

Симметричные закономерности в генетическом коде

Д. Р. Дуплий, С. А. Дуплий, Н. А. Чашин¹

Харьковский национальный университет им. М. Н. Каразина
Пл. Свободы, 4, Харьков, 61077, Украина

¹ Институт молекулярной биологии и генетики НАН Украины
Ул. Академика Заболотного, 150, Киев, 03143, Украина

В работе рассматриваются формальные аспекты проблемы генетического кода. Особое внимание уделено математическим подходам к описанию свойств кода. На основании неравномерной вырожденности генетического кода введена новая числовая характеристика нуклеотида — степень детерминации. Предложена модель дуплетного (ромбического) кода в виде матрицы, полученной внешним произведением, в которой наблюдаются симметричные закономерности. Обсуждаются также различные алгебраические и суперсимметричные модели триплетного генетического кода.

Введение. Выяснение способов хранения и понимание механизмов передачи генетической информации представляют собой важнейшую задачу современного естествознания. Результаты многочисленных исследований в области молекулярной генетики дают объемный материал для всестороннего анализа и математического моделирования [1, 2].

Главным достижением молекулярной генетики можно без преувеличения назвать расшифровку биологического кода, т. е. установление соответствия между последовательностью аминокислот в белке и последовательностью нуклеотидов в молекуле РНК. Это соответствие найдено экспериментально [3, 4], если же его исследовать для поиска симметричных или иных абстрактных законов, то станет возможным более глубокое понимание механизмов наследственной передачи и, вероятно, предсказание новых структурных и динамических объектов. К числу нерешенных проблем генетического кода относится вопрос описания структуры кода как целостной системы [5, 6, 8—11].

Гипотезы систематизации кодонов. Сенсационное известие о расшифровке генетического кода [3] привлекло внимание ученых разных специальностей,

в частности математиков, которые стали рассматривать генетический код как комбинаторную задачу [6—9], применяя такие чисто математические методы анализа, как теория групп [13], построение графов [14], теория векторных пространств [15], теория фракталов [16]. Однако таблица кодирования аминокислот триплетами отражает самый внешний уровень кодирования [17]. Перечитывание одной последовательности символов другой последовательностью наблюдается еще на двух, по крайней мере известных, уровнях. Во-первых, при аминокислотировании тРНК, когда происходит индивидуальное узнавание аминокислоты соответствующим ферментом, что обеспечивает однозначность кодонов. Для каждой аминокислоты имеется единственный фермент, узнающий только одну аминокислоту и все типы изоакцепторных тРНК. Во-вторых, кодирование имеет место при распознавании кодонов мРНК антикодонами тРНК с отклонением от правил комплементарности. Для ряда высокоочищенных тРНК, выделенных из различных источников, было показано, что они способны узнавать несколько различных кодонов [18]. Это означает, что основание в первом положении антикодона должно обладать способностью образовывать пару с несколькими различными основаниями, находящимися в третьем положении соответствующих кодонов. Другими словами, взаимодей-

вие оснований, находящихся в этом положении, не ограничено образованием канонических пар G-C, A-U. Правила, по которым происходит взаимодействие кодон—антикодон, суммированы в гипотезе неоднозначного соответствия [18].

Поиск стереохимического соответствия между аминокислотами и их кодонами впервые проводился Пелком и Велтоном. На основании построенных атомных моделей тринуклеотидов и аминокислот и с учетом возникновения стэкинг-взаимодействий ими найдено достаточно хорошее соответствие во многих, хотя и не во всех случаях. Малая специфичность основания в третьем положении объясняется тем, что оно соответствует группе $H_2N-C-COOH$, имеющейся у всех аминокислот [19].

В [9] выводятся пять семейств аминокислот по количеству атомов водорода на молекулу и наблюдается, что в четырех из них имеются пары аминокислот, у которых Уотсон-Криковская комплементарность (A-U, G-C) для первых оснований кодонов всегда сопровождается комплементарностью типа C-U, A-G для вторых оснований. Комплементарность последнего типа, т. е. оснований, содержащих б-аминогруппу (A и C), в отношении оснований, содержащих, б-карбонильную группу (G и U), может быть постулирована на основании правила Чаргаффа: $A + C = G + U(T)$. Комплементарность оснований по Чаргаффу, как и комплементарность аминокислот в пределах выделенных семейств, возможно, отражает закономерности укладки протяженных молекул нуклеиновых кислот и белков.

На основании этого параллелизма в поведении аминокислот и соответствующих им дуплетов нуклеотидов была предложена идея о структурной аналогии аминокислот и оснований. Предполагается, что на самых ранних этапах органической эволюции, еще до появления нуклеиновых кислот и полипептидов, аминокислоты и основания существовали в виде упорядоченных кристаллических агрегатов, в которых молекулы еще не были соединены ковалентными связями. В частности, автор обращает внимание на то, что в большинстве семейств более низкое значение изоэлектрической точки всегда у той аминокислоты в комплементарной паре, у которой первое основание кодона G или U, т. е. соответственно основание, имеющее более выраженный кислотный характер (относительно C или A). В пятом семействе наблюдается обратная корреляция.

В дополнение к данным, изложенным выше, необходимо подчеркнуть, что распределение 20 общих аминокислот по семействам отражает объективно существующий порядок в распределении дуплетов оснований по кодонам. В самом деле, из

повторения определенных наборов оснований в таблице кажется невозможным вывести какую-нибудь иную систему. Аминокислотные семейства фактически выводятся из совпадения следующих отношений для одних и тех же пар аминокислот: 1) комплементарность Уотсона-Крика или Чаргаффа для первых оснований кодонов; 2) комплементарность Чаргаффа для вторых оснований кодонов; 3) 20 атомов водорода на пару комплементарных аминокислот (исключения из этого правила косвенно подтверждают общий принцип).

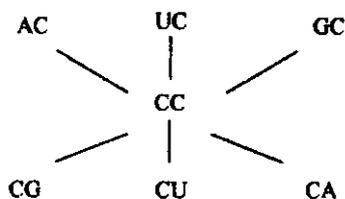
При графическом изображении структур, образованных комбинированием пиримидиновых полукомпартментов, отмечается возникновение зеркальной симметрии в расположении периферических оснований у этой «гибридной» структуры [9]. Симметрия в положении периферических оснований слева и справа — следствие стыковки некомплементарных аминокислот Ile-Phe и Ala-Pro. Такая стыковка может появиться только в результате спонтанной «рекомбинации» при взаимодействии уже сформированных фрагментов исходных структур. В то же время образование «симметричных» (по основаниям-аналогам) пар Ile-Phe, Ala-Pro может означать невозможность дальнейшего продолжения кристаллизации, то есть приращения к структуре новых компонентов («поступающих» в виде комплементарных «асимметричных» пар, например Ile-Ser, Thr-Phe) [9].

Понятие «генетический код» по сути является абстрактным и более широким, чем принятое определение [5, 20]. Своим происхождением оно обязано отчасти моде технотронного века, а отчасти — той наглядности, с которой были представлены соответствия между триплетами и аминокислотами. Но ведь генетический код — это не только система записи одного алфавита посредством другого, эдакая криптограмма или стереохимическая головоломка, а строго функционирующая целостная система, поэтому таблицу кода уместнее назвать генетическим словарем.

Согласно учению А. П. Анохина, в функциональную систему входят «... афферентный синтез, акцептор действия и обратная афферентация о его результатах» [21]. Генетический код обладает всеми этими свойствами, поэтому его можно рассматривать как кибернетическую систему. Обратная афферентация генетической системы, в отличие, например, от дыхательной или нервной, не так скоро и легко наблюдаема. То же можно сказать о структуре генетического кода. Модель топологической структуры кода предложена в работе [14]. Она строится с помощью комбинаторных преобразований Румера [6], который предлагает для систе-

в которой явно видна «сила» каждого дуплета в терминах его степени детерминации, и горизонтальные ряды состоят из равносильных дуплетов. Наглядно, что дуплеты со степенью детерминации больше пяти кодируют одну аминокислоту, со степенью детерминации меньше пяти — две аминокислоты, значение пять — промежуточное: из этого ряда AC и GU детерминируют по одной аминокислоте, а UG и CA — по две. Аминокислоты, кодируемые дуплетами с низкой степенью детерминации (меньше пяти), принадлежат к разным классам по химическому типу радикалов, кроме дуплета GA, детерминирующего Asp и Glu одного химического класса. Примечательно, что в промежуточном случае степень детерминации дуплета пропорциональна числу водородных связей в нем. Пуриновые основания (G и A) образуют по две водородные связи, а пиримидиновые (C и U) — по три. Отметим, что техника векторных пространств для анализа последовательностей ДНК применялась в работе [15], в которой подобие двух последовательностей измерялось внутренним (скалярным) произведением соответствующих векторов.

Закономерности соотношений дуплетов внутри канонической матрицы могут быть представлены в различных интерпретациях ромбического варианта генетического словаря, например, на основе комплементарности кодирующих дуплетов. Так, в [14] на основании анализа канонической матрицы дуплетов построены графы связности для четырех азотистых оснований (тетраэдры с право- и левовращающей симметрией). Основой построения этой структуры, обеспечивающей связь ее элементов, явились единичные переходы азотистых оснований. Возможны переходы между комплементарными пурином-пиримидином $C \leftrightarrow G$, $U \leftrightarrow A$, между гомологичными основаниями $C \leftrightarrow U$, $G \leftrightarrow A$ и между некомплементарными пурином-пиримидином $C \leftrightarrow A$, $G \leftrightarrow U$ (согласно правилу Румера). Во множестве из 16 дуплетов кода, в соответствии с понятием окрестности точки, каждый из них может быть связан единичными переходами (изменением в одном основании) с шестью соседними, что можно изобразить на первом этапе следующим образом:



В развитие этих построений получена структура, изоморфная булеву гиперкубу B^4 [14], в

которой можно выделить четыре октета дуплетов. Связанные в квартеты дуплеты кодируют близкие по структуре аминокислоты: дуплеты квартета AC, UC, UG, AG кодируют в основном аминокислоты, содержащие C-OH и C-SH-группу; квартет GG, CG, UG, AG включает все аргинины кода; квартет GU, CU, UU, AU содержит только неполярные аминокислоты, а квартет AA, GA, CA, UA — только полярные. Таким образом, природа изоморфизма топологической структуры генетического кода булеву гиперкубу оказалась связанной, по [14], с кодированием топологии α -углеродного скелета белковой цепи. Сопоставление с ромбическим вариантом кода (по [22]) позволяет увидеть, что последние три квартета аналогичны рядам с общим основанием. Однако анализ данной структуры показывает, что на самом деле четверки дуплетов замкнуты в цикл. Удачные взаимосвязи аминокислот в структуре белков через отбор систем соответствий должны найти отражение в топологической структуре кода [14, 22]. Иными словами, структура генетического кода должна, по-видимому, представлять собой идеальное пространство, отображающее взаимосвязи аминокислот внутри белкового реального пространства.

Генетический код и симметрия. С математической точки зрения симметрией называется свойство некоторой геометрической фигуры совмещаться с собой при действии некоторой группы ортогональных преобразований. Различают симметрию отражения, или зеркальную, симметрию вращения, симметрию переноса [23]. Геометрические симметрии наиболее наглядные и легко обнаруживаемые в строении животных и растительных организмов [24—26], но они не исчерпывают всего запаса симметрий, существующих в природе. Симметрично не только внешнее строение природных объектов, симметричны и сама природа, и ее законы (в этом случае говорят об инвариантности). Изучая математическую модель той или иной физической системы, физики открывают время от времени новые и неожиданные симметрии, поэтому многообещающим является строгое и абстрактное математическое описание также и биологических систем.

Основатель учения о симметрии Е. Вигнер заметил: «Все законы природы — это условные утверждения, позволяющие предсказывать какие-то события в будущем на основе того, что известно в данный момент, причем для предсказания будущего некоторые аспекты состояния мира в данный момент несущественны» [27]. Следует подчеркнуть, что законы симметрии применимы именно к законам природы, т. е. к корреляциям между событиями, а не к самим событиям. В другой статье

Вигнер определил, что «...функция, которую несут принципы симметрии, состоит в наделении структурой законов природы или установлении между ними внутренней связи так же, как законы природы устанавливают структуру или взаимосвязь в мире явлений» [28].

Таким образом, логично и естественно искать симметричные закономерности в процессах репликации, транскрипции и в самой структуре генетического кода. Существенным продвижением в этом направлении явилось применение теоретико-групповых методов для объяснения эволюции эукариотического генетического кода [29, 30].

Идея заключается в том, что симметрии, наблюдаемые в коде, должны проявлять себя в процессах выбора кодонов для определения различных аминокислот. Так, рассмотрим 64 кодона как 64-размерное неприводимое представление [31] некоторой соответствующим образом выбранной группы Ли [13]. Предполагается, что в начале эволюции они кодировали одну аминокислоту. Наблюдаемые аминокислоты тогда можно трактовать как результат последовательного нарушения симметрии. Раскладывая 64-мерное представление относительно определенного ряда подгрупп, можно получить уменьшающиеся числа подпредставлений с уменьшающимися размерностями. Число представлений, имеющих на каждом шаге разложения, соответствует числу аминокислот, включенных в генетический код. Размерность каждого подпредставления соответствует числу кодонов, которое кодирует соответствующую аминокислоту [29].

Основным предположением, лежащим в основе этого подхода, является тот факт, что среди огромного количества возможных комбинаций кодонов только очень ограниченное их число будет соответствовать картановым симметриям и, следовательно, генерировать эволюцию генетического кода, обусловленную рядом вложенных подгрупп. Среди возможных групп необходимо выделить именно те, которые имеют 64-мерные неприводимые представления, а это ограничивает рассмотрение следующими группами: $SU(2)$, $SU(3)$, $SU(4)$, $Sp(4)$, $Sp(6)$, $SO(13)$, $SO(14)$, G_2 [29]. Далее, поскольку наибольшее количество кодонов, кодирующих одну аминокислоту, равно 6 (γ Arg, Leu, Ser), то наибольшая размерность представлений всех 8 алгебр должна быть нарушена до размерности 6 или меньше. Группы $SO(13)$ и $SO(14)$ не содержат подходящих сохраняющихся симметрий и поэтому не рассматривались, в то время как группа $Sp(6)$ оказалась наиболее многообещающей, поскольку содержит в себе подходящую последовательность динамически нарушенных симметрий [29]

$$\begin{aligned} Sp(6) &\supset Sp(4) \otimes SU(2) && I \\ &\supset SU(2) \otimes SU(2) \otimes SU(2) && II \\ &\supset SU(2) \otimes U(1) \otimes U(1), && III / IV / V \end{aligned} \quad (7)$$

где последовательности I—V можно трактовать как эволюцию генетического кода на ранних этапах органической жизни.

В биохимических терминах динамическая основа алгебраических подходов к генетическому коду лежит в комплементарности нуклеиновых оснований кодонов мРНК и антикодонов тРНК [12]. В целом считается, что эволюция кода представляет собой процесс оптимизации в терминах нарушенных симметрий (7), и ожидается, что корреляция между физико-химическими свойствами кодонов и аминокислот должна отражать их организацию относительно их последовательностей подалгебр. Более того, детали подобных разветвляющихся схем на уровне индивидуальных весовых векторов должны быть в согласии с принятым биологическим пониманием истории кода [5, 11].

Дальнейшее развитие теоретико-групповых математических структур для исследования генетического кода привело к рассмотрению классических супергрупп и супералгебр Ли [32]. В этом подходе предполагается, что типичные представления кодонов также описываются 64-мерными неприводимыми представлениями основных классических супералгебр Ли серии $sl(6/1)$ [10, 33—35] и других подходящих супералгебр, например, $osp(5/2)$ [36]. Была проведена полная классификация всех типичных представлений кодонов — всего 18 представлений для 12 различных супералгебр Ли — с целью нахождения наиболее подходящих для описания вырожденности генетического кода [10, 36]. С одной стороны, как и в несуперсимметричном случае, ни одно из последовательных нарушений суперсимметрии по-прежнему не воспроизводило точно всех численных значений вырожденности, и феномен «замораживания» симметрий (когда в (7) нарушение последних двух $SU(2)$ симметрий до $U(1)$ происходит не для всех кодонов) оставался важной частью этого подхода. В рассмотрении однозначной части приближения допущение возможности «замораживания» симметрий привело к построению трех схем, основанных на ортосимплектической супералгебре $osp(5/2)$ и наиболее точно предсказывающих вырожденность генетического кода [36].

Закключение. Различные типы комплементарности оснований нуклеотидов в генетическом коде используются многими авторами для построения абстрактных моделей, где прослеживаются различ-

ные закономерности и типы симметрий. Можно сделать вывод о том, что, наряду с топологическими и суперсимметричными моделями структуры генетического кода, требует дальнейшего изучения матричная модель дуплетного варианта, использующая понятие степени детерминации кодона. Не исключено, что обобщение существующих абстрактных моделей и сравнение их с экспериментальными данными позволит вскрыть еще один уровень генной организации.

Д. Р. Дуплій, С. А. Дуплій, Н. А. Чащин

Симетрійні закономірності в генетичному кодї

Резюме

У роботі розглянуто формальні аспекти проблеми генетичного коду. Особливу увагу приділено математичним підходам до опису властивостей коду. На підставі нерівномірного виродження генетичного коду запроваджено нову числову характеристику нуклеотиду — ступінь детермінації. Запропоновано модель дуплетного (ромбічного) коду у вигляді матриці, отриманої зовнішнім добутком, у котрій спостерігаються симетричні закономірності. Обговорюються також різні алгебраїчні та суперсиметричні моделі триплетного генетичного коду.

D. R. Duplij, S. A. Duplij, N. A. Chashchin

Symmetric properties of genetic code

Summary

Formal aspects of genetic code problem are considered. Special attention is given to the mathematical approaches for describing the coding process. A new numerical characteristic of nucleotide — determinative degree — is introduced. A model for rhombic code construction as an exterior product is proposed. Various algebraic and supersymmetric models of genetic code are discussed.

СПИСОК ЛІТЕРАТУРИ

1. Девис Р., Бодстайл Д. Методы генной инженерии. Генетика бактерий.—М.: Мир, 1984.—176 с.
2. Гловер Д. Клонирование ДНК. Методы.—М.: Мир, 1988.—538 с.
3. Nirenberg M. W., Matthaei J. H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides // Proc. Nat. Acad. Sci. USA.—1961.—47.—Р. 1588.
4. Стент Г., Кэлиндар Р. Молекулярная генетика.—М.: Мир, 1981.—646 с.
5. Ичас М. Биологический код.—М.: Мир, 1971.—351 с.
6. Румер Ю. Б. Систематизация кодонов в генетическом коде // Докл. АН СССР.—1968.—183, № 1.—С. 225—226.
7. Румер Ю. Б. О систематизации кодонов в генетическом коде // Докл. АН СССР.—1966.—167, № 6.—С. 1393.
8. Румер Ю. Б. О систематизации кодонов в генетическом коде // Докл. АН СССР.—1969.—187, № 187.—С. 937.
9. Суходолец В. В. Смысл генетического кода: реконструкция этапа предбиологической эволюции // Генетика.—1985.—21, № 10.—С. 1589—1599.
10. Bashford J. D., Tsochantjis I., Jarvis P. D. A supersymmetric model for the evolution of the genetic code // Proc. Nat. Acad. Sci. USA.—1998.—95.—Р. 987—992.
11. Maddox J. The genetic code by numbers // Nature.—1994.—367.—Р. 111.
12. Jarvis P. D., Bashford J. D. Systematics of the genetic code and anticodon: history, supersymmetry, degeneracy and periodicity.—Hobart, 1998.—3 p. (Preprint/Univ. of Tasmania, physics/9809030).
13. Куроп А. Г. Теория групп.—М.: ГИЗ ТТЛ, 1953.—467 с.
14. Карасев В. А., Сорокин С. Г. О топологической структуре кода // Генетика.—1977.—33, № 6.—С. 744—751.
15. Bhry T., Cziryk A., Vicsek T., Major B. Application of vector space techniques to DNA // Fractals.—1998.—6, N 3.—Р. 205—210.
16. Devey T. G. Fractals in Molecular Biology.—Oxford: Univ. press, 1997.—327 p.
17. Сингер М., Берг П. Гены и геномы.—М.: Мир, 1998.—373 с.
18. Льюин Б. Гены.—М.: Мир, 1987.—544 с.
19. Welton M. G. E., Pelc S. R. Specificity of the stereochemical relationship between ribonucleic acid-triplets and amino acid // Nature.—1966.—209.—Р. 870—872.
20. Ратнер В. А. Структура и эволюция генетического кода // Итоги науки и техники.—М.: ВНИИТИ, 1985.—С. 158—197.—(Сер. Молекуляр. биология, Т. 21).
21. Горизонтов П. Д. Гомеостаз.—М.: Медицина, 1976.—463 с.
22. Карасев В. А. Ромбический вариант генетического словаря на основе комплементарности кодирующих нуклеотидов // Вестн. Ленингр. ун-та.—1976.—1, N 3.—С. 93—97.
23. Войцеховский М. И. Симметрия // Математическая энциклопедия.—М.: Советская энциклопедия, 1984.—Т. 4.—С. 1150—1151.
24. Петухов С. В. Биомеханика, бионика и симметрия.—М.: Наука, 1981.—239 с.
25. Шафрановский И. И. Симметрия в природе.—Л.: Недра, 1985.—168 с.
26. Депенчук Н. П. Симметрия и асимметрия в живой природе.—Киев: Изд. АН УССР, 1963.—176 с.
27. Визнер Е. Непостижимая эффективность математики в естественных науках // Этюды о симметрии.—М.: Мир, 1971.—С. 182—198.
28. Визнер Е. Симметрия и законы сохранения // Этюды о симметрии.—М.: Мир, 1971.—С. 20—34.
29. Hornos J. E. M., Hornos Y. M. M. Model for the evolution of the genetic code // Phys. Rev. Lett.—1993.—71.—Р. 4401.
30. Findley G. L., Findley A. M., McGlynn S. P. Symmetry characteristics of the genetic code // Proc. Nat. Acad. Sci. USA.—1982.—79, N 22.—Р. 7061—7065.
31. Кириллов Ф. Ф. Элементы теории представлений.—М.: Наука, 1978.—343 с.
32. Кас В. G. Lie superalgebras // Adv. Math.—1977.—26, N 1.—Р. 8—96.
33. Bashford J. D., Jarvis P. D., Tsochantjis J. Supersymmetry and the genetic code // Physical Applications and Mathematical Aspects of Geometry, Groups, and Algebras.—Singapore: World Sci., 1997.—Р. 826—831.
34. Forger M., Sachse S. Lie superalgebras and the multiplet structure of the genetic code I: Codon representations.—Sao Paulo, 1998.—23 p. (Preprint/Inst. de Mat. e Estat., math-ph/9808001).
35. Bashford J. D., Tsochantjis I., Jarvis P. D. Codon and nucleotide assignments in a supersymmetric model of the genetic code // Phys. Lett.—1997.—A233.—Р. 481—488.
36. Forger M., Sachse S. Lie superalgebras and the multiplet structure of the genetic code II: Branching schemes.—Sao Paulo, 1999.—34 p. (Preprint/Inst. de Mat. e Estat., math-ph/9905017).

УДК 575

Поступила в редакцию 14.09.99