

УДК 577.12:519.722

ЭНТРОПИЙНЫЙ ПОДХОД К ОЦЕНКЕ ХАОСА И ПОРЯДКА БИОЛОГИЧЕСКИХ СИСТЕМ

Симонян Г.С., Симонян А.Г.

Ереванский государственный университет, Ереван, e-mail: Sim-gev@mail.ru

Обсуждаются вопросы применения индекса разнообразия Шеннона в биологии, и, в частности, для оценки степени структурированности популяций и биоценозов. С помощью полученного из индекса Шеннона энтропийного индекса G проведен структурный анализ состояния биологических систем на уровне белков, рибонуклеиновой кислоты и клетки. Показано, что в ряду: белок \rightarrow рибонуклеиновая кислота \rightarrow клетка энтропия уменьшается, а геоэкологическая синтропия растет. Для белковых молекул G -функция стремится к единице, что свидетельствует о том, что структурная организация системы является равновесной. Установлено, что для рибонуклеиновой кислоты и клетки энтропийный индекс $G < 0,2$, что свидетельствует о высокой степени свободы и порядка этих систем.

Ключевые слова: белки, РНК, клетка, индекс Шеннона, энтропийный индекс, энтропия, геоэкологическая синтропия

ENTROPIC APPROACH FOR THE EVALUATION OF THE ORDER OF CHAOS AND BIOLOGICAL SYSTEMS

Simonian G.S., Simonian A.G.

Yerevan State University, Yerevan, e-mail: Sim-gev@mail.ru

The problems of application of Shannon diversity index in biology and particularly for evaluation of the degree of structuring of populations and ecological communities are discussed. The structural analysis of the state of biological systems on the protein, RNA and cells level are done by using entropic index G which is obtained from the Shannon index. It is shown that in the order protein \rightarrow RNT \rightarrow cell the entropy decreases and the geoecological sintropy increases. For protein molecules G function tends to unity, indicating that the structural organization of the system is an equilibrium. It is established that entropic index $G < 0,2$ for RNT and cells, which is indicating a high degree of freedom and the order of these systems.

Keywords: protein, RNT, cells, Shannon index, entropic index, entropy, geoecological sintropy

Концепция структурных уровней живых организмов позволяет расположить их в иерархическом соподчинении. Согласно критерию масштабности, биологический уровень включает подуровни: макромолекул (нуклеиновые кислоты, ДНК, РНК, белки), клеточный уровень, микроорганический (одноклеточные организмы), органов и тканей организмов в целом, популяционный, биоценозный и биосферный.

Живые организмы образуются из всевозможных малых органических молекул – мономеров, которые при объединении создают макромолекулы. Особую роль играют три класса мономеров: аминокислоты, нуклеотиды и моносахариды. Это те «кирпичики», из которых затем строятся полимерные макромолекулы – белки, нуклеиновые кислоты и полисахариды. Белки – это органические соединения, входящие в состав всех живых организмов, состоящие из большого числа мономеров. В состав белков входят 20 стандартных аминокислот. Свойства белков определяются пространственной трехмерной структурой их цепей. Изменяя в цепи лишь одну аминокислоту, можно получить молекулу с совершенно другой структурой и иными свойствами.

Огромное разнообразие живых организмов на нашей планете определяется различиями в составе и пространственной форме составляющих их белков. Нуклеотиды – это мономерные звенья цепи нуклеиновых кислот. Они представляют собой химические соединения остатков трех веществ: азотистого основания, углевода и фосфорной кислоты. Существует два вида нуклеиновых кислот – дезоксирибонуклеиновая кислота (ДНК) и рибонуклеиновая кислота (РНК). ДНК содержит генетическую информацию о последовательности аминокислот в полипептидных цепях и определяет структуру белков. РНК отвечает за создание белков. Оказалось, что молекула ДНК состоит из двух мономерных цепей, идущих в противоположных направлениях и спирально закрученных одна вокруг другой. ДНК всего органического мира образованы соединением четырех видов нуклеотидов – аденин (А), гуанин (Г), цитозин (С) и тимин (Т), их расположение в молекуле ДНК дает указание молекуле РНК, как надо строить белок. Молекула РНК одноцепочная. Структура РНК также создается чередованием четырех типов нуклеотидов, но их состав несколько отличается от состава таковых в молекулах

ДНК тем, что вместо тимина (Т) используется урацил (У) [1].

В биологических системах могут идти процессы как с возрастанием, так и уменьшением энтропии. Понятие энтропии имеет множество трактовок в самых разнообразных областях человеческих знаний. Наряду с энтропией Клаузиуса появилась статическая, информационная, математическая, лингвистическая, интеллектуальная и другие энтропии. Энтропия стала базисным понятием теории информации и стала выступать как мера неопределенности некоторой ситуации. Для характеристики меры сложности системы У. Эшби [11] впервые предложил использовать понятие энтропии. Система взаимодействует с внешним миром как единое целое. В целом же, система не теряет своей организованности или высокой упорядоченности. Открытые системы могут обмениваться с окружающими телами энергией, веществом и, что не менее важно, информацией. Чтобы биологическая система действовала и взаимодействовала со средой, она должна потреблять информацию из среды и сообщать информацию среде. Этот процесс называется информационным метаболизмом, который совместно с вещественным и материальным метаболизмом образует полный метаболизм. Впервые понятия энтропия и информация связал Шеннон [7]. С его подачи энтропия – это количество информации, приходящейся на одно элементарное сообщение источника, вырабатывающего статистически независимые сообщения. Получение какого-либо количества информации равно потерянной энтропии. Информационная энтропия для независимых случайных событий x с N возможными состояниями рассчитывается по формуле

$$H = -\sum_{i=1}^N p_i \log_2 p_i,$$

где P_i – вероятность частоты встречаемости некоторого события.

Впервые для оценки степени структурированности биоценозов Мак-Артур в 1955 г. использовал общее уравнение энтропии Шеннона [12], в котором, p_i заменил на $pi = n_i/N$; (где n_i – общее число особей вида i , N – общее число особей во всем биоценозе). В 1957 г. Р. Маргалеф постулировал теоретическую концепцию, согласно которой разнообразие соответствует энтропии при случайном выборе видов из сообщества [13]. В результате этих работ

большое распространение и повсеместное признание получил индекс Шеннона H , иногда называемый информационным индексом разнообразия K . Шеннона [3].

При расчете энтропии H по Шеннону считается, что каждая проба – случайная выборка из сообщества, а соотношение видов в пробе отражает их реальное соотношение в природе. В качестве оценок вероятностей независимых событий pi для формулы (1) могут быть использованы следующие апостериорные отношения: *удельная численность i -го вида*, как частное от деления его численности n_i на общую численность всех видов N , взятых для анализа ($pi = n_i/N$):

$$H = -\sum n_i/N \log_2 (n_i/N).$$

Использование индекса разнообразия Шеннона в экологии и, в частности, для оценки степени структурированности популяций и биоценозов очень детально обсуждается в работах [3, 8, 9]. Особенности функционирования биологических систем разного уровня с точки зрения теории информации были рассмотрены также в работе [10]. Шмальгаузенем были введены понятия о каналах прямой и обратной связи, по которым передается генетическая и фенотипическая информация, рассмотрены закономерности преобразования и кодирования биологической информации.

Целью данной работы является применение индекса Шеннона к биологическим системам, включающим подуровни: макромолекул (нуклеиновые кислоты, белки и РНК) и клеточный уровень.

В последние годы для комплексной оценки качества поверхностных вод сотрудниками кафедры экологической химии ЕГУ используется энтропийный индекс качества воды – G , который получается из индекса Шеннона [4]. G -функция характеризует гидроэкологические системы со стороны соотношения порядка и хаоса, мерами которых являются геоэкологическая синтропия – I [5] и энтропия Шеннона H , соответственно $G = H/I$. Значения G -функции говорят о том, что и в какой мере преобладает в структуре системы: хаос или порядок. Так, если $G < 1$, то в структуре системы преобладает порядок, в противном случае, когда $G > 1$, – хаос. При $G = 1$ хаос и порядок уравниваются друг друга, и структурная организация системы является равновесной.

Методика расчета. Для расчета значений I , H и G пользуемся следующим вычислительным алгоритмом:

1. Определяется число элементов в системе – n .

2. Далее оценивается общая сумма элементов: $N = \sum n$.

3. Потом вычисляются: $\log_2 N$, $n \log_2 n$ и $\sum n \log_2 n$.

4. Рассчитывается геоэкологическая синтропия [11] (I) и энтропия (H):

$$I = \sum n \log_2 n / N \text{ и } H = \log_2 N - I.$$

5. После чего определяется индекс качества воды: $G = H/I$.

В соответствии с целью работы и постановки задачи произведены расчеты функции белковых молекул, РНК, крахмала и живого тела.

Первичные структуры, белковые молекулы и РНК представляют собой, соответственно, последовательности аминокислотных остатков и нуклеотидов, и могут рассматриваться как системы, разделенные на части по одинаковым видам соответствующих аминокислот и нуклеотидов. Например, первичная структура молекулы глюкагона следующая:

His-Ser-Gln-Gly-Thr-Phe-Thr-Ser-Asp-Tyr-Ser-Lys-Tyr-Leu-Asp-Ser-Arg-Arg-Ala-Gln-Asp-Phe-Val-Gln-Trp-Leu-Met-Asn-Thr

Для глюкагона $N = 29$, число нуклеиновых кислот равно 15, а число одинаковых остатков (n) изменяется от 1 до 4. Соответствующие расчеты функций I , H и G приведены в табл. 1, откуда следует, что цепь глюкагона относится к хаотичному типу систем.

В табл. 2 даны значения I , H и G для белковых молекул.

Из табл. 1 и 2 видно, что аминокислотные цепи глюкагона и кортикотропина обладают явной хаотичной структурой ($G > 1$). Цепи же всех остальных белковых молекул относятся к синергетичному типу систем и проявляют тенденцию к уменьшению и приближению $G = 1$ по мере роста в белковой цепи общего числа аминокислотных остатков.

Приведем некоторые статистические данные по геному вируса HIVH3CG [6]. Расчеты приведены в табл. 3.

Так как для РНК $G < 0,2$, то в структуре системы преобладает порядок.

В клетках живых систем найдены многие химические элементы, присутствующие в окружающей среде, однако необходимы для жизни лишь около 20 из них. Эти элементы получили название биогенных.

Таблица 1

Расчет значений I , H и G в структуре молекулы глюкагона

| № п/п | Аминокислота | n | $n \log_2 n$ | № п/п | Аминокислота | n | $n \log_2 n$ |
|---|--------------|---|--------------|-------|--------------|---|--------------|
| 1 | Gly | 1 | 0 | 9 | Leu | 2 | 2 |
| 2 | Ala | 1 | 0 | 10 | Phe | 2 | 2 |
| 3 | Val | 1 | 0 | 11 | Tyr | 3 | 4,76 |
| 4 | Lys | 1 | 0 | 12 | Asp | 3 | 4,76 |
| 5 | His | 1 | 0 | 13 | Gln- | 3 | 4,76 |
| 6 | Met | 1 | 0 | 14 | Thr | 3 | 4,76 |
| 7 | Asn | 1 | 0 | 15 | Ser | 4 | 8 |
| 8 | Arg | 2 | 2 | | | | |
| $N = 29 \sum n \log_2 n = 33,1$ $I = 26:21 = 1,14 \quad H = \log_2 29 - 1,14 = 3,75$ $G = 3,75:1,14 = 3,29$ | | | | | | | |

Таблица 2

Значения N , I , H и G белковых молекул

| Белковые молекулы | N [12] | I | H | G |
|------------------------|----------|------|------|------|
| Глюкагон | 29 | 1,14 | 3,75 | 3,29 |
| Кортикотропин | 39 | 1,45 | 3,84 | 2,65 |
| Цитохром <i>c</i> | 104 | 2,75 | 3,95 | 1,74 |
| Миоглобин | 152 | 3,30 | 3,95 | 1,20 |
| Лютеинизирующий гормон | 205 | 3,57 | 4,11 | 1,15 |
| Тромбин | 259 | 3,81 | 4,21 | 1,10 |
| Фосфоорилаза гликогена | 841 | 5,54 | 4,18 | 0,75 |

Таблица 3

Расчет I, H и G генома вируса (HIVH3CG)

| Нуклеотиды | n | nlog ₂ n |
|--|------|---------------------|
| U | 2168 | 24026,1 |
| C | 1785 | 19281,1 |
| A | 3432 | 40308,3 |
| G | 2364 | 26493,4 |
| $I = \sum n \log_2 n : N = 110108,9 : 9749 = 11,29$ $H = \log_2 N - I = \log_2 9749 - 11,29 = 1,95$ $G = H/I = 1,95 : 11,29 = 0,173$ | | |

Таблица 4

Расчет I, H и G клетки

| Элементы | n | nlog ₂ n |
|--|----|---------------------|
| C | 18 | 75,0 |
| H | 10 | 33,2 |
| O | 70 | 428,8 |
| N | 2 | 2 |
| $I = \sum n \log_2 n : N = 539 : 100 = 5,39$ $H = \log_2 N - I = \log_2 100 - 5,39 = 1,25$ $G = H/I = 1,25 : 5,39 = 0,232$ | | |

В среднем около 70% массы организмов составляет кислород, 18% – углерод, 10% – водород, и 2% – остальные элементы (азот, фосфор, калий, кальций, сера и т.д.) [10].

Из табл. 4 следует, что в структуре клетки также преобладает порядок.

Таким образом, впервые с помощью энтропийного индекса Шеннона проведен структурный анализ состояния биологических систем на уровне белков, рибонуклеиновой кислоты, крахмала и клетки. Установлено, что в ряду: белок → РНК → клетка энтропия уменьшается, а геоэкологическая синтропия растет. Для генома вируса HIVH3CG и клетки $G < 0,2$, что свидетельствует о высокой степени свободы и порядка этих систем. Для белка G-функция стремится к единице, что свидетельствует о том, что структурная организация системы является равновесной.

Список литературы

1. Иванова Т.В., Калинова Г.С., Мягкова А.Н. Общая биология. – М., Просвещение, 2000.
 2. Николаев А.Я. Биологическая химия. – М., МИА, 2001. – 496 с.

3. Розенберг Г.С. Информационный индекс и разнообразие: Больцман, Котельников, Шеннон, Уивер...// Самарская Лука: проблемы региональной и глобальной экологии. – 2010. – Т. 19, № 2. – С. 4–25.

4. Симонян А.Г., Пирумян Г.П. Энтропийный подход к оценке экологического состояния реки. Геология морей и океанов: Материалы XXI Международной научной конференции (Школы) по морской геологии. – М: ГЕОС, 2015. – Т. 4.

5. Симонян Г.С. Оценка состояния гидроэкологических систем в свете синергической теории информации. Материалы Всероссийской научно-практической конференции. Экологическая безопасность и природопользование: наука, инновации, управление. – Махачкала, АЛЕФ, 2013. – С. 275–280.

6. Таланов В.А. Математические модели синтеза пептидных цепей и методы теории графов в расшифровке генетических текстов. – <http://new.math.msu.su/departament/dm/dmnc/PUBL1/gen.htm>.

7. Шеннон К. Работы по теории информации и кибернетике. – М., ИЛ, – 1963. – 830 с.

8. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы системной идентификации. – Тольятти, ИЭВБ РАН, 2003. – 463 с.

9. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы, критерии, решения: в 2-х кн. – М., Наука, 2005. – Кн. 1. – 281 с.

10. Шмальгаузен И.И. Кибернетические вопросы биологии. – Новосибирск: Наука, 1968. – 224 с.

11. Эшби У. Введение в кибернетику. – М., ИЛ, – 1959. – 432 с.

12. MacArthur R.M. Fluctuation of animal populations and measure of community stability// Ecology. – 1955. – V.36, № 3. – P.533–536.

13. Margalef R. Information theory in ecology// Gen. Syst. – 1958. – V. 3. – P. 36–71.