ISSN 0020-1685, Inorganic Materials, 2009, Vol. 45, No. 10, pp. 1077–1080. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © N.N. Kiseleva, 2009, published in Neorganicheskie Materialy, 2009, Vol. 45, No. 10, pp. 1157–1160.

Prediction of Occurrence of AB_3X_3 (X = S, Se, Te)

N. N. Kiseleva

Baikov Institute of Metallurgy and Materials Science, Russian Academy of Sciences, Moscow, Russia
e-mail: kis@ultra.imet.ac.ru

Received July 9, 2008

Abstract—For the first time, we consider the possibility of occurrence of AB_3X_3 composition compounds in the A_2X_3–B_2X (A and B are different elements; X = S, Se, Te) systems under standard conditions. For the prediction, we use an ad hoc information analysis system that combines the complex of databases on properties of inorganic substances and materials and programs for data analysis on the basis of computer training algorithms.

DOI: 10.1134/S002016850910001X

INTRODUCTION

The majority of compounds of the $A^{III}B^{I}_3X_3$ (X = S, Se, Te) composition are semiconductor materials [1–4]. Crystals of the compounds $AsAg_3S_3$ (proustite), $SbAg_3S_3$ (pyrargyrite), and $AsTl_3Se_3$ are also promising for application in nonlinear optics and optoelectronics [5]. The possibilities of using compounds of the $AsAg_3S_3$ and $AsTl_3Se_3$ compositions in acoustooptics are discussed in [5–7].

The aim of this work is to predict the possibility of formation of new $A^{III}B^{I}_3X_3$ (X = S, Se, Te) compounds.

PREDICTION APPROACH

At present, there are two basic approaches to prediction of new inorganic compounds: quantum-mechanical methods and empirical criteria. Despite the advance of physics in the interpretation of properties of solid phases, the efforts to calculate properties of complex inorganic substances on the basis of only the knowledge of properties of their components are inefficient, which is attributed to difficulties of quantum-mechanical calculations of multielectron systems. This was the reason for the development of empirical criteria for the formation of phases with certain properties with the goal of reducing the time and expenditures connected with experimental searches for new substances (for example, the known rules of Hume–Rothery, Laves, Matthias, Goldschmidt, and Darken–Gurry).

Usually, these criteria include only a few parameters of components. The advantages of these criteria are simplicity and clearness. However, the simplest empirical criteria are characterized by a significant disadvantage—incomplete classification of systems, which is attributed, as a rule, to the fact that many other component properties, probably being of no less significance, are not taken into account.

In order to search for classifying rules that cover a wide-ranging ensemble of properties of components of physicochemical properties, we propose to apply methods of computer-assisted data analysis on the basis of computer training algorithms [8]. This approach makes it possible to construct complex multidimensional classifying rules on the basis of processing of large-scale arrays of information from databases (DBs). Herein, the average error of the prediction with the use of formed criteria does not exceed 80% [8].

In order to computerize the procedure of search for multidimensional criteria and their use for prediction of new inorganic compounds, a special programming information analysis system (IAS) is developed [9]. This system comprises a complex of DBs on properties of inorganic compounds [8–10] and a subsystem of data analysis composed of computer training programs that are the most efficient for application in inorganic chemistry and materials science [11, 12].

PREDICTION OF NEW COMPOUNDS OF THE AB_3X_3 COMPOSITION

For computer-assisted analysis, we used information in the DB Phases, which is part of the IAS [8–10], with reference to 117 examples of formation of compounds of the AB_3X_3 (X = S, Se, Te) composition and 58 examples of the absence of compounds of this composition in the A_2X_3–B_2X systems under standard conditions (training sample). In order to describe the compounds in computer memory on the basis of physicochemical concepts on the nature of substances of this type, two subsystems of properties were selected: (i) the properties of the elements A, B, and X (Table 1), whose values were taken from our DB Elements (www.phases.imet-db.ru/elements), which is part of the IAS, and (ii) the properties of simple chalcogenides A_2X_3 and B_2X (standard entropy and enthalpy as well as

Table 1. Properties of elements used for the description of compounds of the AB_3X_3 composition

No.	Property	No.	Property
1	Melting temperature	14	3rd ionization potential (I_3)
2	Covalent radius (r_{cov})	15	Heat conductivity
3	Pseudopotential radius (according to Zunger)	16	Chemical potential according to Miedema (for elements A and B only)
4	Quantum number	17	Enthalpy of melting
5	Valence electron distance (according to Schubert)	18	Electronegativity (according to Pauling)
6	Core electron distance (according to Schubert)	19	Group number (for elements A and B only)
7	Boiling temperature	20	Molar heat capacity
8	Enthalpy of vaporization	21	Solid state entropy
9	Ionic radius (according to Bokii and Belov)	22	Debye temperature (for elements A and B only)
10	Enthalpy of atomization	23	$(r_{covA} - r_{covB})/r_{covC}$
11	Periodic number (according to Mendeleyev–Pettifor) (N_{MP})	24	$(I_{1A} - I_{1B})/I_{1C}$
		25	$(I_{2A} - I_{2B})/I_{2C}$
12	1st ionization potential (I_1)	26	$(I_{3A} - I_{3B})/I_{3C}$
13	2nd ionization potential (I_2)	27	$(N_{MP\,A} - N_{MP\,B})/N_{MP\,C}$

covalent radii and the covalent radius to metallic radius ratio for the elements A, B, and X).

For the data analysis, we used several computer training algorithms included in the developed IAS: the estimate calculation algorithm, the method of binary decision trees, the two-dimensional linear separators, the Fisher linear discriminant, the linear machine algorithm, the neural network training, the method of k nearest neighbors, the genetic algorithm, the method of support vector machine, the dead-end test algorithm, the statistically factored voting, *LoReg* [11], and *ConFor* [12].

In order to estimate the prediction accuracy, we used examinational recognition exercised for the training sample data in two modes: in the absence of sliding control and in its presence. In the first case, after computer training, physicochemical systems from the training sample were recognized. However, this technique of estimating the training quality and subsequent prediction is extremely unreliable. Therefore, we used the second technique of estimating the training quality and prediction—sliding control. In this case, the specified part (f) of objects—descriptions of physicochemical systems—is withdrawn from the training sample; computer training is carried out for the rest of the objects; the withdrawn objects are subjected to recognition. The error rate is recorded. Next, the "control" objects are returned to the training sample; other f of the objects are withdrawn from it. The procedure is repeated until the time that all the objects from the training sample have played the role of "control" ones. This procedure allows a more objective estimation of the accuracy of computer training algorithms for a concrete training sample, especially in the case of pattern recognition methods, for which the undesirable effect of "retraining" is observed, when the algorithm "becomes customized" to only the training sample and subsequently yields poor results for new objects. After the training in the sliding control mode is completed, the average error of the examinational recognition is calculated; subsequently, it is used for the selection of best algorithms for the collective decision making.

The aim of the application of the collective decision strategy is to improve prediction accuracy. The concept of this approach is as follows. First, for the prediction of new compounds, various computer training algorithms are used independently. Next, the results of performance of individual algorithms are processed in a special fashion [11], and a final collective decision is made with respect to attributing compounds to one class or another. In this case, errors of predictions with the use of individual algorithms most often cancel out, which allows more accurate results to be obtained in the majority of cases.

According to the results of the examinational recognition, the following computer training algorithms were selected:

(i) For the formation of classifying criteria comprising properties of elements (Table 1): the linear machine, *LoReg*, k nearest neighbors [11], and *ConFor* [12], for which the prediction accuracy is 85.7, 82.5, 85.7, and 95%, respectively;

(ii) For the search for classifying criteria comprising properties of simple chalcogenides: the linear machine, *LoReg*, Fisher linear discriminant, support vector machine [11], and *ConFor* [12], for which the prediction accuracy is 65.3, 64.5, 64.5, 68.6, and 93.1%, respectively.

Then, results of performance of these algorithms were used in various methods of collective decision making [11] (Bayesian method, detection of compe-

Table 2. Prediction of the ability to form compounds of the AB_3X_3 (X = S, Se, Te) composition

A / B	B	Al	P	Sc	Ti	Cr	Fe	Ga	As	Y	Rh	In	Sn	Sb	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Bi
X – S																														
Li	#1		1	1	#1	1	1	#2	#1			#2		#1	2	2	2	2												#2
Na	#1	#1	1	1	1	1	#1	#2	#1	1	#1	#1	1	#1					1	1	1	1	1	1	1	1	1	1	1	#1
K	#1	1	1	1	1	1	1	#2	#1	1	#1	1	1	#1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	#2
Cu	1	#2	#1	#1		#2	#1		#1	#1	1	#2	#1	#1	#2	#2	2	#2			#1	#1	#1	#1	#1	#1	#1	#1	#1	#1
Rb	#1	1	1	1	1	1	1		1	1	1	#1	1	1	1			1	1	1	1	1	1	1	1	1			1	#1
Ag		#2					#2	#2	#1			#2	#2	#1	#2	#2	#2	#2	2	2	2		2	#2	2					#2
Cs	#1	1	1	1	1	1	#1	#2			1	1	1	#2	1			1	1	1	1	1	1	1	1	1	1	1	1	
Tl	#1		#2	1	1	#1	1	2	#1		1	#1	#1	#1	#2	2	#2	2	2	2	2							1	1	
X – Se																														
Li	1		1	1	1	1	1	2	#2	1	1	2			2	2	2	2					1	1	1	1	1	1	1	
Na	1	#1	1	1	1	1	#1	#1	#1	1	1	1	1	#2									1	1	1	1	1	1	1	1
K	1	#1	1	1	1	1	#1	#1	#1	1	1	1	1	#1		1			1	1	1	1	1	1	1	1	1	1	1	#1
Cu	1		1	#1	1	1	1	#2	#1	#1	1	#2	#1	#1	2	2	2	2		#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1
Rb	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	#1
Ag	1		2			2	2	#2	#1	2		2		#2	2	2	2	2	2	2	2		2	2	2	2	2	2	2	2
Cs	1	1	1	1	1	1	1	#1	1	1	1	1	1	1	1				1	1	1	1	1	1	1	1	1	1	1	#1
Tl	#1		1	1	1	#1	1	2	#1	1	1	#2	#1	#1	2	2	2	2							1	1	1	1	1	#2
X – Te																														
Li	1			1	1		1	2		1	1				2	2	2	2	2						1	1	1	1	1	
Na	1	#1	1	1	1	1	1		1	1	1	1	1	#1				1	1	1	1	1	1		1	1	1	1	1	1
K	1	#1	1	1	1	1	1	#1	1	1	1	1	#1	#1	1				1	1	1	1	1	1	1	1	1	1	1	#1
Cu	1	#2		1	1	1	1	#2	#2	#1	1	#2		#2	2	2	2	2	2	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#2
Rb	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1
Ag	1		2	2		2	2	#2	2	2		#2	2	#2	2	2	2	2	2	2	2	#2	2	2	2	2	2			2
Cs	1	#1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1
Tl	1			1	1		1	2			1	2	#1	#2	#2	2	2	2	2											#2

tence regions and decision templates, Wood's dynamic method, method of logical correction, convex stabilizer algorithm, complex committee methods). The best results of examinational recognition for collective decision making [11] were obtained by the methods based on detection of decision templates—the search for criteria comprising properties of elements (the examinational recognition accuracy being 98.7%)—and the Bayesian method—the search for criteria comprising properties of simple chalcogenides (the examinational recognition accuracy being 98.3%). They were used for the prediction of compounds of the AB_3X_3 composition yet to be obtained. In addition, the prediction requires information only on properties of the chemical elements constituting the system (Table 1) or simple chalcogenides.

Final predictions were obtained on the basis of comparing the results of calculations with the use of formulated criteria comprising properties of elements and simple chalcogenides. If the results were conflicting, the final prediction was considered uncertain.

Table 2 gives examples of predictions of compounds of the AB_3X_3 composition. The following designations are used: *1* denotes the prediction of formation of an AB_3X_3 compound under standard conditions; *2* stands for the prediction of the absence of an AB_3X_3 compound under standard conditions; the # symbol marks the examples whose information was used for computer training; empty cells mean uncertain prediction.

RESULTS AND DISCUSSION

Owing to the inconsistency of information on the occurrence of compounds of the AB_3X_3 composition, the selection of examples for computer training was the most difficult task. For example, the data on the occurrence of compounds of the $LnCu_3X_3$ (Ln = Eu–Lu, X = S, Se) composition are extremely conflicting [2, 13,

14]. Chalcogenide systems of alkali metals with lanthanides are hardly studied at all. It is only natural that the absence of reliable data considerably reduces the prediction accuracy. However, when new data appear, the IAS developed by us [9] will make it possible to easily retarget classifying criteria and to refine predictions.

The next stage in the computer-aided design of compounds of the AB_3X_3 (X = S, Se, Te) composition is concerned with the prediction of their type of crystalline structure under standard conditions and with the estimation of properties (melting temperature, band gap, etc.) with the use of our approach [8, 15].

CONCLUSIONS

Predictions of new compounds of the AB_3X_3 (A and B are different elements; X = S, Se, Te) composition are obtained for the first time with the use of computer training methods. The analysis of the predictions obtained show that there is a large reserve of chalcogenide compounds of the AB_3X_3 composition that have yet to be synthesized. The verification of obtained predictions for compounds with copper, silver, and thallium is of greatest interest for applications in electronics.

ACKNOWLEDGMENTS

1 We are grateful to A.V. Stolyarenko, V.A. Dudarev, V.V. Podbel'skii, V.V. Ryazanov, and O.V. Sen'ko for help in the construction of the IAS.

This work was supported in part by the Russian Foundation for Basic Research, project nos. 06-07-89120, 08-01-90427-Ukr_a, 08-07-00437-a, and 05-03-39009.

REFERENCES

1. Lazarev, V.B., Kish, Z.Z., Peresh, E.Yu., and Semrad, E.E., *Slozhnye khal'kogenidy v sistemakh A^I–B^{II}–C^{VI}* (Complex Chalcogenides in A^I–B^{II}–C^{VI} Systems), Moscow: Metallurgiya, 1993.
2. *Fiziko-khimicheskie svoistva poluprovodnikovykh veshchestv: Spravochnik* (Physicochemical Properties of Semiconductor Substances: Handbook), Moscow: Nauka, 1979.
3. Rustamov, P.G., Aliev, O.M., Einullaev, A.V., and Aliev, I.P., *Khal'kolantanaty redkikh elementov* (Chalcolanthanates of Less-Common Elements), Moscow: Nauka, 1989.
4. Gavrilenko, V.I., Grekhov, A.M., Korbutyak, D.V., and Litovchenko, V.G., *Opticheskie svoistva poluprovodnikov: Spravochnik* (Optical Properties of Semiconductors: Handbook), Kiev: Naukova Dumka, 1987.
5. Bercha, D.M., Voroshilov, Yu.V., and Slivka, V.Yu., *Slozhnye khal'kogenidy i khal'kogengalogenidy (poluchenie i svoistva)* (Complex Chalcogenides and Chalcogenhalides (Obtaining and Properties), Lvov: Vishcha shkola, 1983.
6. Distanov, V.E., Nenashev, B.G., Kirdyashkin, A.G., and Serboulenko, M.G., Proustite Single-Crystal Growth by the Bridgman–Stockbarger Method Using ACRT, *J. Cryst. Growth*, 2002, vol. 235, nos. 1–4, pp. 457–464.
7. Singh, N.B., Suhre, D., Gupta, N., et al., Performance of TAS Crystal for AOTF Imaging, *J. Cryst. Growth*, 2001, vol. 225, nos. 2–4, pp. 124–128.
8. Kiseleva, N.N., *Komp'yuternoe konstruirovanie neorganicheskikh soedinenii. Ispol'zovanie baz dannykh i metodov iskusstvennogo intellekta* (Computer-Assisted Design of Inorganic Compounds. The Use of Databases and Artificial Intelligence Methods), Moscow: Nauka, 2005.
9. Kiselyova, N., Stolyarenko, A., and Ryazanov, V., Information-Analytical System for Design of New Inorganic Compounds, *Int. J. Inf. Theories Appl.,* 2008, vol. 2, no. 4, pp. 345–350.
10. Kiseleva, N., Murat, D., Stolyarenko, A., et al., Databases on Properties of Ternary Inorganic Compounds "Phases" in Internet, *Inf. Resursy Ross.*, 2006, no. 4, pp. 21–23.
11. Zhuravlev, Yu.I., Ryazanov, V.V., and Sen'ko, O.V., *RASPOZNAVANIE. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya* (RECOGNITION. Mathematical Methods. Program System. Practical Application), Moscow: FAZIS, 2006.
12. Gladun, V.P., *Protsessy formirovaniya novykh znanii* (Processes of Formation of New Knowledge), Sofia: SD Pedagog 6, 1995.
13. Gulay, L.D., Shemet, V.Y., Olekseyuk, I.D., et al., Investigation of the R_2S_3–Cu_2S–PbS (R = Y, Dy, Ho, and Er) Systems, *J. Alloys Compd.,* 2007, vol. 431, nos. 1–2, pp. 77–84.
14. Gulay, L.D., Wolcyrz, M., Pitraszko, A., and Olekseyuk, I.D., Investigation of theTm_2Se_3–Cu_2Se–PbSe and Lu_2Se_3–Cu_2Se–PbSe Systems at 870 K, *Pol. J. Chem.*, 2006, vol. 80, no. 10, pp. 1703–1714.
15. Kiselyova, N.N., Stolyarenko, A.V., Gu, T., and Lu, W., Computer-Aided Design of New Wide Bandgap Semiconductors with Chalcopyrite Structure, *Perspekt. Mater.*, 2007, Spec. Issue: Proc. IX Russian–Chinese Symp. New Materials and Technologies, pp. 351–355.

SPELL: 1. Dudarev