

УДК 541+544.01

DOI: 10.31040/2222-8349-2021-0-1-111-115

**ИНФОРМАЦИОННАЯ ЭНТРОПИЯ ХИМИЧЕСКИХ РЕАКЦИЙ –  
НОВАЯ МЕТОДОЛОГИЯ ДЛЯ ЦИФРОВОЙ ОРГАНИЧЕСКОЙ ХИМИИ\***

© Д.Ш. Сабиров

Развитие методологии вычисления структурных дескрипторов, являющихся базовыми понятиями математической и цифровой химии, связано с расширением области их применения. В частности, актуальной задачей является разработка методов вычисления дескрипторов, описывающих строение ансамблей частиц. Информационная энтропия широко используется для описания химической структуры молекул, однако ее применение к молекулярным ансамблям и химическим реакциям ограничено. Нами предложен новый подход к вычислению информационной энтропии молекулярного ансамбля, который учитывает не только вклады отдельных молекул, но и корпоративную энтропию, возникающую при объединении молекул в ансамбль. Такой подход позволяет по-новому рассчитывать изменение информационной энтропии в химических реакциях, представляемых как превращение одного ансамбля (исходных веществ) в другой (молекулярный ансамбль продуктов). Новый способ расчета дает возможность оценить по отдельности вклады в химическую реакцию, связанные с изменением строения и размера молекул.

Ключевые слова: информационная энтропия, молекулярный ансамбль, химическая реакция, цифровая химия.

**Введение.** Цифровая химия – формирующееся направление математической химии, связанное с использованием методов компьютерного моделирования для решения таких химических задач, как поиск соединений с прогнозируемыми свойствами, анализ кинетических схем многостадийных химических процессов, оценка сложности строения молекул и молекулярного подобию, исчисление и генерация изомеров и т.п. Одним из ключевых инструментов цифровой химии является использование дескрипторов для описания строения химических соединений, и этот вопрос вполне проработан на концептуальном уровне для изолированных молекул. Вместе с тем использование структурных дескрипторов (например, информационной энтропии) для описания молекулярных ансамблей и химических процессов до сих пор остается открытым.

В лаборатории математической химии ИНК УФИЦ РАН было найдено, что информационная энтропия молекулярного ансамбля складывается не только из значений информационной энтропии молекул – и включает

дополнительное слагаемое, названное корпоративной энтропией  $H_{\Omega}$ , которая возникает при объединении молекул в ансамбль. Корпоративная энтропия не зависит от строения молекул и определяется только их размером. Эта особенность описания строения молекулярных ансамблей открывает новые возможности для цифрового описания химических реакций органических соединений.

**Формулировка проблемы.** Широкое распространение для описания строения молекул получили структурные дескрипторы, являющиеся инвариантами молекулярных графов (дескрипторы теоретико-графового типа) и так называемые логарифмические меры сложности (дескрипторы теоретико-множественного типа) [1]. Последние могут быть применены не только к отдельно взятым молекулам, но и к их ансамблям. Исторически информационная энтропия ( $h$ ) была первым дескриптором этого типа, который получил широкое распространение в математической химии. Этот параметр был введен Клодом Шенноном в теорию

\* По материалам доклада «Цифровая химия – взгляд в будущее» на Объединенном ученом совете УФИЦ РАН от 22.01.2021 г.

коммуникации в 1948 г. [2] и затем нашел широкое применение для оценки сложности молекулярных объектов, которые могут быть представлены в виде множеств элементов – атомов и/или химических связей [3, 4]. Ранее в ИНК УФИЦ РАН «классический» информационно-теоретический подход использовался для изучения строения фуллеренов и их производных [5, 6], аллотропных модификаций кислорода [7], органических соединений межзвездной среды [8].

Был известен только один способ расчета изменения информационной энтропии в химических реакциях (подход Карремана [9]), построенный на аналогии информационной и термодинамической энтропии. Изменение информационной энтропии в химической реакции предлагалось рассчитывать как разность значений, соответствующих продуктам и реагентам [9]:

$$\Delta h_R = \sum h_{(\text{prod})} - \sum h_{(\text{react})}. \quad (1)$$

Мы обнаружили, что такой способ расчета  $\Delta h_R$  приводит к контринтуитивным результатам, что послужило отправной точкой поиска корректного подхода. Было очевидно, что некорректность этого способа расчета связана с неприменимостью простого суммирования значений  $h$ . Классическим примером некорректности таких количественных оценок являются бимолекулярные ансамбли, состоящие из молекул с нулевой ( $h_1 = 0$ ) и ненулевой энтропией ( $h_2 \neq 0$ ). В подходе Карремана информационная энтропия таких ансамблей равна  $h_{ME} = h_2$ , т.е. игнорируется наличие еще одной молекулы (хотя с химической точки зрения очевидно, что ансамбль двух частиц должен быть сложнее, чем любая из этих частиц по отдельности) [10].

**Решение – новый подход.** Информационная энтропия каждой отдельной молекулы ( $h_i$ ) рассчитывается с учетом числа типов неэквивалентных атомов в ее структуре ( $n$ ) и их заселенностей ( $N_j$ ;  $N = \sum N_j$  – общее число атомов в молекуле), то есть

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

Доля атомов, приходящаяся на каждый тип рассматриваемых молекул:

$$\omega_i = \frac{N_i}{\sum_{i=1}^m N_i}. \quad (3)$$

В общем случае расчет информационной энтропии ансамбля  $h_{ME}$  из  $m$  молекул выполняется аналогично формуле (2):

$$h_{ME} = - \sum_{i=1}^m \sum_{j=1}^n \frac{N_{ij}}{N} \log_2 \frac{N_{ij}}{N}. \quad (4)$$

Последнее выражение не устанавливает в явном виде связь между информационной энтропией ансамбля и составляющих его молекул. Поэтому мы преобразовали его для установления искомой зависимости [11]:

$$h_{ME} = H_{\Omega} + \sum_{i=1}^m \omega_i h_i, \quad (5)$$

где первый член называется кооперативной энтропией, возникающей при объединении молекул в ансамбль:

$$H_{\Omega} = - \sum_{i=1}^m \omega_i \log_2 \omega_i. \quad (6)$$

Эта величина не зависит от строения молекул и определяется только их размером (числом атомов в молекуле). Учет  $H_{\Omega}$  позволяет снять противоречия, появляющиеся в подходе Карремана [10, 11].

Поскольку химическую реакцию можно интерпретировать как изомеризацию одного ансамбля в другой [12], изменение информационной энтропии рассчитывается как разность значений, соответствующих молекулярным ансамблям продуктов и реагентов:

$$\Delta h_R = h_{ME(\text{prod})} - h_{ME(\text{react})}. \quad (7)$$

После небольших математических преобразований в величине  $\Delta h_R$  можно выделить отдельные вклады в информационную энтропию химического процесса, связанные с (*a*) изменением строения молекул и/или (*b*) их размеров, обозначаемые как энтропия реорганизации ( $H_{reorg}$ ) и энтропия перераспределения атомов между молекулами ( $H_{redistr}$ ) [11]:

$$\Delta h_R = H_{reorg}^{str} + H_{reorg}^{str+size} + H_{redistr}^{size}, \quad (8)$$

где (верхние индексы *str*, *size* и *str + size* указывают на зависимость слагаемых от строения молекул, их размера либо от обоих факторов одновременно):

$$H_{reorg}^{str} = \sum_i^{prod} h_i - \sum_j^{react} h_j \quad (9)$$

$$H_{reorg}^{str+size} = \sum_j (1 - \omega_j) h_j - \sum_i^{prod} (1 - \omega_i) h_i \quad (10)$$

$$H_{redistr}^{size} = H_{\Omega}^{prod} - H_{\Omega}^{react}. \quad (11)$$

Отметим, что в подходе Карремана [9] вычисляется только энтропия реорганизации (ср.: уравнения (1) и (9)). Концептуальная схема развиваемого подхода представлена на рис. 1.

Новый подход был успешно апробирован для цифрового описания классических реакций органических соединений (реакций атомизации, изомеризации, Дильса–Альдера и др.) [10, 11]. Для однотипных реакций развиваемый подход дает разные оценки информационной энтропии в зависимости от химической природы взаимо-

действующих веществ и образующихся продуктов, т.е. его использование открывает возможности «машинной» классификации химических реакций. Например, информационная энтропия реакций образования простых эфиров зависит от того, одинаковые или разные молекулы спирта в ней участвуют [10]; информационная энтропия олигомеризации фуллерена по-разному «реагирует» на четность числа мономерных звеньев в зависимости от строения образующихся наноструктур [13] (рис. 2).

**Выводы.** Разработан новый подход к вычислению информационной энтропии молекулярных ансамблей и химической реакции, открывающий новые возможности для цифрового описания реакций органических соединений.

Помимо применения этого подхода к изучению конкретных химических процессов, нами исследуются сопутствующие концептуальные вопросы, такие как выполнимость в новом подходе закона Гесса, условия изоэнтропийности молекул и ансамблей и др. Отметим, что в настоящее время информационная энтропия в ряде исследований трактуется шире, чем просто структурный дескриптор. В работах [14, 15] обсуждаются варианты ее физико-химической интерпретации в контексте самоорганизации вещества; возможна связь информационной и конфигурационной энтропии [16]. Кроме того, этот параметр имеет отношение к количественной оценке материальных ресурсов, необходимых для хранения информации, в том числе в задачах цифрового кодирования химических объектов [17].

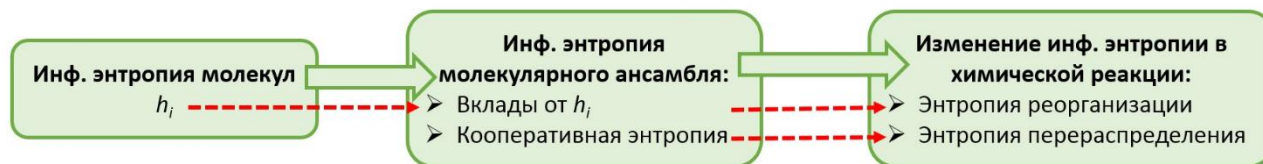


Рис. 1. Концептуальная схема нового подхода к вычислению информационной энтропии химических реакций

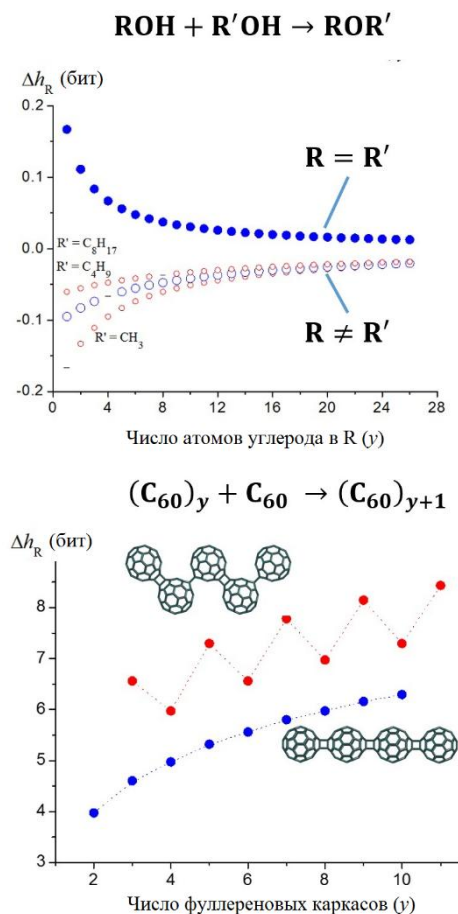


Рис. 2. Примеры применения нового подхода к изучению реакций органических соединений

*Исследования по заявленной тематике выполняются при частичной финансовой поддержке Совета по грантам Президента РФ (проект МД-874.2021.1.3).*

### Литература

1. Станкевич М.И., Станкевич И.В., Зефиоров Н.С. Топологические индексы в органической химии // *Успехи химии*. 1988. Т. 62. С. 337–366.

2. Shannon C.E. A mathematical theory of communication // *The Bell System Technical Journal*. 1948. V. 27. P. 379–423.

3. Barigye S.J., Marrero-Ponce Y., Pérez-Giménez F., Bonchev D. Trends in information theory-based chemical structure codification // *Molecular Diversity*. 2014. V. 18. P. 673–686.

4. Basak S.C., Gute B. D. Mathematical biodescriptors of proteomics maps: Background and applications // *Current Opinion in Drug Discovery and Development*. 2008. V. 11. P. 320–326.

5. Sabirov D.Sh., Ōsawa E. Information entropy of fullerenes // *J. Chem. Inf. Model*. 2015. V. 55. P. 1576–1584.

6. Sabirov D.Sh., Terentyev A.O., Sokolov V.I. Activation energies and information entropies of helium penetration through fullerene walls. Insights into the formation of endofullerenes  $nX@C_{60/70}$  ( $n = 1$  and  $2$ ) from the information entropy approach // *RSC Advances*. 2016. V. 6. P. 72230–72237.

7. Sabirov D.Sh., Shepelevich I.S. Information entropy of oxygen allotropes. A still open discussion about the closed form of ozone // *Comput. Theor. Chem*. 2015. V. 1073. P. 61–66.

8. Sabirov D.Sh. Information entropy of interstellar and circumstellar carbon-containing molecules: Molecular size against structural complexity // *Comput. Theor. Chem*. 2016. V. 1097. P. 83–91.

9. Karreman G. Topological information content and chemical reactions // *Bull. Math. Biophys.* 1955. V. 17. P. 279–285.

10. Sabirov D.Sh. Information entropy changes in chemical reactions // *Comput. Theor. Chem*. 2018. V. 1123. P. 169–179.

11. Sabirov D.Sh. Information entropy of mixing molecules and its application to molecular ensembles and chemical reactions // *Comput. Theor. Chem*. 2020. V. 1187. P. 112933.

12. Ugi I., Gillespie P. Representation of chemical systems and interconversions by be matrices and their transformation properties // *Angew. Chem. Int. Ed. Engl.* 1971. V. 10. P. 914–915.

13. Sabirov D.Sh. Information entropy change in [2+2]-oligomerization of the  $C_{60}$  fullerene // *Int. J. Chem. Model*. 2017. V. 9 (4). P. 203–213.

14. Алесковский В.Б. Информация как фактор самоорганизации и организации вещества // *Журнал общей химии*. 2002. Т. 72, № 4. С. 611–616.

15. Таланов В.М., Иванов В.В. Структура как источник информации о химической организации вещества // *Журнал общей химии*. 2013. Т. 83, № 12. С. 1937–1948.

16. Champion Y., Thurieu N. The sample size effect in metallic glass deformation // *Sci. Rep.* 2020. V. 10. P. 10801.

17. Nielsen M.A., Chuang I.L. Quantum computation and quantum information // Cambridge University Press. 2001. 822 p.

### References

1. Stankevich M.I., Stankevich I.V., Zefirov. Topological indexes in organic chemistry. *Uspekhi khimii*, 1988, vol. 62, pp. 337–366.

2. Shannon C.E. A mathematical theory of communication. *The Bell System Technical Journal*, 1948, vol. 27, pp. 379–423.

3. Barigye S.J., Marrero-Ponce Y., Pérez-Giménez F., Bonchev D. Trends in information theory-based chemical structure codification. *Molecular Diversity*, 2014, vol. 18, pp. 673–686.

4. Basak S.C., Gute B. D. Mathematical biodescriptors of proteomics maps: Background and applications. *Current Opinion in Drug Discovery and Development*, 2008, vol. 11, pp. 320–326.

5. Sabirov D.Sh., Ōsawa E. Information entropy of fullerenes. *J. Chem. Inf. Model*, 2015, vol. 55, pp. 1576–1584.

6. Sabirov D.Sh., Terentyev A.O., Sokolov V.I. Activation energies and information entropies of helium penetration through fullerene walls. Insights into the formation of endofullerenes  $nX@C_{60/70}$  ( $n = 1$  and  $2$ ) from the information entropy approach. *RSC Advances*, 2016, vol. 6, pp. 72230–72237.

7. Sabirov D.Sh., Shepelevich I.S. Information entropy of oxygen allotropes. A still open discussion about the closed form of ozone. *Comput. Theor. Chem.*, 2015, vol. 1073, pp. 61–66.

8. Sabirov D.Sh. Information entropy of interstellar and circumstellar carbon-containing molecules: Molecular size against structural complexity. *Comput. Theor. Chem.*, 2016, vol. 1097, pp. 83–91.

9. Karreman G. Topological information content and chemical reactions. *Bull. Math. Biophys.*, 1955, vol. 17, pp. 279–285.

10. Sabirov D.Sh. Information entropy changes in chemical reactions. *Comput. Theor. Chem.*, 2018, vol. 1123, pp. 169–179.

11. Sabirov D.Sh. Information entropy of mixing molecules and its application to molecular ensembles and chemical reactions. *Comput. Theor. Chem.*, 2020, vol. 1187, p. 112933.

12. Ugi I., Gillespie P. Representation of chemical systems and interconversions by be matrices and their transformation properties. *Angew. Chem. Int. Ed. Engl.*, 1971, vol. 10, pp. 914–915.

13. Sabirov D.Sh. Information entropy change in [2+2]-oligomerization of the C<sub>60</sub> fullerene. *Int. J. Chem. Model.*, 2017, vol. 9 (4), pp. 203–213.

14. Aleskovskii V.B. Information as a factor of self-organization and organization of matter. *Zhurnal obshchey khimii*, 2002, vol. 72, no. 4, pp. 611–616.

15. Talanov V.M., Ivanov V.V. Structure as a source of information about chemical organization of

matter. *Zhurnal obshchey khimii*, 2013, vol. 83, no. 12, pp. 1937–1948.

16. Champion Y., Thureau N. The sample size effect in metallic glass deformation. *Sci. Rep.*, 2020, vol. 10, pp. 10801.

17. Nielsen M.A., Chuang I.L. *Quantum computation and quantum information*. Cambridge University Press, 2001. 822 p.



## **INFORMATION ENTROPY OF CHEMICAL REACTIONS: A NOVEL METHODOLOGY FOR DIGITAL ORGANIC CHEMISTRY**

© **D.Sh. Sabirov**

Institute Petrochemistry and Catalysis – Subdivision of the Ufa Federal Research Centre  
of the Russian Academy of Sciences,  
141, prospect Oktyabrya, 450075, Ufa, Russian Federation

The development of the computational methodology of structural descriptors, which are the basic concept of mathematical and digital chemistry, involves expanding the field of their application. In particular, the development of methods for calculating descriptors relating to the structure of the ensembles of particles is an urgent task. Information entropy is widely used to describe the chemical structures of molecules but its application to molecular assembles and chemical reactions is limited.

We have proposed a new approach to calculating the information entropy of a molecular ensemble: besides the contributions of individual molecules, it takes into account the corporative entropy arising when molecules form an ensemble. This approach provides a fresh view on calculating the change in information entropy in chemical reactions approximated as the transformation of one ensemble (reactants) into another (molecular ensemble of products). The new method of calculation allows separate assessing the contributions to a chemical reaction associated with a change in the molecular structure and molecule size.

**Key words:** information entropy, molecular ensemble, chemical reaction, digital chemistry.