting formation enthalpies of alloys within framework of Miedema's Theory. *Comput. Phys. Commun.* **209**, 58–69 (2016).
- Zhang, R. F. & Liu, B. X. Proposed model for calculating the standard formation enthalpy of binary transition-metal systems. *Appl. Phys. Lett.* **81**, 1219–1221 (2002).
- Zhang, R. F., Sheng, S. H. & Liu, B. X. Predicting the formation enthalpies of binary intermetallic compounds. *Chem. Phys. Lett.* **442**, 511–514 (2007).
- Zhang, R. F. & Rajan, K. Statistically based assessment of formation enthalpy for intermetallic compounds. *Chem. Phys. Lett.* **612**, 177–181 (2014).
- Guo, Q. T. & Kleppa, O. J. The standard enthalpies of formation of the compounds of early transition metals with late transition metals and with noble metals as determined by Kleppa and co-workers at the University of Chicago—A review. *J. Alloy. Compd.* **321**, 169–182 (2001).
- Meschel, S. V., Chen, X. Q., Kleppa, O. J. & Nash, P. The standard enthalpies of formation of some intermetallic compounds of early 4d and 5d transition metals by high temperature direct synthesis calorimetry. *CALPHAD* **33**, 55–62 (2009).
- Meschel, S. V. & Kleppa, O. J. Thermochemistry of some binary alloys of noble metals (Cu, Ag, Au) and transition metals by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **350**, 205–212 (2003).
- Meschel, S. V. & Kleppa, O. J. Thermochemistry of some binary alloys of silver with the lanthanide metals by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **376**, 73–78 (2004).
- Meschel, S. V. & Kleppa, O. J. Thermochemistry of some binary alloys of gold with the lanthanide metals by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **363**, 242–247 (2004).
- Meschel, S. V. & Kleppa, O. J. Thermochemistry of some binary alloys of copper with the lanthanide metals by high-temperature direct synthesis calorimetry. *J. Alloy. Compd.* **388**, 91–97 (2005).
- Meschel, S. V. & Kleppa, O. J. Thermochemistry of some binary alloys of Samarium with the noble metals (Cu, Ag, Au) by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **416**, 93–97 (2006).
- Meschel, S. V. & Kleppa, O. J. The standard enthalpies of formation of some intermetallic compounds of transition metals by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **415**, 143–149 (2006).
- Meschel, S. V., Nash, P. & Chen, X. Q. The standard enthalpies of formation of binary intermetallic compounds of some late 4d and 5d transition metals by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **492**, 105–115 (2010).
- Meschel, S. V., Nash, P., Gao, Q. N., Wang, J. C. & Du, Y. The standard enthalpies of formation of some binary intermetallic compounds of lanthanide–iron systems by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **554**, 232–239 (2013).
- Meschel, S. V., Nash, P., Gao, Q. N., Wang, J. C. & Du, Y. Standard enthalpies of formation of some Lanthanide–Cobalt binary alloys by high temperature direct synthesis calorimetry. *J. Alloy. Compd.* **578**, 465–470 (2013).
- Wang, J. W., Guo, Q. T. & Kleppa, O. J. Standard enthalpies of formation of some Th alloys with Group VIII elements (Co, Ni, Ru, Rh, Pd, Ir and Pt), determined by high-temperature direct synthesis calorimetry. *J. Alloy. Compd.* **313**, 77–84 (2000).
- Curtarolo, S., Morgan, D. & Ceder, G. Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys. *CALPHAD* **29**, 163–211 (2005).
- Curtarolo, S. *et al.* AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
- Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
- Pilania, G., Wang, C. C., Jiang, X., Rajasekaran, S. & Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci Rep* **3**, 2810 (2013).
- Curtarolo, S. *et al.* The aflow ab-initio binary alloy library (2016), <http://materials.duke.edu/apool.html> (Date of access: 11/03/2016).
- Allred, A. L. Electronegativity values from thermochemical data. *J. Inorg. Nucl. Chem.* **17**, 215–221 (1961).
- Huheey, J. E., Keiter, E. A. & Keiter, R. L. *Inorganic chemistry: principles of structure and reactivity*. (Harper Collins College Publishers, 1993).
- Teatum, E., Gschneidner, K. A. & Waber, J. *Compilation of calculated data useful in predicting metallurgical behavior of the elements in binary alloy systems*. Vol. 2345 (Los Alamos Scientific Laboratory of the University of California, 1960).

44. Cahn, R. W. & Haasen, P. *Physical metallurgy*. (North-Holland Publishing, 1996).
45. Teatum, E., Gschneidner, K. & Waber, J. Compilation of Calculated Data Useful in Predicting Metallurgical Behavior of the Elements in Binary Alloy Systems. (Los Alamos Scientific Lab., United States, 1968).
46. Martynov, A.I. & Batsanov, S. S. New approach to calculating atomic electronegativities. *Russ. J. Inorg. Chem.* **25**, 1737–1740 (1980).
47. Zunger, A. Pseudopotential Viewpoint of the Electronic and Structural Properties of Crystals. In *Structure and Bonding in Crystals* (ed. Michael O'Keeffe & Alexandra Navrotsky) 73 (Academic Press, 1981).
48. Pettifor, D. G. The structures of binary compounds. I. Phenomenological structure maps. *J. Phys. C: Solid State Phys.* **19**, 285 (1986).
49. Chelikowsky, J. R. & Phillips, J. C. Quantum-defect theory of heats of formation and structural transition energies of liquid and solid simple metal alloys and compounds. *Phys. Rev. B* **17**, 2453 (1978).
50. Moody, J. & Darken, C. J. Fast learning in networks of locally-tuned processing units. *Neural Comput.* **1**, 281–294 (1989).
51. Haykin, S. *Neural networks and learning machines*. (Prentice Hall, 2008).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NFSC) with Nos. 51471018 and 51672015, National Key Research and Development Program of China (Nos. 2016YFC1102500 and 2017YFB0702100), National Thousand Young Talents Program of China and Fundamental Research Funds for the Central Universities. D. L. acknowledges support by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project “IT4 Innovations National Supercomputing Center – LM2015070” and the project No. 17-23964 S of the Grant Agency of the Czech Republic. We would also like to thank Prof. Dr. S. Veprek for constructive comments and suggestions to the manuscript, and Prof. S. Curtarolo for his comments and providing online HTPF data.

Author Contributions

R.F.Z., X.F.K., S.H.Z. and S.H.S. performed the data collection, data mining, data verification and thermodynamic calculations. H.T.W. and S.S. performed the data analysis by radial basis function artificial neural network (RBF-ANN) algorithm and other approaches. D.L., K.R. and T.C.G. was involved in data analysis, verification and paper organization. All authors contributed the idea and participated in the scientific discussions, manuscript comments and corrections.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-09704-1](https://doi.org/10.1038/s41598-017-09704-1)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017