Studies of Patterns and Statistical Trends in a Database of Inorganic Chemical Reactions—Some Interesting Observations on Stoichiometry

R. Subramaniam,*,† N. K. Goh,‡ and L. S. Chia‡

Singapore Science Centre, Science Centre Road, Singapore 2260, Republic of Singapore, and National Institute of Education, Nanyang Technological University, 469 Bukit Timah Road, Singapore 1025, Republic of Singapore

Received November 1, 1994⊗

Studies of patterns and statistical trends in a database of inorganic chemical reactions have yielded some interesting observations on stoichiometry. An empirical relationship linking the number of reactants, products, and elements constituting these substances in a reaction is advanced. The suitability and reliability of an approach based on this relationship for predicting the number of products formed in a chemical reaction is assessed. Also, the relationship has made possible the definition of numerical bounds for the number of products which can be formed in general in chemical reactions and the classification of reactions in a hierarchy of at least five classes. Other implications of the relationship are also discussed.

INTRODUCTION

A vast amount of information on chemical reactions is now available in computer-readable reaction databases,[1] including CD-ROM databases.[2] The availability of such databases has greatly aided chemists in the retrieval of useful information for their research.

The cost of gaining access to any of the comprehensive computer-readable reaction databases is not inexpensive, especially if the search routine is rather protracted and has to be balanced against competing priorities in the disbursement of research funds allocated for a project. If the scope of the work is well-defined and exploratory in nature at the beginning and entails the need for gaining access to a comprehensive database, it is possible to delay entry by working on a limited database compiled manually on paper for that purpose. Such databases can provide a useful starting point for further investigations. The present work is an example in this regard and demonstrates the utility of assembling a limited database for obtaining useful information about certain aspects of chemical reactions through the observation of simple patterns and statistical trends.

The study was initiated as an attempt to explore certain aspects of the stoichiometry of chemical reactions. For example, is it possible to predict the number of products which will be formed in a chemical reaction given an ensemble of reactants? Our literature search brought to light no references in the primary chemical literature on approaches for predicting the number of products which would be formed in a chemical reaction, though references exist for predicting the types of products which could be formed.[3] It was in the process of examining Porges' article[4] on the algebraic method of balancing chemical equations and the stipulation that solvability is assured as long as the number of substances in a reaction do not exceed the number of independent equations of balance by 2 or more, that a lead in this direction became apparent to us. Further work led to the compilation of a database of chemical reactions, and the examination of patterns and statistical trends in this database which could aid the prediction process. It soon became apparent that in the process of addressing the original brief, other interesting aspects of stoichiometry began to emerge—for example, the existence of numerical bounds for the number of products which could be formed in chemical reactions and the existence of a hierarchy of reaction classes based on the difference between the number of substances (reactants and products) and the number of elements constituting these substances in a chemical reaction. Other related implications also surfaced.

METHODS

It is a formidable task indeed to compile a database of chemical reactions, given the complexity and diversity in the nature of inorganic and organic reactions. As the study was exploratory in nature, we restricted ourselves to inorganic reactions and assembled a database of 140 inorganic chemical equations, culled from standard textbooks and journals. It was ensured that there is a good mix of simple and complicated equations, including the various reaction types such as acid–base, redox, decomposition, metathesis, etc. For each equation, three parameters were extracted—number of reactants, number of products, and number of elements constituting these substances. From these data, patterns were sought which could provide order amidst the prevailing diversity and complexity as well as statistical trends demarcating these patterns. In particular, an empirical relationship involving the aforementioned three parameters was sought.

RESULTS AND DISCUSSION

(1) Formulation and Test of Empirical Relationship. It was found that the difference between the number of substances and the number of elements in the equations surveyed is within ±2, that is, it is one of the following values: −2, −1, 0, 1, 2. Table 1 presents examples of equations satisfying the various numerical relationships.

† Singapore Science Center.
‡ Nanyang Technological University.
⊗ Abstract published in *Advance ACS Abstracts*, March 15, 1995.

0095-2338/95/1635-0428$09.00/0

Table 1. Reactions Typifying the Relationship between the Number of Substances and the Number of Elements in a Chemical Equation

	equation	no. of elements	no. of reactants	no. of products	no. of substances	no. of substances − no. of elements
1.	$2KMnO_4 + 5H_2O_2 + 3H_2SO_4 \rightarrow K_2SO_4 + 2MnSO_4 + 8H_2O + 5O_2$	5	3	4	7	+2
2.	$2KMnO_4 + 10FeSO_4 + 8H_2SO_4 \rightarrow K_2SO_4 + 2MnSO_4 + 5Fe_2(SO_4)_3 + 8H_2O$	6	3	4	7	+1
3.	$2KClO_3 \rightarrow 2KCl + 3O_2$	3	1	2	3	0
4.	$Fe_2(SO_4)_3 + 3Ba(OH)_2 \rightarrow 2Fe(OH)_3 + 3BaSO_4$	5	2	2	4	−1
5.	$3NH_4Cl + Na_3[Co(NO_2)]_6 \rightarrow (NH_4)_3[Co(NO_2)_6] + 3NaCl$	6	2	2	4	−2

Some of the interesting statistical trends observed in this limited survey were as follows:

- For 62.7% of the equations, the difference is +1.
- For 21.6% of the equations, the difference is 0.
- For 9.7% of the equations, the difference is −1.
- For 4.5% of the equations, the difference is +2.
- For 1.5% of the equations, the difference is −2.

The following general expression for a chemical equation can now be formulated:

$$\text{no. of substances} - \text{no. of elements} = \pm 2$$

Denoting e as number of elements, r as number of reactants, and p as number of products, we have

$$r + p - e = \pm 2$$

or

$$p = e - r \pm 2$$

The empirical correlation advanced is a general expression to show that the number of products formed in an inorganic chemical reaction satisfies the stipulated relationship. Test of this relationship on a number of inorganic chemical equations randomly culled from the literature showed that it is remarkably obeyed. One of the reviewers of this article, Professor A. T. Balaban, has brought to our attention that exceptions to the above relationship are some disproportionation reactions as well as all isomerizations (e.g., cis–trans) of inorganic complex compounds formed from more than four elements.

(2) Predicting the Number of Products Formed in an Inorganic Chemical Reaction. We now proceed to explore the reliability and accuracy of the empirical correlation for predicting the number of products formed in a reaction, given an ensemble of reactants. The use of the correlation is, however, complicated by the fact that the degree of discrimination afforded is not good since the number of products can be any one of five possibilities. The statistical survey also points out that the variation is not evenly weighted in the positive and negative directions. However, recognizing the fact that for the majority of the chemical equations surveyed (94%), the variation is within ±1, the prediction range can therefore be significantly narrowed down to one of three possibilities for most reactions if ±1 is used in place of ±2 in the empirical correlation. A prediction to an accuracy of ±1 is considered significant, given the complexity and diversity in the stoichiometry of many chemical reactions. The statistically most probable variation is +1, and this could, of course, be used in place of ±1 or ±2. However, the use of +1 as the variation restricts its applicability to a significant percentage of the reactions. A distinction therefore ought to be made between the statistically most probable variation (+1) to be used in the correlation and the variation (±1) that encompasses a majority of reactions. In other words, the expression $p = e - r \pm 1$ is valid for a majority of the reactions surveyed.

Also, variations of +2 and −2 are rather rare. In the case where the number of substances exceed the number of elements by 2, it is usually for redox reactions involving hydrogen peroxide, chloric acid, and potassium chlorate. These reagents are not particularly stable and may give rise to more than one reduction product.[5] Cognizance of the latter observation can therefore enhance the reliability of the technique for predicting the number of products.

In order to test the reliability of the technique for predicting the number of products formed in chemical reactions, we chose three test equations. The test equations involve either several reactants or several elements within the reactants, and two of these equations are among the most complicated equations which we have encountered in the literature. For each ensemble of reactants in the test equations, the empirical correlation is first used to examine whether the number of products falls within the range of ±2. Having established this fact, we sought to further explore whether the prediction range can be narrowed by the use of ±1 in place of ±2.

Table 2 presents the results of this exercise. Even for seemingly complicated reactions, the relationship is obeyed. In all three test equations, the number of products formed is within ±1. Indeed, if the statistically most probable variation of +1 is used in the correlation, the prediction is on target!

Though the approach to prediction is based on an empirical correlation derived from the observation of patterns and statistical trends at work on a limited database of inorganic chemical reactions, that does not by itself detract from the significance of the results as chemists have traditionally learned from observations, and empirical considerations are commonly factored in a number of theoretical treatments in chemistry. In particular, the simplicity of the approach and the significance of the results have to be seen in the light of the complexity and diversity in the nature of chemical reactions. Thus, even though mechanistic and post reaction analysis techniques are commonly available for product elucidation, the empirical correlation advanced can be a useful adjunct in the process.

It is worth reiterating here that the approach for prediction uses only two readily extracted parameters from a reactant ensemble—number of reactants and the number of elements contained therein. More significantly, no thermodynamic, kinetic, or other chemical considerations are used here!

(3) Numerical Bounds for the Formation of Products in Inorganic Chemical Reactions. The correlation expression $p = e - r \pm 2$ circumscribes the domain for the number of products which could be formed in general in inorganic chemical reactions—that is, $e - r - 2$ to $e - r + 2$. Though the domain could be narrowed to some extent according to the arguments advanced in the previous section, it is interesting to note the existence of a limit on the number of products which could be formed in chemical reactions. The

Table 2. Test Equations for Evaluating Validity of Empirical Correlation in Predicting the Number of Products Formed in Chemical Reactions

	equation	no. of elements	no. of reactants	no. of products
1.	$6(NH_4)_4UO_2(CO_3)_3 \rightarrow 24NH_3 + 2U_3O_8 + 18CO_2 + 12H_2O + O_2$	5	1	5
2.	$10HAuCl_3 + 7K_4Fe(CN)_6 \rightarrow 2KAu(CN)_4 + 7KAu(CN)_2 + KAu(CN)_2Cl_2 + 18KCl + 10HCl + 4Fe(CN)_3{\cdot}3Fe(CN)_2$	7	2	6
3.	$88H_2 + 15Ca(CN)_2 + 6NaAlF_4 + 10FeSO_4 + 3MgSiO_3 + 6KI + 2H_3PO_4 + 6PbCrO_4 + 12BrCl + 3CF_2Cl_2 + 20SO_2 \rightarrow 6PbBr_2 + 6CrCl_3 + 3MgCO_3 + 6KAl(OH)_4 + 10Fe(SCN)_3 + 2PI_3 + 3Na_2SiO_3 + 15CaF_2 + 79H_2O$	19	11	9

dependence of the limit on the number of reactants and elements contained therein is something which has not been given proper consideration in the general scheme of things.

Though the combinatorial dynamics of the elements interspersed among the reactants may be thought to give rise to a number of products during reaction, the parallel operation of thermodynamic considerations, kinetic factors, valence fulfillment requirements when species couple together to form products, preferential bond cleavage, etc. ensure that only a lesser number of products can form.

(4) Laboratory Implications. In laboratory studies of new chemical reactions, knowledge of this interesting relationship proposed can be useful in the course of the investigations. For example, if the number of products formed in a reaction does not satisfy the stipulated relationship, further investigations may be in order. Also, for a given ensemble of reactants for which reaction is known to occur, the relationship offers an intuitive grasp of the number of products that are likely to be formed. This is especially useful for multireactant systems or multielement based reactant ensembles (see Table 2 for second and third equations), where it is clear that a formidable task awaits the chemist in the laboratory!

(5) Hierarchy of Five Reaction Classes. The database study points to a hierarchy of at least five reaction classes, based on the difference between the number of substances (reactants and products) and constituting elements in a reaction. The enumeration of reaction types based on this approach confers another useful tool for classifying the numerous inorganic chemical reactions in the literature as well as providing a logical format to observe statistical trends. Compared to conventional classifications such as acid–base reactions, redox reactions, metathesis, etc., a computer-based classification of reactions contained in comprehensive databases based on this approach would be far more easier because of the numerical basis involved.

The database study suggests that the statistically most prevalent reaction type is where the number of substances exceeds the number of constituting elements by 1. At the other end of the spectrum is the reaction type where the number of elements exceeds the number of substances by 2.

The proposed classification can thus provide an alternative framework for looking at the general scheme of chemical reactions. Indeed, for chemists, the problem of classifying chemical reactions is intrinsically complex. As noted by Satchell,[6] "Many proposals exist for the classification of chemical reactions. Some classifications have been devised simply to provide an author with a convenient framework for some limited purpose; others have been intended to be more rigorous and inclusive. Most have features in common, but the different schemes almost never agree when examined in detail: in short, there exists no consensus about the proper classification of reactions."

(6) Other Aspects. The simplicity of the empirical correlation suggests possible pedagogical implications. When chemical equations are encountered by students (and even chemists), there is a tendency to focus on the chemistry involved. It can be instructive for them to observe the interesting relationship existing between the number of substances and the number of elements contained therein in the equation as well as to determine whether the number of products formed matches that predicted by the correlation, a procedure which hardly takes a minute!

FUTURE WORK

The results of the present work on a limited database of inorganic reactions are promising and offer further scope for refinement and better understanding. Needless to say, further research is needed to fine-tune the procedure for predicting the number of products which would be formed in a chemical reaction.

We are currently exploring the feasibility of gaining access to one of the comprehensive computer-readable reaction databases in order to obtain further indications of the patterns at work and their statistical trends. Such access will offer for study a broad spectrum of chemical reactions and also reveal any exceptions to the general correlation advanced. Incorporation of mathematical ideas based on the works of Sellers[7] and Bonchev[8] is also being considered. The validity of the empirical correlation is also being tested on organic and nuclear chemical reactions.

CONCLUSION

Analysis of a limited database of inorganic chemical reactions through the observation of simple patterns and statistical trends has shed useful light on some aspects of the stoichiometry of reactions.

ACKNOWLEDGMENT

We thank the reviewers for their helpful comments on our article.

REFERENCES AND NOTES

(1) Blakes, J. E.; Dana, R. C. CASREACT: More than a Million Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 394–399.
(2) Hicks, M. G. Beilstein Current Facts in Chemistry: A Large Chemical Database on CD-ROM. *Anal. Chim. Acta* **1992**, *265*, 291–300.
(3) Sello, G. Reaction Prediction: The Suggestions of the Beppe Program. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 713–717.
(4) Porges, J. A Question of Balancing. *J. Chem. Educ.* **1945**, *22*, 266–267.
(5) Hall, W. T. Non-Stoichiometric Equations. *J. Chem. Educ.* **1944**, *21*, 201–202.
(6) Satchell, D. P. N. The Classification of Chemical Reactions. *Naturwiissenschaften* **1977**, *64*, 113–121.
(7) Happel J.; Sellers, P. H. The Classification of Chemical Mechanisms from a Geometric Viewpoint. *Stud. Phys. Theor. Chem.* **1983**, *28*, 420–429.
(8) Bonchev, D.; Mekenyan, O. New Results in the field of Mathematical Chemistry and Chemical Information Sciences. *Khim Ind. (Sofia)* **1988**, *60*, 414–416.

CI940116C