



# Information entropy changes in chemical reactions

Denis Sh. Sabirov

Institute of Petrochemistry and Catalysis, Russian Academy of Sciences, 450075 Ufa, Russia

## ARTICLE INFO

### Article history:

Received 17 October 2017

Received in revised form 28 November 2017

Accepted 29 November 2017

Available online 2 December 2017

### Keywords:

Information entropy  
Molecular complexity  
Molecular ensemble  
Chemical reaction

## ABSTRACT

Currently, information entropy is used mainly for the description of the complexity of chemical structures. Recent advances in its application to fullerenes and interstellar compounds revealed that it could provide valuable information about chemical processes taking place under extreme conditions. We have applied the known Karreman's algorithm to calculating information entropies of chemical reactions. This algorithm implies that information entropy of molecular ensemble equals the sum of the information entropies of its members and, consequently, information entropy change in a reaction is defined similar to common thermodynamic functions. Unfortunately, this approach sometimes leads to counterintuitive results. Therefore, in the present work, we introduce and justify a novel methodology for assessing information entropy changes in chemical reactions. It treats the molecular ensemble (which may correspond to reactants or products) as a whole and its information entropy generally does not equal the sum of the information entropies of its members. A comparison of the novel algorithm with the previous one has been performed on the examples of atomization, cycloaddition, isomerization, and exchange reactions. As shown, our approach allows avoiding ambiguous interpretations and provides accurate estimates of the changes in complexity under chemical reactions.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Application of the information approach (in its topological form) to the description of chemical structures has a long story and there are numerous examples. Initially, this approach was developed by Shannon [1] and Hartley [2] for mathematical description of information transmission and then used for numerically assessing complexity of diverse objects including graphs [3]. Since molecules can be approximated as graphs, in which vertices and edges represent atoms and chemical bonds, information entropy was recognized as an efficient structural descriptor [4–8]. Indeed, it is applied (both in original [4,7–11] and modified [6,11,12] forms) to the specific needs of chemometrics and cheminformatics and enters such hot topics as, for example, drug screening [13]. The information-entropy estimates are sensitive to the changes in molecular structure that is used in QSAR–QSRP studies [10,11,14] and rational schemes for classification of natural compounds [9,15,16].

On the other hand, information entropy has been narrowly applied to chemical and physicochemical processes. The list of such works was initiated with Karreman's study [17] on the topological changes upon chemical reactions and continued by Zhdanov, who applied the information approach to the selected

reactions of organic compounds. We should note the disadvantages of these two works. Karreman [17] treated entirely hypothetical cases far from real chemical reactions whereas Zhdanov's work [9] dealt mainly with *a posteriori* analysis of chemical reactions, viz. the empirical yields of the products were analyzed in terms of information entropy. Simultaneously, the Shannon approach was used to investigate physicochemical aspects of thinking [18] and pure physical processes [19,20]. Recently, it has been invoked to estimate the energetic disorder of conformational complexity [21] and its applicability to self-assembly processes under non-equilibrium conditions has been considered [22–24].

Previously, we have demonstrated that the information approach in its discrete form is useful for analysis of fullerenes [25,26], oxygen allotropes [27], and variegated set of interstellar molecules [28]. For example, buckminsterfullerene C<sub>60</sub> dominates over the great amount of other fullerene structures under the arc-discharge synthesis and interstellar conditions and it is the only member of the fullerenes family with zero information entropy [25]. Most of the non-detected (hypothetical) fullerenes have the information entropies larger than corresponding value of C<sub>70</sub>, the second most abundant fullerene. It is assumed that, being a descriptor for molecular complexity, information entropy may be deeper introduced into the theory of chemical reactions. A computation scheme for chemical reactions worked out by Bertz [29] is one of the advanced schemes based on the C index deduced

E-mail address: [diozno@mail.ru](mailto:diozno@mail.ru)

from information entropy. This approach is recommended for assessment of complexity of organic synthetic strategies and supported by typical examples of synthetic reactions.

As for application of the information approach in its original form, Karreman's work [17] seems to be the only one in this regard. It implies treating the information entropy change in a chemical reaction like the heat effect and other thermodynamic reaction parameters. Trying to incorporate this approach into our mechanistic study on fullerene [25] and endofullerene [26] formation, we have previously found that information entropy of a chemical reaction  $\Delta h_R$  can be calculated in an alternative manner. To the best of our knowledge, the modes of calculation of  $\Delta h_R$  have not been previously compared and discussed. However, this issue deserves a particular attention because it deals with the correct use and interpretation of the information approach applied to chemical processes. In the present work, we introduce and theoretically justify our mode of the  $\Delta h_R$  calculation and compare it with the previous computational scheme of Karreman applied to some typical reactions (mainly of organic compounds).

## 2. Theoretical background and justification of novel algorithm

### 2.1. Information entropy of the molecule

According to the original Shannon approach, a chemical structure is considered as a set of  $N_1$  atoms of the 1st type,  $N_2$  atoms of the 2nd type, ...  $N_n$  atoms of the  $n$ -th type, where  $n$  is a number of atom types (or cardinality of the set) and  $N = \sum N_j$  is a total number of atoms in the molecule. The probability ( $p_j$ ) to find the atom of  $j$ -th type is the weight of each subset and equal

$$p_j = \frac{N_j}{\sum_{j=1}^n N_j} \quad (1)$$

Information entropy  $h$  (or Shannon entropy) is introduced as a sum of the logarithms of the probabilities

$$h = -\sum_{j=1}^n p_j \log_2 p_j \quad (2)$$

or

$$h = -\sum_{j=1}^n \frac{N_j}{\sum_{j=1}^n N_j} \log_2 \frac{N_j}{\sum_{j=1}^n N_j} \quad (3)$$

The use of logarithm to the base of two allows expressing information entropy in bits (and these equations work regardless of the type of the base). Discriminating atoms by atom types, we take into account their positions in the molecular graph. Atoms of different elements are attributed to different atom types by default. Here-with, atoms of different elements equally contribute into the information entropy of the molecule. Atoms of the same element belong to different atom types (subsets) if they occupy different positions in the graph. The inequivalence of the vertices depends on the connectivity but we do not explicitly consider the type of the bond (single, double, triple) in terms of our approach.

Conformationally flexible molecules contain the rotating fragments, so several conformations may correspond to one molecular graph, depending on spatial arrangement of the fragments. We follow the common convention neglecting this conformational complexity [29,30]. As a significant (and acceptable) simplification, the peripheral atoms of chemical groups possessing free rotation, e.g.  $CX_3-$ ,  $-CX_2-$ , and  $NX_2-$  ( $X$  = hydrogen or halogen), are treated as equivalent.

Thus, information entropy characterizes structural complexity of the molecule and depends on its symmetry and number of the constituting atoms. The calculated  $h$  values of some typical

molecules are shown in Table 1. High symmetry and low elemental diversity decrease the information entropy of the molecule. These factors cause more uniform distribution of the constituent atoms over the atoms types. Low symmetry, enlarging molecular size, and number of chemical elements making up the molecule increase its information entropy.

A special case of Eq. (3) is presented by low-symmetry molecules with a singular population of all atom types ( $N_j = 1$ ,  $n = N$ ) and includes heteronuclear diatomic molecules AB, linear molecules with  $C_{\infty v}$  symmetry (HCN,  $HC\equiv CCl$ ,  $HC\equiv CCN$ , etc.), and all species having no symmetry ( $C_1$  point symmetry group). Their information entropies are equal to:

$$h_{\max} = -\log_2 \frac{1}{N} \quad (4)$$

The  $h_{\max}$  value is the highest structural complexity that may be reached with a given set of the atoms used to design the molecule [28].

### 2.2. Information entropy of molecular ensemble

To rationalize the application of the information approach to chemical reactions, we consider the reactants and products of chemical reaction as two isomeric molecular ensembles. According to the definition [31–33], isomeric molecular ensembles (MEs) is a set of different combinations of molecules made up with the fixed types of chemical elements and numbers of atoms. In terms of Karreman's approach [17], information entropy of ME is expressed as a sum of the  $h$  values of its isolated components:

$$h_{ME}^K = \sum_{i=1}^m v_i h_i \quad (5)$$

where  $h_i$  and  $v_i$  are the information entropies and number of molecules in ME, respectively ( $m$  is a number of types of molecules; and  $v = \sum v_i$  is a total number of molecules in ME). This mode of calculation may lead to counterintuitive results. Let us consider a molecular ensemble made up with  $v$  identical molecules so that  $m = 1$  and  $h_1 = h$ . According to Eq. (5), the information entropy of such ensemble equals  $vh$  and linearly increases when it includes new molecules of the same type. However, it is not obvious that the structural complexity of such ensemble increases too. To demonstrate the ambiguity of the situation, let us imagine the NMR measurements of these MEs. We choose this experimental technique as it allows efficient sortation of the atoms by atom types. Regardless of the number of the molecules, we should obtain identical sets of NMR shifts for the ensemble described above. Analogous ambiguity arises if the ensemble consists of the compound with  $h_1 \neq 0$  and simple diatomic  $X_2$  with  $h_2 = 0$ , e.g.,  $CH_4 + O_2$ ,  $C_2H_4 + Cl_2$ . According to

**Table 1**  
Information entropies of typical molecules.

Molecule	Partition	$h$
CH <sub>4</sub>	1 × 4 + 1 × 1	0.722
CH <sub>3</sub> Cl	1 × 3 + 2 × 1	1.371
C <sub>2</sub> H <sub>6</sub>	1 × 6 + 1 × 2	0.811
C <sub>2</sub> H <sub>4</sub>	1 × 4 + 1 × 2	0.918
C <sub>2</sub> H <sub>2</sub>	2 × 2	1.000
CH <sub>3</sub> OH	1 × 3 + 3 × 1	1.792
CH <sub>3</sub> CH <sub>2</sub> OH	1 × 3 + 1 × 2 + 4 × 1	2.419
CH <sub>3</sub> OCH <sub>3</sub>	1 × 6 + 1 × 2 + 1 × 1	1.224
CH <sub>3</sub> COOH	1 × 3 + 5 × 1	2.406
C <sub>6</sub> H <sub>6</sub>	2 × 6	1.000
C <sub>60</sub> ( $I_h$ )	1 × 60	0
C <sub>70</sub> ( $D_{5h}$ )	3 × 10 + 2 × 20	2.236
HCl	2 × 1	1.000
O <sub>3</sub>	1 × 2 + 1 × 1	0.918
H <sub>2</sub> O	1 × 2 + 1 × 1	0.918

Download English Version:

<https://daneshyari.com/en/article/7839159>

Download Persian Version:

<https://daneshyari.com/article/7839159>

[Daneshyari.com](https://daneshyari.com)