

[F0004]

Molecular Diversity Assessment: Calculation

Shu-Kun Lin

Molecular Diversity Preservation International (MDPI), Saengergasse 25, CH-4054 Basel, Switzerland. Tel: +41 79 3223379, Fax +41 61 3028918, E-mail: Lin@mdpi.org

With [biographical summary](#)

Received: 7 July 1997 / revised: 7 August 1997

- [1. Introduction](#)
- [2. The formula](#)
- [3. Similarity Definition](#)
- [5. Conclusion](#)
- [References and Notes](#)

Abstract: A new pairwise similarity formula has been defined and used for molecular diversity assessment based on a novel method. It is demonstrated here that the logarithmic relations of entropy and indistinguishability give the expected diversity values which decrease with the increase in species similarities. The calculation by *Agrafiotis* would be successful if the present pairwise similarity is used.

Keywords: Species diversity, entropy, similarity, diversity calculation.

1. Introduction

With the development of high throughput screening technology in recent years, the acquisition of molecular and biomolecular samples by collection and combinatorial synthesis has now become the bottleneck in the process of new drug discovery [1]. It suddenly occurred to us that a collection (or a library) of chemical substances (compound samples) was no less precious than a collection (or a library) of chemical information (chemistry books) [2] and the molecular diversity assessment became an urgent topic [1].

A recent article by *Agrafiotis* [3] on calculations using our method [4] demonstrated that a correct definition of the pairwise similarity is the necessary condition to ensure expected results of the diversity assessment. This calculation [3] did not give expected results, which is attributable to a improper definition of pairwise similarity. This similarity scale [3] implies that the maximum distinguishable species is always only 2, instead of N which is required in the original method [4], for assessing diversity of a collection of N chemical samples.

Our method [4] was based on a new theory (Figure 1) [5-7] constructed after rejection of Gibbs paradox. In this paper, simple examples and a newly defined pairwise similarity formula suitable for assessing molecular diversity has been given and used for species diversity calculation.

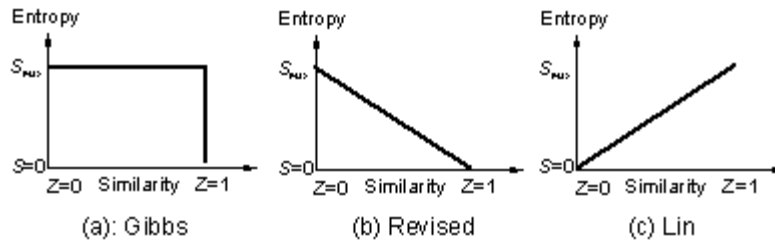


Figure 1. Correlation of entropy of mixing with similarity.

2. The formula

For convenience, our method [4] is briefly summarized here. The diversity index (D) is defined as the ratio of the information (I) and the maximum information (I_{\max}), as given by eq 1.

$$D = \frac{I}{I_{\max}} \quad (1)$$

This method is based on a new theory where entropy is clearly defined as information loss by the following relation:

$$I = S_{\max} - S \quad (2)$$

In this equation, entropy is given by the familiar expression

$$S = - \sum_{i=1}^* p_i \ln p_i \quad (3)$$

where p_i is the probability of the i th microstate with the property that

$$\sum_{i=1}^* p_i = 1 \quad (4)$$

while the maximum entropy is

$$S_{\max} = \ln w \quad (5)$$

where w is the indistinguishability number which is the number of microstates of indistinguishable property.

The apparent indistinguishability number of microstates is defined as

$$w_a = \exp\left(- \sum_{i=1}^* p_i \ln p_i\right) \quad (6)$$

and eq 3 becomes

$$S = \ln w_a \quad (7)$$

which is the logarithmic relation of entropy and indistinguishability.

Practically, in order to record information, a system composed of N "unit devices" is used. In computer science or in our daily information recording as well, these "unit devices" are N individuals (such as symbols) assembled on a media such as a piece of paper. These individuals appear as M attributes, based upon which it is said that the system has M species, such as the two species 0 and 1 in the binary system [4].

Because the assessment of diversity of N chemical samples is our only concern here, the individual number

(N) and the maximum species number (M) are designated as the same: $N = M$. This can be envisaged as N holes in microplates used for high throughput screening containing N compound samples [9].

If these N compounds are all distinguishable, they can be used to record the maximum information as given by eq 8. If red ink is used to represent 0 and blue ink 1, and two bottles of these different inks are used, 2 bits of information can be recorded if the number of individual (N) is 2. There will be 4 ($w = 2^2 = 4$) distinguishable microstates, see Figure 2). The maximum information is

$$I_{\max}(N, N) = \ln w = \ln N^N = N \ln N \quad (8)$$

It is said that this is the maximum information because one can still intentionally use only a small part of the available species to record only smaller amount of information. In eq 8, w is the number of *distinguishable* microstates [8]:

$$w = N^N \quad (9)$$

The corresponding entropy has the minimum value which is zero:

$$S_{\min}(N, N) = \ln 1 = 0 \quad (10)$$

This extreme case is illustrated in Figure 2 ($N=2, w=4$) and Figure 3 ($N=3, w=27$) [8].

```

00 11
01 10
    
```

Figure 2. A binary system of distinguishable species ($M_s = 2$) with $N=2$ which gives 4 microstates.

```

AAA BBB CCC
AAB BAB CAA
AAC BAA CAC
ABA BBA CBB
ABB BBC CBC
ACA BCB CCA
ACC BCC CCB
ABC BAC CBA
ACB BCA CAB
    
```

Figure 3. A trinary system of distinguishable species ($M=3$) with ($N=3$ ($M=3$)) which gives 27 microstates.

Let all the N samples in the N bottles be the samples of extremely similar (or the same) property. Then there will be still $w = N^N$ microstates (or N^N assemblages) constructed by the N^N times of different combinatorial sequences of assembling to form solid structures. However, because they are all virtually *indistinguishable* microstates, there is always the minimum information and the maximum entropy (eqs 11 and 12):

$$I_{\min}(N, N) = \ln 1 = 0 \quad (11)$$

$$S_{\max}(N, N) = \ln w = \ln N^N = N \ln N \quad (12)$$

Suppose you have accidentally installed two bottles of red ink for a printer. Even though exactly the same amount of effort is taken to prepare the 4 microstates, i.e., the four microstates are prepared in a same way as that of Figure 2 by using inks from two individual bottles, there will be 4 ($w = 2^2 = 4$) indistinguishable microstates (see Figure 4). Similarly, whether we factually take the same sample from one sample bottle or different sample bottles, we always have w *indistinguishable* microstates, if they are virtually the same compound in all the N bottles; see Figure 4 ($N=2, w=4$, if all species are factually 0) and Figure 5 ($N=3, w=27$, if all species are B). The maximum microstate indistinguishability number is therefore

$$w = N^N \quad (13)$$

These two extreme sets of distinguishable and indistinguishable samples which give minimum (zero) and maximum entropy values respectively (eqs 10 and 12), already illustrated that our method of entropy calculation is different from the classical statistical mechanics and the classical information theory (see also Figures 1).

00 00
00 00

Figure 4. A binary system with $N=2$ which gives 4 indistinguishable microstates with the highest species indistinguishability ($\alpha_s=2$).

BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB
BBB BBB BBB

Figure 5. A trinary system with ($N=3$) which gives 27 indistinguishable microstates because of the highest species indistinguishability ($\alpha_s=3$). An example is shown in the last row of Table 1, where the property is represented by the symbol "B".

Generally, suppose the N individuals used to construct N^w microstates are only mutually similar to a certain extent and they are neither distinguishable (eqs 8 and 10) nor indistinguishable (eqs 11 and 12). Instead of using eq 1 directly, eq 14 is used to calculate entropy.

$$S(N, N) = - \sum_{j=1}^N \sum_{i=1}^N (p_{ij} \ln p_{ij}) \quad (14)$$

The N^2 pairwise similarities A_{ij} in the table

$$\begin{matrix} A_{11} & A_{12} & \dots & \dots & A_{1N} \\ A_{21} & A_{22} & \dots & \dots & A_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ A_{N1} & A_{N2} & \dots & \dots & A_{NN} \end{matrix} \quad (15)$$

have values limited between 0 and 1 and are given by pairwise comparison among the N individuals according to one and only one systematically followed standard of comparison for all the values A_{ij} ($i=1, \dots, N, j=1, \dots, N$). Then a normalization factor c is required.

$$c = \frac{1}{\sum_{i=1}^N A_{ii}} \quad (16)$$

It follows that

$$A_{ij} = c A_{ij} \quad (17)$$

Then eq 14 will give the same results as given by eqs 10 and 12 respectively under the two extreme conditions.

We agree with *Agrafiotis* [3,10] that, in principle, the general equation (eq 3) should be directly used, where w is simply replaced by N^w . The obvious disadvantage of using eq 1 directly is that the sum runs over all the N^w microstates (see Figures 2-5). The calculation of these terms of enormous number N^w , which

can be an astronomical figure, is impractical. Normally N is 100000, the size of a compound sample library or sublibrary. In eq 14, the number is substantially reduced to totally N^2 terms of $p, \ln p$.

Secondly, we are not really interested in using the chemical samples to record information by taking the sample bottles as "unit devices". Therefore, we will not perform experiment or calculation to characterize the chemical structural and other physicochemical properties of all these N^2 *microstates*. Instead, we measure (or calculate from the known structures) the properties of the N *molecules*, based on which the pairwise similarities are to be easily calculated. This means that, instead of considering similarities and probabilities among N^2 *microstates*, only N^2 probabilities $P_{AA}, P_{AB}, P_{AC}, \dots$, calculated from the pairwise similarities among the N *samples* A, B, C, ..., etc., will be considered.

The first column of Table 1 showed several sets of three imaginary compound samples (A, B and C) plotted against a uniform property scale as used by *Agrafiotis* [3]. If the properties of the samples are the same, the points will coincide and these samples will be regarded as the same samples. If their distances are very short and they are very close, they are regarded as very similar.

The probability p , calculated from eq 17 means the probability of finding the j th individual as the i th species. The diversity of these species is unknown and yet to be assessed; they are presumably similar to each other to a certain extent [4]. Therefore, the comparisons are *not* performed between the N samples and a set of *a priori* known set of distinguishable prototypes; the comparisons are performed among the N samples themselves. The normalization factor c is required because these values are subject to the constraint:

$$\sum_{i=1}^N p_i = 1 \quad (18)$$

Using of the logarithmic relations of entropy (eq 3) , here,

$$S = \ln w_{\bullet} = N \ln \sigma_{\bullet} \quad (19)$$

it is easily found that the apparent species indistinguishability number σ_{\bullet} . For the examples shown in Figures 4 and 5, $\sigma_{\bullet} = N$. Generally,

$$\sigma_{\bullet} = \sqrt[N]{e^S} \quad (20)$$

Easily, the apparent species number M_{\bullet} can be calculated.

$$\begin{aligned} M_{\bullet} &= \exp \left(\ln N + \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^N p_{ij} \ln p_{ij} \right) \\ &= \frac{N}{\sigma_{\bullet}} \end{aligned} \quad (21)$$

Using this method properly, the molecular diversity as expressed by the diversity index D and several related parameters can be calculated. To compare the diversity of several selections of a sublibrary of compounds from all available sources, and to acquire the same number of samples of the highest diversity for many different screening purposes, the sublibrary of minimized entropy S is the choice.

3. Similarity Definition

Before calculating the molecular diversity of a library of N compound samples, the similarities for all the mutual pairwise comparisons among all the N individuals should be clearly defined. Whether it is a proper definition of similarities can be quickly checked first by the following criteria of the two extremes and by using eq 14: (a) It should be able to give the definition of the N individuals of maximum apparent distinguishability. The entropy of this system is the minimum which is zero (eq 10). (b) It should be able to give the definition of the N individuals of minimum apparent distinguishability. This means there can be N indistinguishable species. The entropy of this system is the maximum (eq 12).

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|------|------|---|--|---|---|------|------|------|---|------|------|---|---|---|-----|-----|-----|---|--|------|---|---|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| <table border="1"> <tr><td>A</td><td>C</td><td></td></tr> <tr><td></td><td>B</td><td></td></tr> <tr><td>0.00</td><td>0.50</td><td>1.00</td></tr> </table> | A | C | | | B | | 0.00 | 0.50 | 1.00 | <table border="1"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> </table> | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | <table border="1"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0.5</td><td>0.5</td></tr> <tr><td>0</td><td>0.5</td><td>0.5</td></tr> </table> | 1 | 0 | 0 | 0 | 0.5 | 0.5 | 0 | 0.5 | 0.5 | 1.39 | 4.00 | 1.59 | 1.89 | 1.91 | 0.58 | | | |
| A | C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.00 | 0.50 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.5 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.5 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <tr><td>A</td><td></td><td>C</td></tr> <tr><td></td><td></td><td>B</td></tr> <tr><td>0.00</td><td>0.50</td><td>1.00</td></tr> </table> | A | | C | | | B | 0.00 | 0.50 | 1.00 | <table border="1"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> </table> | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | <table border="1"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0.5</td><td>0.5</td></tr> <tr><td>0</td><td>0.5</td><td>0.5</td></tr> </table> | 1 | 0 | 0 | 0 | 0.5 | 0.5 | 0 | 0.5 | 0.5 | 1.39 | 4.00 | 1.59 | 1.89 | 1.91 | 0.58 | | | |
| A | | C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | B | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.00 | 0.50 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.5 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.5 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <tr><td>A</td><td>C</td><td>B</td></tr> <tr><td></td><td></td><td></td></tr> <tr><td>0.00</td><td>0.50</td><td>1.00</td></tr> </table> | A | C | B | | | | 0.00 | 0.50 | 1.00 | <table border="1"> <tr><td>1</td><td>0</td><td>0.5</td></tr> <tr><td>0</td><td>1</td><td>0.5</td></tr> <tr><td>0.5</td><td>0.5</td><td>1</td></tr> </table> | 1 | 0 | 0.5 | 0 | 1 | 0.5 | 0.5 | 0.5 | 1 | <table border="1"> <tr><td>0.67</td><td>0</td><td>0.33</td></tr> <tr><td>0</td><td>0.67</td><td>0.33</td></tr> <tr><td>0.25</td><td>0.25</td><td>0.50</td></tr> </table> | 0.67 | 0 | 0.33 | 0 | 0.67 | 0.33 | 0.25 | 0.25 | 0.50 | 2.31 | 10.1 | 2.16 | 1.39 | 0.99 | 0.30 | | | |
| A | C | B | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.00 | 0.50 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 1 | 0.5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.5 | 0.5 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.67 | 0 | 0.33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.67 | 0.33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.25 | 0.25 | 0.50 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <tr><td></td><td>C</td><td></td></tr> <tr><td></td><td>B</td><td></td></tr> <tr><td></td><td>A</td><td></td></tr> <tr><td>0.00</td><td>0.50</td><td>1.00</td></tr> </table> | | C | | | B | | | A | | 0.00 | 0.50 | 1.00 | <table border="1"> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | <table border="1"> <tr><td>1/3</td><td>1/3</td><td>1/3</td></tr> <tr><td>1/3</td><td>1/3</td><td>1/3</td></tr> <tr><td>1/3</td><td>1/3</td><td>1/3</td></tr> </table> | 1/3 | 1/3 | 1/3 | 1/3 | 1/3 | 1/3 | 1/3 | 1/3 | 1/3 | 3.30 | 27.0 | 3.00 | 1.00 | 0.00 | 0.00 |
| | C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | B | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.00 | 0.50 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1/3 | 1/3 | 1/3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1/3 | 1/3 | 1/3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1/3 | 1/3 | 1/3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

5. Conclusion

This paper has described the application of a new theory based on the rejection of Gibbs paradox of entropy of mixing and assembling for calculating species diversity. A similarity scale used by Agrafiotis is modified here. As we have demonstrated, if the pairwise similarity is properly defined following very simple criteria, calculations will generate satisfactory and expected results.

Finally, the widespread conceptual confusion between information loss of dynamic mixing and information loss in static assembling, even though has been considered by us [5-7, 13-15], will be discussed in more detail elsewhere.

References and Notes

1. Warr, W. A. Combinatorial chemistry and molecular diversity. An overview. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 134-140; and useful internet linkers and leading literature sources cited therein.
2. (a) Lin, S. -K. Chemical information. *Chem. Eng. News* **May 26, 1997**, *4*.
(b) Lin, S. -K. [Preserving and exploiting molecular diversity: deposit and exchange of chemical information and chemical samples](#). *Molecules* **1997**, *2*, 1-2.
(c) Lin, S. -K. [Molecular Diversity Preservation Strategies: The MDPI Project](#), paper presented at the 36th IUPAC Congress, Geneva, Switzerland, August 17 - 22, 1997. (Abstract published in *Chimia*, **1997**, *51*, 544)
3. Agrafiotis, D. K. On the use of information theory for assessing molecular diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 576-580.
4. Lin, S. -K. Molecular diversity assessment: Logarithmic relations of information and species diversity and logarithmic relations of entropy and indistinguishability after rejection of Gibbs paradox of entropy of mixing. *Molecules* **1996**, *1*, 57-67. (The electronic version of reprint of this published paper can be downloaded from <http://www.mdpi.org/lin.htm>).
5. Lin, S. -K. Understanding structural stability and process spontaneity based on the rejection of the Gibbs paradox of entropy of mixing. *J. Mol. Struct. Theochem* **1997**, *398*, 145-153.
6. Lin, S. -K. Correlation of entropy with similarity and symmetry. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 367-376.
7. Lin, S. -K. Gibbs paradox of entropy of mixing: experimental facts, its rejection and the theoretical consequences. *J. Theor. Chem.* **1996**, *1*, 135-150. (The electronic version of reprint of this published paper can be downloaded from <http://www.mdpi.org/lin.htm>).
8. This already explained clearly the relation of "higher diversity higher information". However, Agrafiotis [3] claims that "increase in diversity results in an *increase* in entropy" (italics belong to Agrafiotis [3]).

9. Normally N can be either greater or smaller than M . For example, a harddisk of N bits has N individuals with M equals 2. Normally $N \gg M$. If one puts 200 Chinese characters in a typical letter written in Chinese where N equals 200 and $M > N$ (here the species number M is the total number of different Chinese characters normally used which is 10000).

10. Whether $-\sum_{i=1}^N p_i \ln p_i$ and $-\sum_{j=1}^M \sum_{i=1}^N (p_{ij} \ln p_{ij})$ are compatible or not under the restrictions specified in the context should be mathematically justified. For the two extremes they give the same results as shown in eqs 10 and 12.

11. Johnson, M. A.; Maggiora, G. M. eds. *Concepts and Applications of Molecular Similarity*, New York: Wiley-Interscience Publication, 1990. 12. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*, Weinheim: VCH, 1993; 172. 13. Lin, S. -K. *Informational temperature concept and the nature of self-organization*. Paper presented at the Second International Congress on Theoretical Chemical Physics, New Orleans, Louisiana, April 9-13 1996.

14. (a) Lin, S. -K. *The nature of the chemical process. Part 2. Mixing and separation*. Paper presented at The Fourth World Congress of Theoretically Oriented Chemists, Jerusalem, Israel, July 7-12, 1996. (b) Lin, S. -K. *The nature of the chemical process. Part 3. Self-organization in hierarchical structures*. Paper presented at The Fourth World Congress of Theoretically Oriented Chemists, Jerusalem, Israel, July 7-12, 1996. (c) Lin, S. -K. *The nature of the chemical process. Part 4. Formation of the chemical bond*. Paper presented at The Fourth World Congress of Theoretically Oriented Chemists, Jerusalem, Israel, July 7-12, 1996. (d) Lin, S. -K. *The nature of the chemical process. Part 5. Deformation in energy transduction*. Paper presented at the Fourth World Congress of Theoretically Oriented Chemists, Jerusalem, Israel, July 7-12, 1996.

15. (a) Lin, S. -K. *Symmetry breaking problem resolved*. Paper presented at the American Physical Society 1997 March Meeting, Kansas City, MO, March 17-21, 1997. Abstract downloadable from <http://www.aps.org/BAPSMAR97/abs/S4580004.html>. (b) Lin, S. -K. *Symmetry breaking problem resolved*. *Bull. Am. Phys. Soc.* **1997**, 42, 679.

Shu-Kun Lin

Shu-Kun born March 24, 1957 in Hanchuan, Hubei Province, China. He graduated from Wuhan University in 1982, majoring inorganic chemistry. He studied physical chemistry in the Chinese Academy of Sciences (1982-1986, MSc in 1985) and in the USA (1.1987-7.1989). He obtained his doctorate in organic chemistry at the Swiss Federal Institute of Technology (ETH-Zurich) in 1992 after three years studies at the group of Prof. Dr. Bernhard Jaun, Lab. f. org. Chemie-ETH-Zurich. He then worked at Ciba-Geigy Ltd. for ca. three years, first as a postdoctoral research associate in organic synthesis and worked on two drug discovery programmes in the CNS area for 1.5 years, secondly for the Dyestuffs Division of Ciba-Geigy Ltd., receiving training in dyestuff R&D and Production. He initiated a molecular diversity preservation project and founded [MDPI \(Molecular Diversity Preservation International\)](#), a nonprofit and independent international organization in Switzerland. In 1995 he launched the first journal [Molecules](#) that encourages authors to deposit their compound samples. He is the main author of 40 publications. He has two children (a daughter Qian-Qian, born July 27, 1986, and a son, Di-Fan, born June 1, 1996). His personal website is <http://www.mdpi.org/lin.htm>. Present Address: Dr. Shu-Kun Lin, Molecular Diversity Preservation International (MDPI), Saengergasse 25, CH-4054 Basel, Switzerland. Tel. +41 79 322 3379, Fax +41 61 302 8918, E-mail: Lin@mdpi.org.

Comments

During 1-30 September 1997, all comments on this poster should be sent by e-mail to ecsoc@listserv.arizona.edu with **F0004** as the message subject of your e-mail. After the conference, please send all the comments and reprints requests to the author(s).

