

Анализ и моделирование биологических цифровых информационных систем

Ратушняк А.С.*, Проскура А.Л., Шевырина В.А.

Федеральный исследовательский центр информационных и вычислительных технологий,

Новосибирск, Россия

** ratushniak.alex@gmail.com*

Ключевые слова: моделирование возникновения живых систем; энтропия; негэнтропия; молекулярные машины

Мотивация и цель: Для продвижения работ в области бионейроинформатики, для разработки методов и подходов к профилактике, диагностике и коррекции нейроиндуцированных патологий необходимо понимание базовых физических функций живых систем. Для этого необходимы представления о физической сущности и механизмах возникновения и эволюции негэнтропийных молекулярных комплексов. Работа посвящена, прежде всего, анализу существующих методов, поиску подходов к моделированию возникновения, эволюции, созданию моделей первичных протобиологических молекулярных комплексов (protobionts) [<https://en.wiktionary.org/wiki/protobiont>]. Анализ фундаментальных основ жизни (поведения) возможен только на основе моделирования простейшей живой системы [1, 2].

Методы и алгоритмы: Сейчас одним из ведущих постулатов считается, что основной функцией живых систем является возможность саморепликации. При этом существует понимание, что вероятность спонтанного возникновения самореплицирующейся системы беспрельдно низка. А без репликации (с некоторыми ошибками в пределах порога Эйгена) эволюционный процесс, в рамках таких представлений, практически мало реален.

В рамках другого подхода к возникновению protobionts существуют представления о том, что базовой функцией живых систем является поддержание собственной устойчивости в среде после спонтанного формирования достаточно простых молекулярных конструкций. Такие конструкции, состоящие из молекул, способных выполнять логические операции, могли бы на основе информационных процессов прогнозировать состояние окружающей среды (опережающее отражение [3]) и избегать понижающих устойчивость воздействий и/или приобретать дополнительную устойчивость. Додарвиновская эволюция при этом может состоять в отборе средой наиболее устойчивых из спонтанно возникающих protobionts.

Для определения реальных путей возникновения и функционирования биосистем необходимо создание их имитационных моделей. Это облегчает анализ результатов и позволяет решать одну из важнейших задач – определение базовых функций биосистем. Однако при попытках создания моделей даже самых простейших из существующих биообъектов возникают значительные проблемы. Одной из ведущих проблем является «проклятие размерности», т. е. многокомпонентность модели. Не менее значима оценка параметров и интеграция данных. Особенно при попытках создания моделей внутриклеточных информационных сетей. Отсутствие исчерпывающих и количественных данных и потребность в вычислитель-

ных ресурсах еще больше усложняют процесс моделирования. При существующем подходе не удастся создать достаточно полную модель даже одной клетки, состоящей из примерно 10^{15} атомов, при этом реакции, к примеру, сворачивания белка и ферментативные реакции происходят за пико- и фемтосекунды. Самые крупные из существующих модельных систем, доступных для используемых методов, состоят из 10^{10} атомов [2, 4], при этом временные диапазоны, в которых возможно изучение таких систем с помощью суперкомпьютеров, ограничены наносекундами. Чтобы смоделировать клетку целиком, нужна модель в 10^5 раз больше, которая, соответственно, будет работать в 10^6 раз дольше. Нужная компьютерная модель должна быть как минимум в 10^{11} раз масштабнее, чем те, что у нас есть на сегодняшний день. При существующем подходе понятно, что набор свойств объекта моделирования на порядки больше, чем свойств модели. В обозримом будущем радикального развития вычислительных мощностей не ожидается. Поэтому для решения огромного множества вопросов необходим поиск новых подходов и концепций. Наше исследование сосредоточено на решении этих проблем – выборе методов и алгоритмов моделирования биологических цифровых информационных систем.

Результаты: В данной работе мы попытались создать модель наипростейших молекулярных конструкций (агентов – негэнтропийных молекулярных комплексов), отвечающих представлениям о живых системах как об устройствах, обладающих активной устойчивостью в среде. При этом задавались параметры среды и агента. «Правила» для внешней среды: Уровень «энтропии». Наличие достаточного количества и разнообразия молекулярных компонентов для спонтанного формирования агентов, с определенной вероятностью. Наличие значимых и ассоциированных (или причинно-следственно), связанных со значимыми, подпороговых информационных сигналов (разной модальности, например, свет, звук и др.). При этом задается соотношение количества значимых и подпороговых сигналов. В среде (исходно) количество пар значимых и подпороговых сигналов может быть равно или больше, чем рецепторов у агента.

«Правила» для агентов: Агенты возникают в среде спонтанно. Задаются следующие параметры: nFr – рецепторы значимых сигналов (n – к-во); nfr – рецепторы ассоциированных сигналов (n – к-во); nE – эффекторы (n – к-во); M – объем ассоциативной памяти совпадения значимых и подпороговых сигналов; nL – запас энергии (n – к-во жизней) ($n > 1$); x – коэффициент вычитания из L при срабатывании fr ; y – коэффициент количества повторений f , не сочетанного с F , до удаления из памяти $F + f = E$; z – максимум nL ; v – коэффициент увеличения nL при достижении L значения z . При каждом возникновении агентов значение коэффициентов обновляется (или сохраняется – наследуется).

Предельно простые супрамолекулярные конструкции флуктуационно формируются в природе, вероятно, с некоторым «запасом», избытком упорядоченности – негэнтропии N . Взаимодействие с процессами, происходящими в окружающей энтропийной среде, приводило к уменьшению этой упорядоченности, суммарной негэнтропии в системе. Приобретенная при этом информация в дальнейшем «используется» молекулярной системой для прогнозирования подобных событий на время Δt . Благодаря этому происходит уменьшение энтропии с некоторым «информационным коэффициентом» K_i , зависящим от объема полученной информа-

ции и определяющим глубину возможного прогнозирования. Разработаны алгоритмы функционирования и программные коды моделей негэнтропийных биоагентов.

Выводы: Опираясь на модель биоагента (protobionts), основанную на таких физических представлениях, и учитывая преемственность, эмерджентность организации биологических систем, в дальнейшем представляется возможным попытаться моделировать их эволюцию наращивать сложность до пределов аппаратных ограничений.

Финансирование: Работа выполнена в рамках государственного задания Минобрнауки России для Федерального исследовательского центра информационных и вычислительных технологий.

Analysis and modeling of biological digital information systems

Ratushnyak A.S.*, Proskura A.L., Shevyrina V.A.

Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia

* *ratushniak.alex@gmail.com*

Ключевые слова: modelling emergence of living systems; entropy; negentropy; molecular machines

Motivation and Aim: To advance work in the field of bioneuroinformatics, to develop methods and approaches to the prevention, diagnosis and correction of neuro-induced pathologies, it is necessary to understand the basic physical functions of living systems. This requires ideas about the physical essence and mechanisms of the emergence and evolution of negentropic molecular complexes. The work is devoted, first of all, to the analysis of existing methods, the search for approaches to modeling the emergence, evolution, and creation of models of primary protobiological molecular complexes (protobionts) [<https://en.wiktionary.org/wiki/protobiont>]. Analysis of the fundamental principles of life (behavior) is possible only on the basis of modeling the simplest living system [1, 2].

Methods and Algorithms: Now one of the leading postulates is that the main function of living systems is the ability to self-replicate. At the same time, there is an understanding that the probability of spontaneous emergence of a self-replicating system is infinitely low. And without replication (with some errors within the Eigen threshold), the evolutionary process, within the framework of such ideas, is practically unrealistic.

Another approach to the emergence of protobionts is the idea that the basic function of living systems is to maintain their own stability in the environment after the spontaneous formation of fairly simple molecular structures. Such structures, consisting of molecules capable of performing logical operations, could be able, based on information processes, to predict the state of the environment (anticipatory reflection [3]) and avoid impacts that reduce stability and/or acquire additional stability. Before Darwinian evolution, it may consist in the selection by the environment of the most stable of spontaneously arising protobionts. To determine the real ways in which biosystems arise and function, it is necessary to create their simulation models. This facilitates the analysis of the results and allows us to solve one of the most important tasks – determining the basic functions of biosystems. However, when trying to create models of even the simplest existing biological objects, significant problems arise. One of the leading problems is the “curse of dimensionality”, i. e. multicomponent model. Equally important is parameter estimation

and data integration. Especially when trying to create models of intracellular information networks. The lack of comprehensive and quantitative data and, at the same time, the need for computing resources further complicate the modeling process. With the existing approach, it is not possible to create a sufficiently complete model of even one cell consisting of approximately 10^{15} atoms, while reactions, for example, protein folding and enzymatic reactions, occur in pico and femto seconds. The largest model systems existing today, available for calculation methods, consist of 10^{10} atoms [2, 4], while the time ranges in which it is possible to study such systems using supercomputers are limited to nanoseconds. To simulate an entire cell, you need a model 10^5 times larger, and which, accordingly, will work 10^6 times longer. The required computer model would have to be at least 10^{11} times larger than what we have today. With the existing approach, it is clear that the set of properties of the modeling object is orders of magnitude larger than the properties of the model. Radical developments in computing power are not expected in the near future. Therefore, to solve a huge variety of issues, it is necessary to search for new approaches and concepts. Our research focuses on addressing these issues. Selection of methods and algorithms for modeling biological digital information systems.

Results: In this work, we tried to create a model of the simplest molecular structures (agents – negentropic molecular complexes), corresponding to the concept of living systems as devices with active stability in the environment. At the same time, the parameters of the environment and the agent were set.

“Rules” for the external environment: Level of “entropy”. The presence of a sufficient number and variety of molecular components for the spontaneous formation of agents, with a certain probability. The presence of significant and associated (or causally) related to significant, subthreshold information signals (of different modalities, for example light, sound, etc.). In this case, the ratio of the number of significant and subthreshold signals is set. In the environment (initially), the number of pairs of significant and subthreshold signals may be equal to or greater than the number of receptors on the agent.

“Rules” for agents: Agents arise spontaneously in the environment. The following parameters are set: nFr – receptors of significant signals (n – number); nfr – receptors of associated signals (n – number); nE – effectors (n – number); M is the volume of associative memory for the coincidence of significant and subthreshold signals; nL – energy reserve (number of lives) ($n > 1$); x – subtraction coefficient from L when fr is triggered; y – coefficient of the number of repetitions of f not combined with F before deleting $F+f=E$ from memory; z – maximum nL ; v is the coefficient of increase in the quantity F and f when L reaches the value z . With each occurrence of agents, the value of the coefficients is updated (or saved – inherited)

Extremely simple supramolecular structures are formed fluctuation ally in nature, probably with some “reserve”, an excess of order – negentropy N . Interaction with processes occurring in the surrounding entropic environment led to a decrease in this order, the total negentropy in the system. The information acquired in this case is subsequently “used” by the molecular system to predict similar events at time Δt . Due to this, entropy decreases with a certain “information coefficient” K_i , which depends on the amount of information received and determines the depth of possible prediction. Functioning algorithms and program codes for models of negentropy bioagents have been developed.

Conclusion: Based on the model of bioagents (protobionts) based on such physical concepts and taking into account the continuity and emergence of the organization of biological systems, in the future it seems possible to try to simulate their evolution and increase complexity to the limits of hardware limitations.

Funding: The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation for Federal Research Center for Information and Computational Technologies.

Список литературы/References

1. Thornburg Z.R., Bianchi D.M., Brier T.A. et al. Fundamental behaviors emerge from simulations of a living minimal cell. *Cell*. 2022;185(2):345-360.e28. doi 10.1016/j.cell.2021.12.025
2. Singharoy A., Maffeo C., Delgado-Magnero K.H. et al. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell*. 2019;179(5):1098-1111.e23. doi 10.1016/j.cell.2019.10.021
3. Anokhin P.K. Anticipatory reflection of reality. *Problems of Philosophy*. 1962;7:97-106 (in Russian)
4. Karr J.R., Sanghvi J.C., Macklin D.N. et al. A whole-cell computational model predicts phenotype from genotype. *Cell*. 2012;150(2):389-401. doi 10.1016/j.cell.2012.05.044