

НАУЧНАЯ СТАТЬЯ

УДК 544.16

**РАСЧЕТ И СВОЙСТВА 3D-МОЛЕКУЛЯРНОЙ ЭНТРОПИИ ШЕННОНА
НА ОСНОВЕ ГИСТОГРАММ МЕЖАТОМНЫХ РАССТОЯНИЙ****Вениамин Юрьевич Григорьев¹, Людмила Дмитриевна Григорьева²,
Олег Викторович Тиньков³**

¹ Институт физиологически активных веществ Федерального государственного бюджетного учреждения науки Федерального исследовательского центра проблем химической физики и медицинской химии Российской академии наук, лаборатория количественной связи структура – активность, Черноголовка, Россия

² Московский государственный университет имени М.В. Ломоносова, факультет фундаментальной физико-химической инженерии, Москва, Россия

³ Приднестровский государственный университет имени Т.Г. Шевченко, кафедра фармакологии и фармацевтической химии медицинского факультета, Тирасполь, Приднестровье, Молдова

Автор, ответственный за переписку: Вениамин Юрьевич Григорьев,
beng@ipac.ac.ru

Аннотация. С использованием гистограмм межатомных расстояний (ГМА) рассчитаны энтропии Шеннона (H) для ряда молекул. Учитывая, что величины H , полученные на основе обычных ГМА, не отражают в должной мере геометрическую структуру молекул, предложено использовать модифицированные гистограммы межатомных расстояний путем расчета межатомных расстояний в интервале 0–10 Å и низкоинтенсивного заполнения нулевых интервалов. Выявлены основные структурные факторы, влияющие на величину H . Исследована возможность использования энтропии Шеннона в качестве молекулярного дескриптора при разработке новых функциональных материалов.

Ключевые слова: энтропия Шеннона, гистограмма, молекулярное моделирование, QSAR, QSPR

DOI: 10.55959/MSU0579-9384-2-2025-66-2-124-137

Финансирование. Часть работы выполнена в рамках Государственного задания Института физиологически активных веществ Российской академии наук 2024 года (тема № FFSG-2024-0019).

Для цитирования: Григорьев В.Ю., Григорьева Л.Д., Тиньков О.В. Расчет и свойства 3D-молекулярной энтропии Шеннона на основе гистограмм межатомных расстояний // Вестн. Моск. ун-та. Сер. 2. Химия. 2025. Т. 66. № 2. С. 124–137.

ORIGINAL ARTICLE

CALCULATION AND PROPERTIES OF 3D MOLECULAR SHANNON ENTROPY USING INTERATOMIC DISTANCE HISTOGRAMS

Veniamin Y. Grigorev¹, Ludmila D. Grigoreva², Oleg V. Tinkov³

¹ Institute of Physiologically Active Compounds at Federal Research Center of Problems of Chemical Physics and Medicinal Chemistry, Russian Academy of Sciences, 142432, Severniy proezd 1, Chernogolovka, Moscow region, Russia

² Department of Fundamental Physical and Chemical Engineering, Moscow State University, 119991, Leninskiye Gory 1/51, Moscow, Russia

³ Department of Pharmacology and Pharmaceutical Chemistry, Medical Faculty, Transnistrian State University, 3300, October 25 str. 128, Tiraspol, Transdnestria, Moldova

Corresponding author: Veniamin Y. Grigorev, beng@ipac.ac.ru

Abstract. Interatomic distance histograms (IDH) are used to calculate the Shannon entropies (H) of a number of molecules. Considering that the H values obtained on the basis of conventional IDH do not adequately reflect the geometric structure of molecules, it was proposed to use modified histograms of interatomic distances by calculating interatomic distances in the range of 0–10 Å and filling zero bins with low intensity. The main structural factors affecting the H values are revealed. The possibility of using the Shannon entropy as a molecular descriptor in the development of new functional materials is studied.

Keywords: Shannon entropy, histogram, molecular modeling, QSAR, QSPR

Financial Support. Part of this work was supported by the budget of the Institute of Physiologically Active Compounds of the Russian Academy of Sciences (IPAC RAS) Targets – 2024 (topic No. FFSG-2024-0019).

For citation: Grigorev V.Y., Grigoreva L.D., Tinkov O.V. Calculation and Properties of 3D Molecular Shannon Entropy Using Interatomic Distance Histograms // Vestn. Mosk. un-ta. Ser. 2. Chemistry. 2025. T. 66. № 2. S. 124–137.

Концепция энтропии [1] играет большую роль в современном естествознании. Этим вопросом занимались многие выдающиеся ученые. Известны, например, энтропия Клаузиуса в термодинамике, энтропия Больцмана в статистической физике, энтропия Шеннона в информационной теории, энтропия Колмогорова в теории динамических систем и энтропия фон Неймана в квантовой механике. Понятие энтропии широко используется и применяется в разных областях науки [2, 3], таких как физика [4], химия [5], биология [6], лингвистика [7], экономика [8], инженерия [9], био/хемоинформатика [10] и др. К основными видами энтропии относятся энтропия в термодинамике и статистической механике, информационная энтропия и энтропия динамических систем [11]. Особое место занимает энтропия Шеннона (H) [12], поскольку в этом случае удалось идентифицировать и количественно охарактеризовать информацию, что

послужило основой для новой интерпретации энтропии: энтропия – это информация, которой не хватает для полного описания микросостояния системы.

Следует отметить успехи, которые были достигнуты при использовании информационной энтропии для количественного описания химической структуры молекул и физико-химических процессов, для структурного изучения кристаллов и исследования молекулярных ансамблей [13]. При этом отмечается, что способ использования информационной энтропии зависит от уровня детализации химической структуры. Существуют различные методы для описания молекулярных структур, которые учитывают эмпирические формулы, молекулярные графы, геометрию или конформационную динамику. В настоящее время в качестве основного подхода используется представление структуры в виде

молекулярных графов. Однако следует отметить некоторые недостатки при использовании информационной энтропии, например, при этом подходе игнорируются семантические аспекты информации. Кроме того, при использовании графов для характеристики структуры молекул принимаются во внимание только их топологические свойства и не учитываются трехмерные аспекты информации.

Для оценки энтропии на основе вероятностной функции распределения (PDF) используются различные методы [14], включая простой и часто применяемый метод гистограмм [15]. При конструировании гистограмм [16] необходимо выбрать начало отсчета и ширину интервала (w). Причем нужно отметить, что величина интервала оказывает существенное влияние на энтропию Шеннона [17, 18]. Использование слишком больших интервалов приводит к уменьшению значимой информации, а слишком малые интервалы ведут к увеличению шума. В литературе приведены различные рекомендации и формулы [19–22] по выбору оптимальных значений w и коррекции величин H . При этом в основном опираются на размер выборки и некоторую меру дисперсии PDF. В случае анализа экспериментальных данных разумным ориентиром при выборе ширины интервала может служить точность наблюдения [14].

Нужно подчеркнуть, что молекулярные гистограммы (спектры) межатомных расстояний (ГМА) давно и успешно используются для исследования связи «структура – свойство» на основе методологии QSAR/QSPR. Так, в одной из ранних публикаций [23] для описания структуры соединений было использовано молекулярное преобразование, являющееся аналогом функции, описывающей интенсивность рассеивания молекулой рентгеновского излучения. Дальнейшее развитие этой идеи привело к появлению функции радиального распределения, получаемой непосредственно из ГМА [24] и дескрипторов 3D-MoRSE [25]. В последующем для исследования свойств биологически активных веществ была разработана компьютерная программа MOLTRA [26]. Ее отличительной чертой являлось то, что для описания химических соединений использовалась информация, включающая в себя сведения о геометрической структуре и энергии внутримолекулярных взаимодействий. В литературе [27, 28] можно найти примеры успешного применения 3D-дескрипторов на основе молекулярной гео-

метрии при создании QSAR/QSPR-моделей. В настоящее время ГМА продолжает использоваться для представления структуры молекул при создании компьютерных моделей на основе методов машинного обучения [29]. При этом следует отметить, что ГМА ранее не применялась для расчета энтропии Шеннона.

Цель настоящей работы – изучение свойств 3D-молекулярной энтропии Шеннона, рассчитанной на основе гистограмм межатомных расстояний.

Объекты и методы исследования

В качестве объектов исследования использовали геометрические структуры двух групп соединений. Первая группа состояла из 25 различных органических соединений, структура которых в газовой фазе была определена экспериментально с использованием методов электронной дифракции и микроволновой спектроскопии [30]. Вторая группа соединений включала в себя 75 алканов, структура которых была оптимизирована без проведения полного конформационного анализа на основе полуэмпирического квантово-химического метода PM3 (пакет программ HyperChem [31], RMS = 0,001 ккал/(Å моль), алгоритм Полака – Рибьера, стартовые молекулы для оптимизации сгенерированы программой из 2D-структур).

Для расчета молекулярной энтропии Шеннона (H) применяли гистограммы межатомных расстояний. При выборе величин w исходили из того, что они должны соответствовать точности измерения/расчета исследуемых межатомных расстояний. В частности, точность определения экспериментальных межатомных расстояний для малых молекул может составлять 0,001 Å [32], а в случае больших молекул (белков) при разрешении 1 Å точность определения межатомных расстояний может достигать 0,01 Å [33]. С учетом этого в качестве рабочей области w целесообразно рассматривать величины 0,001–0,01 Å. Границами ГМА служили величины $r_{\min} = 0$ Å и $r_{\max} = 10$ Å с учетом максимального размера малых молекул, используемых в качестве лекарств [34]. В случае больших биомолекул этот диапазон может быть расширен.

Энтропию Шеннона определяли на основе оценки максимального правдоподобия [35]:

$$H = - \sum_{i=1}^m p_i \log_2 p_i, \quad (1)$$

где p_i – вероятность попадания межатомного расстояния в i -й интервал; m – число интервалов с $p_i > 0$.

Для расчета линейных регрессионных моделей применяли программу SVD из библиотеки программ IMSL [36]. В качестве статистических характеристик моделей использовали: n – число молекул, R^2 – квадрат коэффициента линейной корреляции, R_{cv}^2 – квадрат коэффициента линейной корреляции в условиях перекрестного контроля, $gmse$ – среднее квадратическое отклонение, $gmse_{cv}$ – среднее квадратическое отклонение в условиях перекрестного контроля, s – стандартное отклонение, Δ – стандартная ошибка коэффициента, R_p^2 – метрика для оценки случайной корреляции, которую вычисляли по формуле [37]:

$$R_p^2 = R \cdot (R^2 - R_{rand}^2)^{0,5},$$

где R_{rand}^2 – квадрат рандомизированного коэффициента линейной корреляции.

Для внутренней валидации моделей применяли пятикратный перекрестный контроль. Рандомизацию проводили с использованием 100 итераций. При анализе зависимости энтропии Шеннона от ширины интервала использовали ряд $w = 0,001 \text{ \AA}$; $w = 0,002 \text{ \AA}$; $w = 0,004 \text{ \AA}$; $w = 0,01 \text{ \AA}$; $w = 0,02$; $w = 0,04 \text{ \AA}$; $w = 0,1 \text{ \AA}$; $w = 0,2 \text{ \AA}$; $w = 0,4 \text{ \AA}$; $w = 1,0 \text{ \AA}$.

Результаты и их обсуждение

Ранее [38] уже отмечались недостатки энтропии Шеннона при ее использовании. В частности, энтропия Шеннона отражает только син-

таксические аспекты информации, игнорируя ее семантические аспекты [39]. В соответствии с этим при расчете H рассматриваются только события с ненулевой вероятностью [40]. В случае ГМА это означает, что во внимание не принимаются нулевые межатомные расстояния. С учетом вышесказанного прямое использование уравнения (1) по отношению к обычным ГМА выглядит проблематичным. Так, например, совершенно неразличимыми становятся изоструктурные молекулы H_2O и H_2S ($H = 0,918$ бит, $m = 2, w = 0,001 \text{ \AA}$) или уксусной и трифторуксусной кислоты ($H = 4,084$ бит, $m = 19, w = 0,001 \text{ \AA}$). В случае молекулярных ГМА очевидно, что значимая информация о геометрической структуре молекул содержится как в интервалах с $p > 0$, так и в интервалах с $p = 0$. Одним из путей для преодоления отмеченных недостатков может служить та или иная модификация гистограмм [41–43]. Для повышения информативности и устойчивости оценок энтропии Шеннона был использован единый опорный диапазон межатомных расстояний от 0 до 10 \AA , а также проведена модификация обычных гистограмм путем преобразования нулевых интервалов в ненулевые с интенсивностью, равной $1/m_0$, где m_0 – общее число интервалов, заключенных между 0 и максимальным межатомным расстоянием. Пример расчета модифицированной ГМА представлен на рис. 1.

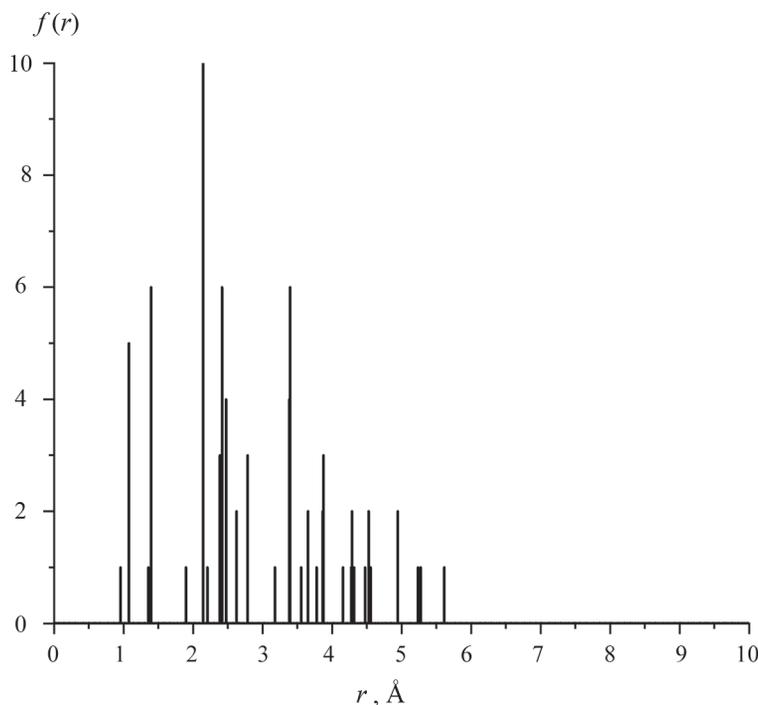


Рис. 1. Модифицированная гистограмма межатомных расстояний фенола ($w = 0,01 \text{ \AA}$)

Результаты расчетов энтропии Шеннона для 25 молекул на основе модифицированных гистограмм при значениях w , равных 0,001; 0,002; 0,004 и 0,01 Å представлены в табл. 1.

Полученные результаты (рис. 2) ясно указывают на отсутствие корреляционной связи между H и m . Это дает возможность предположить, что энтропия Шеннона может успешно отражать геометрическую структуру мо-

лекул. Так, в случае изоструктурных молекул H_2O и H_2S величина H ($w = 0,001$ Å) составляет 4,138 и 4,225 бит, а в случае уксусной и трифторуксусной кислот 4,568 и 4,665 бит соответственно. При рассмотрении свойств молекулярной энтропии Шеннона нельзя обойти стороной вопрос о влиянии ширины интервала гистограммы на величину H . В этих целях мы проанализировали зависимость H от w в виде

Т а б л и ц а 1

Молекулярные энтропии Шеннона (H , бит) и число интервалов (m), полученные из модифицированных гистограмм, при различной ширине интервалов (w , Å)

Соединение	CAS RN	w , Å							
		0,001		0,002		0,004		0,01	
		m	H	m	H	m	H	m	H
1,3-Бутадиен	106-99-0	5591	5,107	2795	5,084	1397	5,016	559	4,980
1,3-Бутадиин	460-12-8	5999	4,033	2999	3,970	1499	3,905	600	3,817
2,2'-Дихлоробифенил	13029-08-8	9295	5,955	4647	5,950	2323	5,945	929	5,861
3-Метилацетонитрил	630-18-2	4474	4,694	2237	4,606	1118	4,369	447	4,296
2-Бутанон	78-93-3	5599	4,936	2799	4,922	1399	4,873	560	4,851
Уксусная кислота	64-19-7	3884	4,568	1942	4,532	971	4,494	388	4,438
Ацетонитрил	75-05-8	3176	3,708	1588	3,644	794	3,407	317	3,487
Анилин	62-53-3	5899	5,179	2949	5,167	1474	5,076	589	5,014
Бензол	71-43-2	5000	3,434	2500	3,419	1250	3,403	500	3,381
Хлорбензол	108-90-7	5619	3,991	2809	3,976	1404	3,959	562	3,937
Циклогексан	110-82-7	5021	5,127	2510	4,814	1255	4,336	502	4,316
Диметоксиметан	109-87-5	6482	4,828	3241	4,815	1620	4,801	648	4,730
Диметиловый эфир	115-10-6	4149	3,627	2074	3,599	1037	3,570	414	3,530
Диметилсульфон	67-71-0	4651	4,692	2325	4,673	1162	4,653	465	4,622
Нафталин	91-20-3	7158	5,427	3579	5,420	1789	5,413	715	5,349
Фенол	108-95-2	5624	5,248	2812	5,183	1406	4,925	562	4,719
Пропан	74-98-6	4372	4,501	2186	4,482	1093	4,390	437	4,290
Пропен	115-07-1	4267	5,114	2133	5,085	1066	5,054	426	4,954
Пропин	74-99-7	4231	3,925	2115	3,878	1057	3,830	423	3,763
Пиридин	110-86-1	4886	5,453	2443	5,255	1221	5,052	488	4,637
Тетрацианоэтилен	670-54-2	6007	4,144	3003	4,121	1501	4,098	600	3,892
Тетрагидрофуран	109-99-9	4298	5,549	2149	5,509	1074	5,377	429	5,275
Толуол	108-88-3	5895	5,321	2947	5,266	1473	5,020	589	5,033
Трифторуксусная кислота	76-05-1	4051	4,665	2025	4,629	1012	4,496	405	4,441
Триметиламин	75-50-3	4191	4,102	2095	4,089	1047	4,075	419	3,951

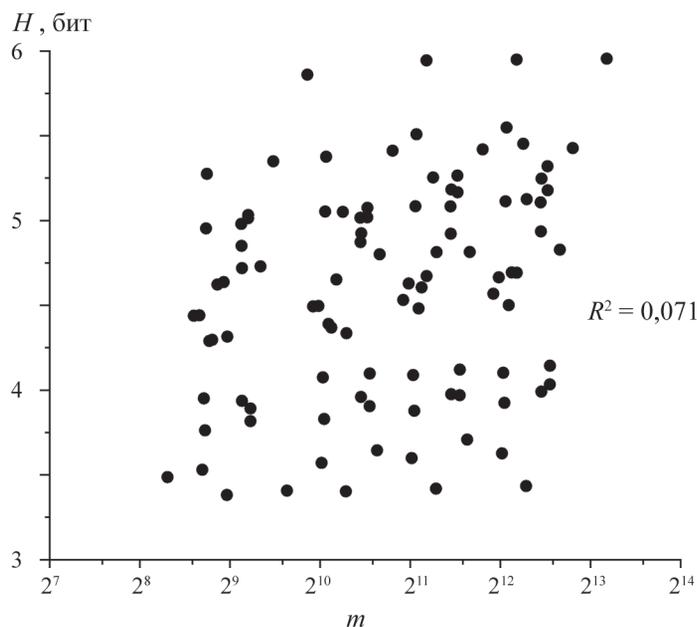


Рис. 2. Зависимость энтропии Шеннона, рассчитанной на основе модифицированных гистограмм, от числа интервалов

Т а б л и ц а 2

Статистические характеристики зависимости энтропии Шеннона (H , бит) от ширины интервала (w , Å) для модели $H = a + b \cdot \log_2(w)$

Соединение	CAS RN	n	R^2	s	a	$\pm\Delta$	b	$\pm\Delta$
1,3-Бутадиен	106-99-0	5	0,975	0,012	4,725	0,028	0,039	0,004
1,3-Бутадиин	460-12-8	3	0,999	0,001	3,394	0,002	0,064	0,001
2,2'-Дихлоробифенил	13029-08-8	3	0,999	0,001	5,904	0,001	0,005	0,001
3-Метилацетонитрил	630-18-2	8	0,972	0,072	3,288	0,067	0,145	0,010
2-Бутанон	78-93-3	5	0,962	0,010	4,683	0,023	0,026	0,003
Уксусная кислота	64-19-7	3	0,999	0,001	4,197	0,004	0,037	0,001
Ацетонитрил	75-05-8	7	0,935	0,071	2,684	0,087	0,103	0,012
Анилин	62-53-3	6	0,940	0,048	4,390	0,080	0,084	0,011
Бензол	71-43-2	3	0,999	0,001	3,279	0,001	0,016	0,001
Хлорбензол	108-90-7	3	0,999	0,001	3,834	0,001	0,016	0,001
Циклогексан	110-82-7	3	0,986	0,067	1,214	0,428	0,395	0,048
Диметоксиметан	109-87-5	3	0,999	0,001	4,691	0,002	0,014	0,001
Диметиловый эфир	115-10-6	3	0,999	0,001	3,344	0,002	0,028	0,001
Диметилсульфон	67-71-0	3	0,999	0,001	4,495	0,003	0,020	0,001
Нафталин	91-20-3	3	0,999	0,001	5,354	0,001	0,007	0,001
Фенол	108-95-2	7	0,985	0,051	3,694	0,062	0,159	0,009
Пропан	74-98-6	9	0,944	0,150	2,819	0,111	0,191	0,018
Пропен	115-07-1	3	0,999	0,001	4,814	0,005	0,030	0,001
Пропин	74-99-7	3	0,999	0,001	3,451	0,003	0,048	0,001
Пиридин	110-86-1	3	0,999	0,002	3,454	0,014	0,201	0,002
Тетрацианоэтилен	670-54-2	3	0,999	0,001	3,916	0,002	0,023	0,001
Тетрагидрофуран	109-99-9	4	0,970	0,027	4,698	0,092	0,087	0,011
Толуол	108-88-3	7	0,927	0,089	4,149	0,108	0,121	0,015
Трифторуксусная кислота	76-05-1	6	0,969	0,026	4,026	0,043	0,064	0,006
Триметиламин	75-50-3	3	0,999	0,001	3,967	0,002	0,014	0,001

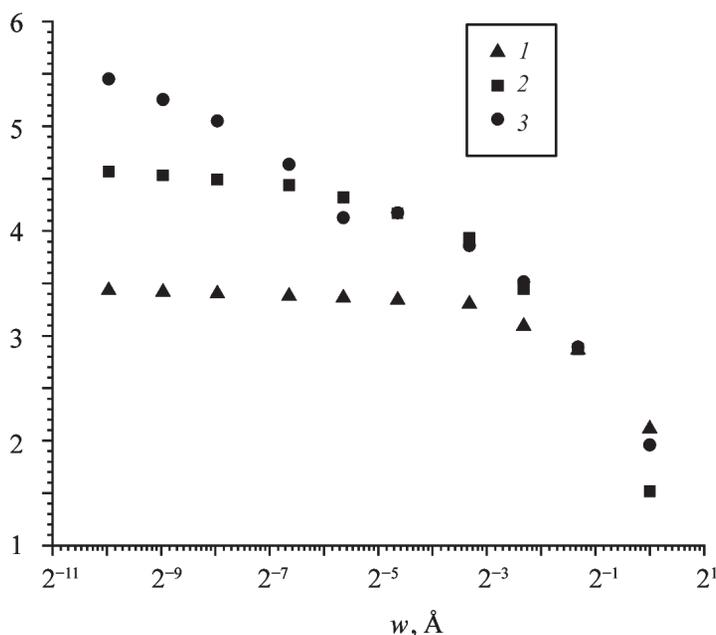


Рис. 3. Зависимость энтропии Шеннона, рассчитанной на основе модифицированных гистограмм, от ширины интервала (1 – бензол, 2 – уксусная кислота, 3 – пиридин)

$H = a + b \cdot \log_2(w)$ (рис. 3, табл. 2). Из представленных данных следует, что H уменьшается с увеличением интервала гистограммы. В целом эта зависимость имеет нелинейный характер. Однако при небольших значениях величины w можно выделить прямолинейные участки, статистические характеристики которых представлены в табл. 2. Полученный результат вполне согласуется с литературными данными [44] о существовании линейной зависимости между энтропией Шеннона и масштабным параметром. Очевидно, что угловой коэффициент b варьирует в достаточно больших пределах и может выступать в качестве меры молекулярной «чувствительности». Таким образом, наряду с величиной H он может служить еще одним 3D-молекулярным дескриптором. При этом обращает на себя внимание тот факт, что молекулы бензола и пиридина, имеющие сходное электронное строение и близкое число атомов, обладают различными величинами энтропии Шеннона (соответственно 3,434 и 5,453 бит при $w = 0,001 \text{ \AA}$) и величинами b (соответственно 0,016 и 0,201). Вероятно, это связано с глубокими различиями в структуре гистограмм (рис. 4). Известно, что энтропия Шеннона достигает своего максимума в случае одинаковой интенсивности анализируемой переменной. ГМА пиридина имеет большое число одинаковых величин с $f(r) = 1$ и

$f(r) = 2$. Это, вероятно, и приводит к наблюдаемым различиям в значениях H и b .

Для выяснения структурных факторов, влияющих на величину H , была использована группа соединений, состоящая из 75 алканов $C_1 \dots C_9$. Результаты расчетов приведены в табл. 3 и на рис. 5.

Из анализа полученных данных следует, что энтропия Шеннона возрастает с увеличением числа атомов (N) в молекуле. Однако при одном и том же числе атомов величины H могут значительно варьировать. Так, для молекул ($N = 17$) 2-метилбутана, пентана и 2,2-диметилпропана энтропия Шеннона составила 6,067; 5,035 и 3,384 бит соответственно. Это можно объяснить тем, что рассматриваемые соединения относятся к разным точечным группам симметрии. Первая молекула относится к группе C_1 , вторая – к C_{2v} , а третья – к T_d . Отметим при этом, что группа C_1 включает в качестве операции симметрии $\{E\}$, C_{2v} содержит $\{E, C_2, \sigma_v\}$, а T_d состоит из $\{E, C_3, C_2, \sigma, S_4\}$, т.е. с увеличением числа элементов симметрии в группе величина H уменьшается. Это связано с уменьшением числа уникальных межатомных расстояний (при сохранении их общего числа). Это приводит, в свою очередь, к упрощению дискретных PDF межатомных расстояний и, как следствие, к уменьшению энтропии Шеннона. Сходные результаты были получены при анализе информационной энтропии с использованием

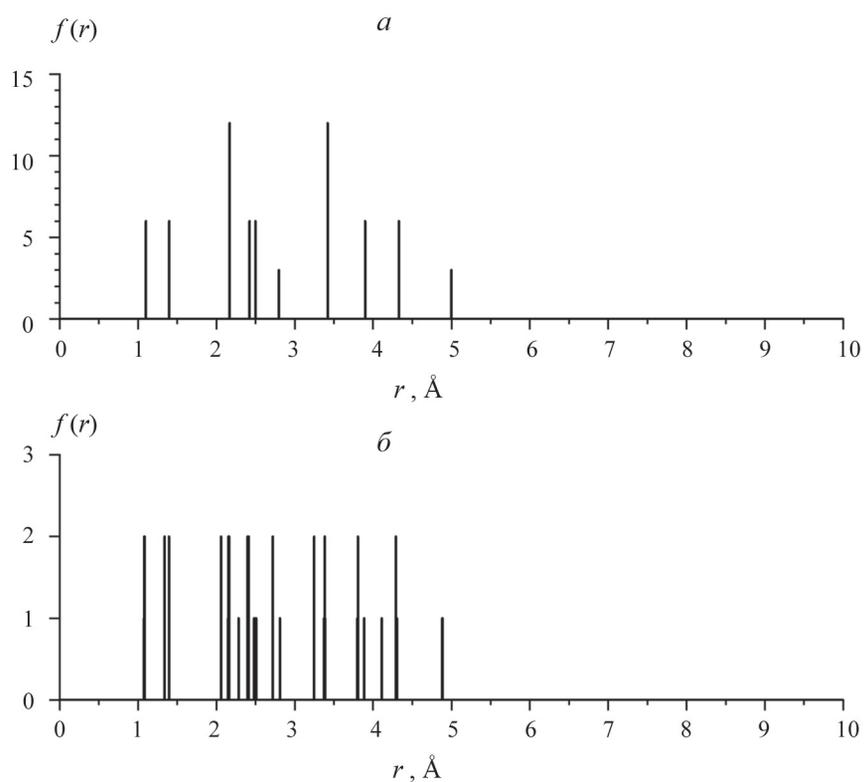


Рис. 4. Модифицированные гистограммы межатомных расстояний бензола (а) и пиридина (б) при $w = 0,001$ Å

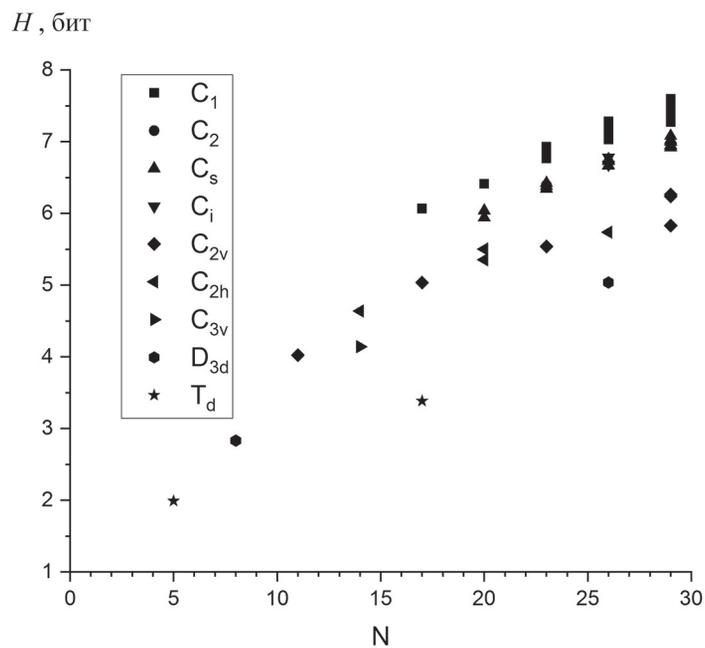


Рис. 5. Зависимость энтропии Шеннона, рассчитанной на основе модифицированных гистограмм ($w = 0,01$ Å), от числа атомов и точечной группы симметрии молекул

Т а б л и ц а 3

Число атомов (N), точечные группы симметрии (S) и энтропии Шеннона (H , бит) для алканов $C_1 \dots C_9$, полученные из модифицированных гистограмм, при ширине интервала $w = 0,01 \text{ \AA}$

Соединение	CAS RN	N	S	H
Метан	74-82-8	5	T_d	1,990
Этан	74-84-0	8	D_{3d}	2,832
Пропан	74-98-6	11	C_{2v}	4,025
Бутан	106-97-8	14	C_{2h}	4,638
2-Метилпропан	75-28-5	14	C_{3v}	4,141
Пентан	109-66-0	17	C_{2v}	5,035
2-Метилбутан	78-78-4	17	C_1	6,067
2,2-Диметилпропан	463-82-1	17	T_d	3,384
Гексан	110-54-3	20	C_{2h}	5,355
2-Метилпентан	107-83-5	20	C_1	6,414
3-Метилпентан	96-14-0	20	C_s	6,039
2,2-Диметилбутан	75-83-2	20	C_s	5,943
2,3-Диметилбутан	79-29-8	20	C_{2h}	5,502
Гептан	142-82-5	23	C_{2v}	5,539
2-Метилгексан	591-76-4	23	C_1	6,896
3-Метилгексан	589-34-4	23	C_1	6,769
2,2-Диметилпентан	590-35-2	23	C_s	6,382
2,3-Диметилпентан	565-59-3	23	C_1	6,928
2,4-Диметилпентан	108-08-7	23	C_s	6,423
3,3-Диметилпентан	562-49-2	23	C_2	6,393
3-Этилпентан	617-78-7	23	C_1	6,771
2,2,3-Триметилбутан	464-06-2	23	C_s	6,345
Октан	111-65-9	26	C_{2h}	5,738
2-Метилгептан	592-27-8	26	C_1	7,167
3-Метилгептан	589-81-1	26	C_1	7,118
4-Метилгептан	589-53-7	26	C_s	6,671
3-Этилгексан	619-99-8	26	C_1	7,122
2,2-Диметилгексан	590-73-8	26	C_s	6,732
2,3-Диметилгексан	584-94-1	26	C_1	7,222
2,4-Диметилгексан	589-43-5	26	C_1	7,282
2,5-Диметилгексан	592-13-2	26	C_i	6,782
3,3-Диметилгексан	563-16-6	26	C_s	6,757
3,4-Диметилгексан	583-48-2	26	C_i	6,681
3-Этил-2-метилпентан	609-26-7	26	C_1	7,206
3-Этил-3-метилпентан	1067-08-9	26	C_1	7,156
2,2,3-Триметилпентан	564-02-3	26	C_1	7,095

Окончание табл. 3

2,2,4-Триметилпентан	540-84-1	26	C ₁	7,031
2,3,3-Триметилпентан	560-21-4	26	C ₁	7,150
2,3,4-Триметилпентан	565-75-3	26	C _s	6,664
2,2,3,3-Тетраметилбутан	594-82-1	26	D _{3d}	5,036
Нонан	111-84-2	29	C _{2v}	5,828
2-Метилоктан	3221-61-2	29	C ₁	7,390
3-Метилоктан	2216-33-3	29	C ₁	7,400
4-Метилоктан	2216-34-4	29	C ₁	7,273
3-Этилгептан	15869-80-4	29	C ₁	7,433
4-Этилгептан	2216-32-2	29	C ₁	7,354
2,2-Диметилгептан	1071-26-7	29	C _s	6,998
2,3-Диметилгептан	3074-71-3	29	C ₁	7,471
2,4-Диметилгептан	2213-23-2	29	C ₁	7,594
2,5-Диметилгептан	2216-30-0	29	C ₁	7,557
2,6-Диметилгептан	1072-05-5	29	C _s	7,001
3,3-Диметилгептан	4032-86-4	29	C _s	7,024
3,4-Диметилгептан	922-28-1	29	C ₁	7,369
3,5-Диметилгептан	926-82-9	29	C _s	7,080
4,4-Диметилгептан	1068-19-5	29	C _{2v}	6,264
3-Этил-2-метилгексан	16789-46-1	29	C ₁	7,391
4-Этил-2-метилгексан	3074-75-7	29	C ₁	7,517
3-Этил-3-метилгексан	3074-76-8	29	C ₁	7,473
3-Этил-4-метилгексан	3074-77-9	29	C ₁	7,479
2,2,3-Триметилгексан	16747-25-4	29	C ₁	7,485
2,2,4-Триметилгексан	16747-26-5	29	C ₁	7,437
2,2,5-Триметилгексан	3522-94-9	29	C ₁	7,355
2,3,3-Триметилгексан	16747-28-7	29	C ₁	7,503
2,3,4-Триметилгексан	921-47-1	29	C ₁	7,500
2,3,5-Триметилгексан	1069-53-0	29	C ₁	7,513
2,4,4-Триметилгексан	16747-30-1	29	C ₁	7,518
3,3,4-Триметилгексан	16747-31-2	29	C ₁	7,512
3,3-Диэтилпентан	1067-20-5	29	C _s	6,950
2,2-Диметил-3-этилпентан	16747-32-3	29	C ₁	7,449
2,3-Диметил-3-этилпентан	16747-33-4	29	C ₁	7,427
2,4-Диметил-3-этилпентан	1068-87-7	29	C ₁	7,570
2,2,3,3-Тетраметилпентан	7154-79-2	29	C _s	6,918
2,2,3,4-Тетраметилпентан	1186-53-4	29	C ₁	7,449
2,2,4,4-Тетраметилпентан	1070-87-7	29	C _{2v}	6,240
2,3,3,4-Тетраметилпентан	16747-38-9	29	C _s	6,947

молекулярных графов. В частности, для ряда фуллеренов [45] было установлено, что информационная энтропия при фиксированной симметрии растет с увеличением числа атомов и падает с увеличением числа вращательной симметрии при фиксированном числе атомов.

Для количественной оценки влияния структурных факторов на молекулярную энтропию Шеннона были сконструированы простые линейные модели на основе соединений с одинаковой точечной группой симметрии:

$$C_1: H = 4,273(\pm 0,115) + 0,110(\pm 0,004) N \quad (2)$$

$$n = 41; R^2 = 0,946; \text{rmse} = 0,074; R_{cv}^2 = 0,939; \text{rmse}_{cv} = 0,078; R_p^2 = 0,933$$

$$C_s: H = 3,885(\pm 0,113) + 0,107(\pm 0,004) N \quad (3)$$

$$n = 16; R^2 = 0,978; \text{rmse} = 0,051; R_{cv}^2 = 0,971; \text{rmse}_{cv} = 0,059; R_p^2 = 0,945$$

$$C_{2h} + C_{2v}: H = 3,167(\pm 0,232) + 0,103(\pm 0,010) N \quad (4)$$

$$n = 10; R^2 = 0,927; \text{rmse} = 0,179; R_{cv}^2 = 0,854; \text{rmse}_{cv} = 0,253; R_p^2 = 0,872$$

$$D_{3d}: \text{угловой коэффициент} = (5,036 - 2,832)/(26 - 8) = 0,122 \quad (n = 2) \quad (5)$$

$$T_d: \text{угловой коэффициент} = (3,384 - 1,990)/(17 - 5) = 0,116 \quad (n = 2) \quad (6)$$

Следует отметить, что регрессионные модели (2–4) обладают удовлетворительными статистическими характеристиками и могут быть использованы для анализа зависимости «структура – свойство». Из-за малого количества данных модели (5)–(6) имеют оценочный характер. Анализируемые точечные группы симметрии можно разбить на две категории: низшую (C_1 , C_s , C_{2h} , C_{2v}) (I) и высшую (D_{3d} , T_d) (II). Нетрудно заметить, что между этими категориями и величинами угловых коэффициентов моделей (2)–(6) существует соответствие:

$$I < II \text{ и } 0,107 \text{ (среднее)} < 0,119 \text{ (среднее)}.$$

Т а б л и ц а 4

Межплоскостные расстояния (d_{001} , нм) Na^+ -монтмориллонита, модифицированного четвертичными алкиламмониевыми катионами, и молекулярные дескрипторы катионов: протонодонорный дескриптор (ΣC_d), энтропии Шеннона (H , бит; $w = 0,01 \text{ \AA}$), полученные на основе различных областей ГМА

Катион	d_{001}	ΣC_d	H_{tot}	H_{val}	H_{unb}
$C_4H_9N(CH_3)_3^+$	1,42	0,00	6,729	2,135	6,574
$C_8H_{17}N(CH_3)_3^+$	1,42	0,00	7,475	1,979	7,467
$C_{12}H_{25}N(CH_3)_3^+$	1,66	0,00	7,795	1,840	7,817
$C_{14}H_{29}N(CH_3)_3^+$	1,77	0,00	7,912	1,782	7,938
$C_{16}H_{33}N(CH_3)_3^+$	1,81	0,00	7,980	1,732	8,004
$C_{18}H_{37}N(CH_3)_3^+$	1,85	0,00	8,122	1,687	8,156
$C_4H_9NH_3^+$	1,36	-7,50	5,730	2,481	5,312
$C_6H_{13}NH_3^+$	1,36	-7,50	6,250	2,308	6,066
$C_8H_{17}NH_3^+$	1,36	-7,50	6,579	2,142	6,501
$C_{10}H_{21}NH_3^+$	1,45	-7,50	6,866	2,009	6,853
$C_{12}H_{25}NH_3^+$	1,70	-7,50	7,044	1,902	7,054
$C_{14}H_{29}NH_3^+$	1,75	-7,50	7,170	1,816	7,172
$C_{16}H_{33}NH_3^+$	1,75	-7,50	7,305	1,745	7,314
$C_{18}H_{37}NH_3^+$	1,85	-7,50	7,401	1,684	7,411
$(C_6H_{13})_2NH_2^+$	1,40	-6,00	7,674	2,119	7,722
$(C_8H_{17})_2NH_2^+$	1,80	-6,00	8,436	1,946	8,464
$(C_9H_{19})_2NH_2^+$	1,90	-6,00	8,768	1,875	8,803
$(C_6H_{13})_3NH^+$	1,80	-4,31	8,935	2,085	9,063
$(C_8H_{17})_3NH^+$	1,90	-4,31	9,415	1,915	9,553

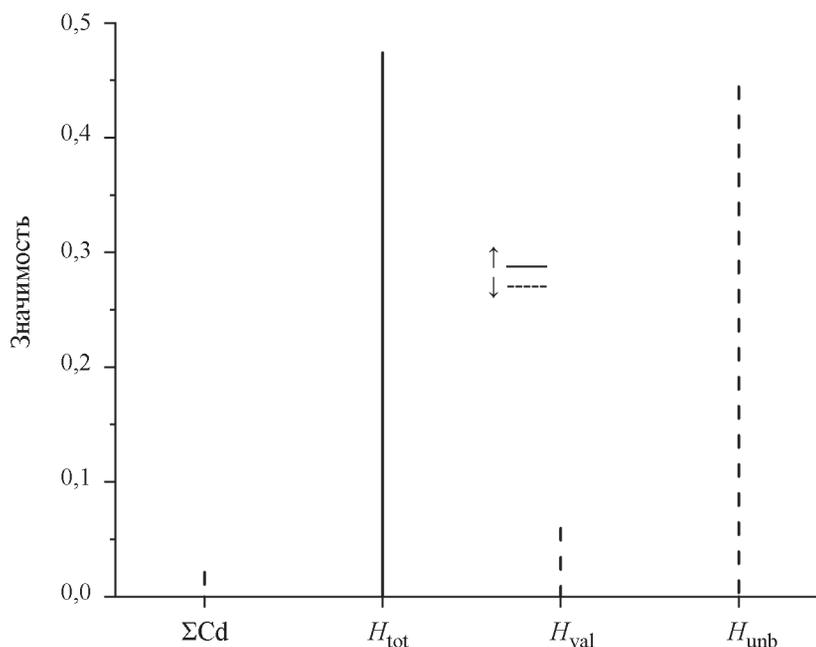


Рис. 6. Относительная значимость (0...1) и вклад (↑/↓) молекулярных дескрипторов

Выявленные данные о зависимости молекулярной энтропии Шеннона от числа атомов и точечной группы симметрии во многом схожи с результатами исследования фрактальной размерности молекул [46]. Очевидно, это является результатом тесной связи информационно-энтропийных и фрактальных свойств объектов различной природы [44, 47–49].

Полученные в настоящей работе 3D-молекулярные энтропии Шеннона могут найти применение при разработке новых функциональных материалов. Так, с инженерной точки зрения, важной задачей является получение органомодифицированных глин с заданными химическими и механическими свойствами, включая Na^+ -монтмориллонит, модифицированный четвертичными аммониевыми катионами [50]. Для конструирования QSPR-модели, связывающей межплоскостное расстояние d_{001} (нм) монтмориллонита с молекулярными дескрипторами катионов, использовали выборку [51, 52] из 19 соединений (табл. 4).

В качестве молекулярных дескрипторов применяли протоноакцепторный дескриптор ΣC_d , а также значения энтропии Шеннона, рассчитанные для различных областей ГМА при $w = 0,01 \text{ \AA}$. В частности, для расчета H_{tot} (общая область) использовали все межатомные расстояния в молекуле, для H_{val} (валентная область) – только валентные межатомные расстояния, а для H_{unb}

(несвязанная область) – только расстояния между несвязанными атомами за исключением пар атомов, находящихся в ван-дер-ваальсовом контакте.

$$d_{001} = 1,38(\pm 0,28) - 0,019(\pm 0,005)\Sigma C_d + 1,47(\pm 0,31) H_{tot} - 0,81(\pm 0,11) H_{val} - 1,23(\pm 0,28) H_{unb};$$

$$n = 19; R^2 = 0,925; \text{rmse} = 0,06; R^2_{cv} = 0,859; \text{rmse}_{cv} = 0,08; R^2_p = 0,806. \quad (7)$$

Полученная линейная QSPR-модель имеет удовлетворительные статистические характеристики и может быть вполне интерпретирована (рис. 6). Из представленных данных следует, что наибольший вклад в модель (7) вносят величины H_{tot} (увеличение d_{001}) и H_{unb} (уменьшение d_{001}), вклад остальных дескрипторов незначителен. Величина H_{tot} характеризует общий размер внедряемого в монтмориллонит катиона, рост которого приводит к увеличению межплоскостных расстояний. Величину H_{unb} можно рассматривать в качестве меры молекулярной «рыхлости», которая растет с увеличением размера катиона и способствует уменьшению d_{001} . Таким образом, энтропийные дескрипторы оказывают существенное влияние на формирование межплоскостных расстояний в модифицированном Na^+ -монтмориллоните.

Заключение

В настоящей работе на основе гистограмм межатомных расстояний были рассчитаны

3D-молекулярные энтропии Шеннона. Учитывая, что величины H , полученные с использованием обычных ГМА, не отражают в должной мере геометрическую структуру молекул, была проведена модификация гистограмм, которая заключалась в расчете межатомных расстояний в интервале 0–10 Å и низкоинтенсивном заполнении интервалов с нулевой вероятностью. При анализе свойств энтропии Шеннона было установлено, что зависимость H от $\log_2(w)$ может быть выражена в виде линейных моделей в рамках интервала $w = 0,001$ –1 Å. Угловым коэффициентом этой модели зависит от структуры и может, таким образом, выступать

в качестве молекулярного 3D-дескриптора. Основными структурными факторами, влияющими на величину энтропии Шеннона, являются число атомов и точечная группа симметрии молекулы. Полученные данные могут быть использованы для описания структуры молекул при разработке новых функциональных материалов. В качестве дальнейшего направления исследований можно выделить изучение связи энтропии Шеннона с фрактальной размерностью молекул. Также перспективным выглядит исследование зависимости между молекулярной энтропией Шеннона и термодинамической энтропией.

СПИСОК ЛИТЕРАТУРЫ

- Wehrl A. // *Rev. Mod. Phys.* 1978. Vol. 50. P. 221.
- Popovic M.E. // *Therm. Sci.* 2018. Vol. 22. P. 1163.
- Ribeiro M., Henriques T., Castro L., Souto A., Antunes L., Costa-Santos C., Teixeira A. // *Entropy*. 2021. Vol. 23. P. 222.
- Heusler S., Dür W., Ubben M.S., Hartmann A. // *J. Phys. A: Math. Theor.* 2022. Vol. 55. P. 404006.
- Yan X.H., Li J.S., Zhang W.R., Zhang Y. // *Mater. Chem. Phys.* 2018. Vol. 210. P. 12.
- Roach T.N.F. // *Entropy*. 2020. Vol. 22. P. 1335.
- Vera J., Fuentealba D., Lopez M., Ponce H., Zariquiey R. // *EPL*. 2021. Vol. 136. P. 68003.
- Vozna L.Yu. // *Journal of Heterodox Economics*. 2016. Vol. 3. P. 1.
- Feng X., Yue Y., Qiu J., Jain H., Zhou S. // *Fundam. Res.* 2022. Vol. 2. P. 783.
- You Q., Li C., Sun J., Palade V., Pan F. // *Mol. Inform.* 2023. Vol. 42. P. 2200080.
- Amigó J.M., Keller K., Unakafova V.A. // *Discrete Continuous Dyn. Syst. Ser. B*. 2015. Vol. 20. P. 3301.
- Shannon C.E. // *Bell Syst. Tech. J.* 1948. Vol. 27. P. 379.
- Sabirov D.S., Shepelevich I.S. // *Entropy*. 2021. Vol. 23. P. 1240.
- Gong W., Yang D., Gupta H.V., Nearing G. // *Water Resour. Res.* 2014. Vol. 50. P. 5003.
- Watts S.J., Crow L. // URL: <https://arxiv.org/abs/2210.02848>.
- Scott D.W. // *WIREs Comp. Stat.* 2010. Vol. 2. P. 44.
- Purwani S., Nahar J., Twining C. // *AIP Conference Proceedings*. 2017. Vol. 1868. P. 040008.
- Wallis K. // URL: <https://mpira.ub.uni-muenchen.de/52856/>.
- Birge L., Rozenholc Y. // *ESAIM: PS*. 2006. Vol. 10. P. 24.
- Güçlü B. // *Acta Neurobiol. Exp.* 2005. Vol. 65. P. 399.
- Wand M.P. // *Am. Stat.* 1997. Vol. 51. P. 59.
- Contreras Rodríguez L., Madarro-Capó E.J., Legón-Pérez C.M., Rojas O., Sosa-Gómez G. // *Entropy*. 2021. Vol. 23. P. 561.
- Soltzberg L.J., Wilkins C.L. // *J. Am. Chem. Soc.* 1977. Vol. 99. P. 439.
- Новиков В.П., Раевский О.А. // *Хим.-фарм. журн.* 1982. Т. 16. С. 574.
- Schuur J.H., Selzer P., Gasteiger J. // *J. Chem. Inf. Comput. Sci.* 1996. Vol. 36. P. 334.
- Трепалин С.В., Ярков А.В., Раздольский А.Н., Раевский О.А. // *Современные наукоемкие технологии*. 2014. № 12. С. 33.
- Klein C.T., Kaiser D., Ecker G. // *J. Chem. Inf. Comput. Sci.* 2004. Vol. 44. P. 200.
- Devinyak O., Havrylyuk D., Lesyk R. // *J. Mol. Graph. Model.* 2014. Vol. 54. P. 194.
- Faber F.A., Hutchison L., Huang B., Gilmer J., Schoenholz S.S., Dahl G.E., Vinyals O., Kearnes S., Riley P.F., von Lilienfeld O.A. // *J. Chem. Theory. Comput.* 2017. Vol. 13. P. 5255.
- CRC Handbook of Chemistry and Physics / W.M. Haynes, D.R. Lide and T.J. Bruno (Eds.), CRC Press: Boca Raton, FL, 2017. P. 9.
- HyperChem. // URL: <https://hyperchem.software.informer.com/>.
- Fink M., Schmiedekamp C.W., Gregory D. // *J. Chem. Phys.* 1979. Vol. 71. P. 5238.
- Petsko G.A., Ringe D. // URL: https://farmamol.web.uah.es/New_Science_Press/protein.html.
- Matsson P., Kihlberg J. // *J. Med. Chem.* 2017. Vol. 60. P. 1662.
- Paninski L. // *Neural Comput.* 2003. Vol. 15. P. 1191.
- IMSL. // URL: <https://help.imsl.com/fortran>.
- Mitra I., Saha A., Roy K. // *Mol. Simul.* 2010. Vol. 36. P. 1067.
- Rajaram R., Castellani B., Wilson A.N. // *Complexity*. 2017. Vol. 2017. P. 8715605.
- Lombardi O., Holik F., Vanni L. // *Synthese*. 2016. Vol. 193. P. 1983.
- Schroeder M.J. // *Entropy*. 2004. Vol. 6. P. 388.
- Aljanabi M.A., Hussain Z.M., Lu S.F. // *Math. Probl. Eng.* 2018. Vol. 2018. P. 9801308.

42. Zhuang L., Guan Y. // *Comput. Intell. Neurosci.* 2018. Vol. 2018. P. 3837275.
43. Zou Y., Zhang J., Upadhyay M., Sun S., Jiang T. // *IEEE Access.* 2020. Vol. 8. P. 171218.
44. Zmeskal O., Dzik P., Vesely M. // *Comput. Math. Appl.* 2013. Vol. 66. P. 135.
45. Sabirov D.S., Ōsawa E. // *J. Chem. Inf. Model.* 2015. Vol. 55. P. 1576.
46. Grigorev V.Yu., Grigoreva L.D. // *Mosc. Univ. Chem. Bull.* 2016. Vol. 71. P. 199.
47. Chen Y. // *Phys. A: Stat. Mech. Appl.* 2020. Vol. 553. P. 124627.
48. Mitić V.V., Lazović G.M., Manojlović J.Ž., Huang W.-C., Stojiljković M.M., Facht H., Vlahović B. // *Therm. Sci.* 2020. Vol. 24. P. 2203.
49. Sparavigna A.C. // *Philica.* 2016. hal-01377975.
50. Zhao Q., Burns S.E. // *Langmuir.* 2012. Vol. 28. P. 16393.
51. Heinz H., Vaia R.A., Krishnamoorti R., Farmer B.L. // *Chem. Mater.* 2007. Vol. 19. P. 59.
52. Ahmad M.B., Hoidy W.H., Ibrahim N.A., Al-Mulla E.A.J. // *J. Eng. Appl. Sci.* 2009. Vol. 4. P. 184.

Информация об авторах

Вениамин Юрьевич Григорьев – зав. лабораторией количественной связи структура – активность ИФАВ РАН, докт. хим. наук (beng@ipac.ac.ru);

Людмила Дмитриевна Григорьева – доцент факультета фундаментальной физико-химической инженерии МГУ имени М.В. Ломоносова, канд. физ.-матем. наук (ldg-msu@yandex.ru);

Олег Викторович Тиньков – доцент кафедры фармакологии и фармацевтической химии медицинского факультета Приднестровского государственного университета им. Т.Г. Шевченко, канд. хим. наук (oleg.tinkov.chem@mail.ru).

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Соблюдение этических стандартов

В данной работе отсутствуют исследования человека и животных.

Статья поступила в редакцию 07.06.2023;
одобрена после рецензирования 20.12.2023;
принята к публикации 25.03.2024.