ГЕНЕЗИС НОСИТЕЛЕЙ КОДА И ДИВЕРГЕНЦИЯ ТЕМПЛАТНЫХ СЕМАНТИД ОТ ЭПИСЕМАНТИД

> Выведение жизни из одного простого исходного пункта многим кажется сейчас привлекательным, и даже биохимики так увлечены представлением о роли нуклеиновой кислоты в синтезе белков, что рисуют столь же нелогичную картину, которую можно было бы изложить следующими словами: «Некогда на берегах первичного океана существовала небольшая молекула дезоксирибонуклеиновой кислоты (ДНК); она начала более или менее правильно удваиваться и вызвала образование белка, который способствовал синтезу новых молекул ДНК, а из этих молекул в результате изменчивости возникли организмы». Однако биохимики в своем энтузиазме склонны забывать о том, что первая молекула ДНК сама является уже очень сложным образованием.
>
> ***Дж. Д. Бернал***[1]

Введение. Постановка проблемы.

Проблема происхождения транскрибируемого генетического кода и возникновения наследственности на его основе является неотъемлемой частью современной теории абиогенеза и ключом к раскрытию механизмов эволюции протобиологических систем.

В силу множественности факторов, взаимодействовавших в ходе образования первичных носителей кода, очевидно, что решение данной проблемы невозможно без привлечения сторонних дисциплин. Конкретность условий, необходимых и достаточных для возникновения носителей кода, указывает на необходимость привлечения геохимии, в частности, геохимии углерода, а также палеоклиматологии. В более общем случае, связанном с нахождением прекурсоров носителей в метеоритном материале, необходима апелляция к данным космохимии. Являясь супрамолекулярными

---

[1] Цит. по: Бернал Дж.Д. Молекулярная структура, биохимическая функция и эволюция, в сб.: Теоретическая и математическая биология. М., Мир, 1968, с. 113.

структурами, носители кода любого уровня элементарности не могут быть описаны без привлечения супрамолекулярной химии, действующей на принципах самосборки, и, как следствие – нелинейной физики и неравновесной термодинамики, которые описывают процессы самоорганизации в произвольных системах. Являясь, в той или иной форме, носителями информации, элементарные молекулярные коды не могут быть описаны без привлечения теории информации и теории групп, а также статистической механики[2]. Порождая процессы молекулярной эволюции, экспрессируемые коды адекватно описываются в динамике только при использовании эволюционной кибернетики. Последняя, в свою очередь, приводит к молекулярной логике и теории молекулярных вычислений, так как возникает возможность сопоставления операций, выполнявшихся в кодах раннего и современного строения. Эту цепь причинно-следственных связей можно продолжить (что и будет сделано в настоящей работе), но уже из того, что изложено, следует, что проблема возникновения кода не может быть решена средствами молекулярной биологии как таковой без привлечения методов и концептов весьма отдалённых дисциплин, зачастую отсутствующих в поле компетенции системной биологии. Таким образом, в дальнейшем уже нельзя ограничиваться рамками внутренней непротиворечивости концептов происхождения кода с позиций молекулярной генетики, так как любая теория или концепт, претендующие на комплексный подход к проблеме, обязаны, по определению, отвечать на фундаментальные запросы смежных дисциплин.

В общем случае, говоря о коде и генетических системах, подразумевают соответственно носитель зашифрованной (т.е. несуществующей без

[2] В настоящее время подобный подход к происхождению кода превращается в молекулярно-биологический и биофизический тренд. Для примера можно привести статью Tlusty T. A colorful origin of the genetic code: Information theory, statistical mechanics and the emergence of molecular codes. *Physics of Life Reviews*, Vol. 7, No. 3, pp. 362-376 (2010) и её обсуждение Tagaki H., Kaneko K. Topological, statistical, and dynamical origins of genetic code. Ibid, pp. 379-380.

системы декодирования) информации и системы, осуществляющие её декодирование в форме транскрипции и трансляции. Как правило, существует ещё источник «шума», несущий эволюционный смысл и представленный петлями обратной связи эпигенетического, репарационного и криптогенного (например – RNA editing эукариот) назначения [1]. С позиций теории управляющих систем, указанные функции могут быть рассмотрены в рамках схемы, состоящей из памяти как совокупности дискретных значений, емкость которой исчисляется в килобазах (kbase), и полюсов, на которые подаются и с которых снимаются сигналы взаимодействующих элементов [2]. В таком случае, с формальной точки зрения, для сопоставления различных систем передачи генетической информации логично использовать метод ER-диаграмм[3], используемый в теории управляющих систем. Тогда сравнение современных носителей кода и их предшественников может проводиться по аналогии с концепцией QSAR[4] или QSPR [3] с использованием диаграмм сущность-связь (ERD) для отображения взаимоотно-

[3] Метод ER-диаграмм (ERD, от англ. entity-relationship diagram – диаграммы сущность-связь) – графический способ нотации моделей данных с помощью блоков и векторов связи, описывающих объекты и отношения между ними. Наиболее удобен в канонической нотации П.Чена (Chen P. The entity-relationship model—toward a unified view of data, ACM Transactions on Database Systems, Vol.1, Issue 1, pp. 9-36 (1976)), приближенной к принятому в РФ понятию «блок-схема». Значительно удобнее стандартных диаграмм Насси-Шнейдермана (Nassi I., Shneiderman B. Flowchart Techniques for Structured Programming, SIGPLAN Notices, Vol. 8, Issue 8, pp. 12-26 (1973)), поскольку не накладывает запрет на операции безусловного перехода, что корректно для генетических систем и их абиогенных предшественников.

[4] QSAR – (от англ. quantitative structure-activity relationship – количественная корреляция между структурой и активностью) – концепт установления структурно-функциональных корреляций с помощью присваивания молекулярным структурам некоторой структуры определённых функциональных дескрипторов, связанных с топологическими, электронно-плотностными и иными характеристиками, и последующего моделирования с выявлением функциональных аналогов вещества по его структуре. Специальным случаем QSAR является концепт QSPR (quantitative structure-property relationship – количественная корреляция между структурой и свойствами), в рамках которого возможно применение ERD как способа визуализации и анализа корреляции структура-свойство через нотацию отношений сущность-связь, в том числе – для нестатических структур

шений структуры и функциональных характеристик (SPR) потенциальных носителей кода, что расширяет границы выборки последних.

В границах терминологии Полинга кодирующие полимеры, вне зависимости от состава, обозначаются как «семантиды», в то время как метаболиты относятся к «эписемантидам», рассматривающимся как внешние сигналы [4]. Таким образом, задача происхождения генетического кода может быть интерпретирована как поиск вероятных протобиологических семантид, являвшихся предшественниками современных носителей транскрибируемого кода, и дифференциация классов семантид от классов эписемантид для ранних абиогенетических и филогенетических этапов. Это должно быть сопряжено с выявлением эквивалентных или функционально идентичных диаграмм в различных классах.

Необходимым и достаточным условием для выделения нетождественных классов является, во-первых, выполнение некоторым классом присваиваемой ему функции, а во-вторых, невыполнение им функций, тождественных специфичным для других классов или их надклассов. В связи с этим, построение настоящего материала вынуждено подчинено одновременно и дедуктивной, и индуктивной логике, допускающей верификацию современных представлений о классах семантид и эписемантид как с помощью доказательства существования функциональных межклассовых эквивалентов, так и путём редукции классов, выполняющих эквивалентные операции. Функциональная минимизация является способом реконструкции примитивных самовоспроизводящихся протобиологических систем, поэтому, учитывая недостаточную стерическую дифференциацию эписемантических и семантических молекул на ранних этапах эволюции, логично рассматривать верификацию классов указанных объектов как способ установления путей их функциональной дивергенции. Соответственно, надо иметь критерии идентификации классов для построения филогенетических деревьев молекулярной эволюции.

В соответствии с литературными данными, широко используемыми, но не комплексными критериями функциональной идентификации современных семантид являются:

I. стереохимическая и электронная специфичность комплементарного связывания [5], описываемая, в частности, в терминологии QSAR с помощью топологических (подструктурных) дескрипторов типа MCI (индексов молекулярной связи) и квантовых параметров,

II. матричная функция – способность к построению комплементарных последовательностей на базе селективного связывания с различными субстратами [6], лежащая, в частности, в основе транскрипции и трансляции, в том числе – девиационной [7],

III. самовоспроизведение или репликация как частный случай реализации матричной функции, а также матрично-опосредованная дупликация [8] и мультипликация [9], имеющие в основе гомологичные структурно-химические механизмы,

IV. геометрическая и топологическая структура [10], благоприятствующая корректной транскрипции [11] и репликации [12], и консеквентная целому ряду биологических эффектов [13], влияющих на экспрессию,

V. способность к самовосстановлению – репарации [14], базирующейся на клеточной машинерии[5] ферментативного распознавания [15] и супра-молекулярной химии комплементарной фиксации.

Не подвергая сомнению общепризнанный и повсеместно используемый подход инивидуального рассмотрения каждого из указанных критериев, тем не менее, следует указать на то, что, с позиций физической химии, в основе всех указанных атрибутов носителя лежит один и тот же принцип – принцип координационной (комплементарной) фиксации рецептора и

[5] Термин, калька с английского, внедрен в российскую терминологию А. Финкельштейном и О. Птицыным (Финкельштейн А.В., Птицын О.Б. Физика белка. М., КДУ, 2005, 456 с.)

субстрата [16]. Это значит, что существует множество потениально комплементарных друг другу структур различного состава, зачастую отсутствующих в современных генетических системах, но имевших место в ходе образования кодирующих молекулярных структур на ранних этапах протобиологической эволюции. С точки зрения молекулярной логики живого состояния (*molecular logic of the living state*), способы воспроизведения генетических функций с сохранением биоорганического химизма среды и кодирующей матрицы и с изофункциональным или изоструктурным замещением имеют равное право на существование, поскольку соблюдают единые предикативные отношения и описываются единым классом самовоспроизводящихся клеточных автоматов [17, 18]. Это хорошо соотносится с функциональной геномикой, рассматривающей проблемы наследственности с позиций биоинформатики, не опирающейся на элеменный состав носителя при анализе комплементарных взаимодействий в чиповых структурах, что позволяет использовать синтетические основания и замещённые формы в ходе молекулярно-генетических экспериментов[6] [19].

В таком случае, в контексте происхождения матричного кодирования целесообразно рассматривать любые допустимые, с палеогеохимических позиций, варианты темплатов. В настоящей работе дается литературный анализ по проблеме потенциальных темплатных семантид, относящихся к различным структурно-химическим группам. Рассматриваются также «органоминеральные» механизмы образования носителей с использованием абиогенных неорганических поверхностей в качестве первичного темпла-

[6] Учитывая аддитивный характер фореза и, как следствие, автоматического секвенирования, очевидно, что в ходе интерпретации встаёт ряд проблем, решаемых статистически только принятием *a priori* гомогенности химизма выборки. В результате возникает необходимость в алгоритмах выравнивания последовательностей, недостаточно подобных друг другу, приближенного полуглобального сравнения со штрафами за пропуски, эвристического выбора наиболее важных ячеек из общего массива данных, перенормированного консенсуса и т.д. Более подробно с этим можно ознакомиться в соответствующей литературе. На русском языке см., напр.: Сетубал Ж., Мейданис Ж. Введение в вычислительную молекулярную биологию. М., «Регулярная и хаотическая динамика», 2007, 420 с.

та. В доказательстве некоторых неочевидных положений по необходимости используется математический формализм, указывающий на независимость функциональных моделей от квантовой химии и элементного состава носителя. Это позволяет сопрягать проблему происхождения кода с эволюционной кибернетикой как для протобиологических, так и для предбиологических структур. Данная работа не претендует на исчерпывающе полный обзор существующих трендов рассматриваемого направления, являясь лишь совокупностью литературно-теоретических данных, необходимых для доказательства основных тезисов, иллюстрирующих возможные пути происхождения темплатного/матричного кодирования, репликации и трансляции.

ПРОБЛЕМА ПРОИСХОЖДЕНИЯ НУКЛЕИНОВЫХ КИСЛОТ И НУКЛЕИНОВОГО КОДА. ФУНДАМЕНТАЛЬНЫЕ ВОПРОСЫ ВЕРИФИКАЦИИ ПРОТИВОРЕЧИВЫХ КОНЦЕПЦИЙ.

В первом приближении проблема происхождения генетического кода, что очевидно, сводится к проблеме протобиологического синтеза оснований нуклеиновых кислот под действием абиотических факторов среды. В связи с этим имеет место комплекс вопросов, без ответа на которые любое решение проблемы является бессмысленным, а именно:

1. Тождественны ли современные компоненты носителей кода синтезировавшимся в эоархейских условиях ≈3.6 млрд. лет назад или исходно имели место элементарные прекурсоры, замещенные впоследствии более специфическими семантидами?
2. Возникновение носителей кода происходило в условиях Земли или в космической среде? От решения этого вопроса зависят принимаемые *ab initio* свойства моделей его воспроизведения в первичных условиях.

3. Возникновение современных семантид происходило в протоклетке или автономно от неё? Если автономно, то – каким образом носитель кода был интеркалирован в протоклетку и сумел специфично функционировать в новом для него окружении; если эндогенно, то – какие механизмы привели к возникновению подобных систем?
4. Соответствуют ли современные принципы лабораторного синтеза носителей кода статистически вероятным условиям абиогенеза и синтеза (или биосинтеза) в имевшихся в протобионтах на тот период времени супрамолекулярных ансамблях эписемантид?
5. Облигатно ли все первичные семантиды обладали комплексом свойств современных носителей кода или же для современных носителей существенны факультативные механизмы, являющиеся продуктом эволюции клеточного окружения?
6. Единственны ли нативные нуклеиновые кислоты как носители кода?
7. Какова минимальная совокупность взаимодействующих подсистем, при которой код может быть транскрибируем и в которой целесообразна его экспрессия?

Иными словами, так как свойства детерминированного кода зависят от его происхождения и носят эмерждентный характер по отношению к среде, формировавшей его материальный носитель, очевидно, что без ответов на вопросы об условиях его возникновения принципиально невозможно верифицировать современные представления о механизмах его функционирования и эволюции на первичных этапах молекулярной эволюции. Далее приводится аналитический обзор литературы, позволяющий сформулировать гипотезу, не противоречащую экспериментальным данным смежных дисциплин и не требующую для собственного существования игнорирования каких-либо из имеющихся проблем генезиса кода в естественных условиях.

§ 1.1. Тождественны ли современные нуклеиновые кислоты первичным носителям кода?

Согласно распространенной модели «мир РНК» [20] первыми репликаторами были РНК, катализирующие синтез собственных копий[7]. Начиная с 1961 г., когда на установке для моделирования предбиологического органического синтеза конструкции Миллера [21] при добавлении к стандартной смеси синильной кислоты был получен аденин [22], возник вопрос о возможности синтеза полинуклеотидов в абиогенетических условиях. Исходно в продуктах искрового разряда удалось обнаружить продукты конденсации циановодорода HCN – аденин и гуанин, а также производные цианоацетилена HC=C–C=N – цитозин и урацил (тимин обнаружен не был) [23], которые были опознаны стандартными методами биохимии (без структурного анализа), что утвердило позиции сторонников гипотезы. В то же время, несмотря на обнаружение в подобных условиях также моносахаридов и рибозы [24], помимо того, что это не давало ответа на вопрос о возникновении РНК как таковой, оказалось, что многие химические структуры могли быть упущены в силу методического несовершенства экспериментальной идентификации. В современной химической аналитике для идентификации сложных соединений используются комплексный подход к сбору и интерпретации данных, базирующийся на хемометрических принципах, дающих возможность достоверно идентифицировать структуру определяемого вещества. Так, для качественной идентификации органических соединений неизбежно интегрированное использование масс-спектрометрии, хроматографии, инфракрасной спектрометрии и Фурье-

[7] В дальнейшем изложении подобное самокопирование кода, базирующееся на изоморфизме группировок темплата и субстрата, зачастую называется автоморфизмом. Воспроизведение последовательностей кода на базе отличных группировок или носителей рассматривается как изоморфизм, обусловленный биективными (взаимнооднозначными) отображениями матрицы на субстрат. При этом, поскольку биекция реализуется при композиции инъекции и сюръекции, возможны комбинации нечеткого соответствия, обусловливаемые неполными отображениями кода в ходе сюръективных и инъективных операций.

ИК-спектрометрии, а также, в специальных случаях, методов ядерно-магнитного и электрон-парамагнитного резонанса [25]. Идентификация в опытах по абиогенному синтезу велась методами бумажной хроматографии, спектрофотометрии и радиоавтографии. При этом не было проведено ни одного инструментального структурного анализа. В то же время, принимая во внимание возможность получения экзотических продуктов в столь нетривиальных условиях эксперимента, идентификация компонент по подвижности и аддитивному спектру является, как минимум, недостаточной. Поэтому, данные опытов [22, 24] при всем уважении к их авторам и значимости результата нельзя считать аналитически исчерпывающими.

Требуется качественное повторение опытов Миллера-Юри, Оргела и Оро на новых установках с параллельным детектированием образующихся химических соединений и их структурным анализом. Последний необходимо дополнить методами квантовой химии и молекулярного моделирования, позволяющими расшифровать нестандартные структуры и варианты конформаций. Известно, что применение стандартных методов протеомики при исследовании образцов вещества, полученных на установке Миллера, при идентификации с использованием стандартных библиотек уже дало возможность обнаружить в них не 4, а 22 типа аминокислот [26]. В связи с неочевидностью ограничения вероятных абиогенных продуктов номенклатурой современных биоорганических соединений и, тем более, тех, которые были выявлены архаическими методами полвека назад, возникла потребность в расширенном пересмотре исходных данных разрядных экспериментов [27]. В недавнем времени появилась гипотеза, допускающая синтез различных нуклеиновых кислот и их предшественников в атмосфере восстановительного состава и построение биохимических систем на базе продуктов абиогенного синтеза – т.н. гипотеза HCN world[8] [28].

[8] Подобные идеи в контексте абиогенеза ранее высказывались Пфлюгером в конце XIX века: согласно этой гипотезе, возникновение циановых соединений углерода в условиях

Действительно, если исходить из функциональных предпосылок (см. выше), то «мир РНК» является частным случаем совокупной реализации репликативной, каталитической и структурно-пластической функций, способных реализовываться и на других носителях, подобных РНК. Проиллюстрируем это. Согласно А.С.Спирину, функциональные посылы «мира РНК» – совокупно реализуемые РНК процессы, могут быть сведены к комплексу взаимосвязанных функций, уникальному для РНК – рибозимов; это [29][9]:

I. «Генетическая репликативная функция: структурная возможность копирования (репликации) линейных последовательностей нуклеотидов через комплементарные последовательности.
II. Кодирующая функция: программирование белкового синтеза линейными последовательностями нуклеотидов.
III. Структурообразующая функция: формирование уникальных трехмерных структур. Компактно свернутые молекулы малых РНК принципиально подобны трехмерным структурам глобулярных белков, а более длинные молекулы РНК могут образовывать и более крупные биологические частицы или их ядра.
IV. Функция узнавания: высокоспецифические пространственные взаимодействия с другими макромолекулами (в том числе белками и другими РНК) и с малыми лигандами. Эта функция, пожалуй, глав-

---

раскаленной планеты приводило затем к возникновению белков при взаимодействии циана с образовавшейся аодой при остывании планеты, после чего становился возможен биогенез по Геккелевскому типу. Экспериментально конденсация низкомолекулярных единиц в более крупные в водных растворах, содержащих HCN, в контексте возникновения жизни изучалась Кальвином – см. Calvin M. Chemical Evolution, *Proc. Roy. Soc., Ser. A*, Vol. 288, pp. 441-466 (1965).

[9] Аналогичные идеи изложены им в: Spirin A.S. The RNA World and Its Evolution, *Molecular Biology*, Vol. 39, No. 4, pp. 466-472 (2005); Spirin A.S. Ancient RNA World, *Paleontological Journal*, Vol. 44, No. 7, pp. 737-746 (2010); Spirin A.S. When, Where and in What Environment Could the RNA World Appear and Evolve, *Paleontological Journal*, Vol. 41, No. 5, pp. 481-488 (2007);

ная у белков. Функция узнавания является базой специфического катализа.

V. Каталитическая функция: специфический катализ химических реакций рибозимами. Данная функция аналогична энзиматической функции белков-ферментов»[10].

Граф. 1: Схема функционирования мира РНК до синтеза белка (по Спирину [29]).

Приведенная номенклатура была бы абсолютна только в случае отсутствия комплекса данных функций у других предшественников кода или же при отсутствии альтернативных предшественников нуклеиновых кислот кода как носителей указанных функций. Однако практика говорит об обратном.

Во-первых, существуют нуклеиновые кислоты, отличные от РНК, претендующие на роль прекурсоров первичного кода. В настоящее время известно множество синтетических нуклеиновых кислот (см. таблицу 1), часть которых претендует на роль предбиологических носителей. Многие

[10] Нетрудно заметить, что данный список идентичен критериям идентификации семантид, означенным выше при постановке задачи (см. введение).