

ГЕНЕТИЧЕСКИЙ КОД КАК СИСТЕМА

В. А. РАТНЕР

Новосибирский государственный университет

GENETIC CODE AS A SYSTEM

V. A. RATNER

Deciphering the Genetic Code is one of the most outstanding scientific discoveries of the XXth Century. The Genetic Code is not a random set of t-RNA anticodons and corresponding aminoacids, but a highly organized system which manifests some general features and regularities: unique triplet nature, degeneracy, connectedness, regularity, symmetry. The molecular mechanisms capable to provide the key features of genetic code are discussed.

Расшифровка генетического кода – одно из самых выдающихся научных открытий XX века. Генетический код представляет собой не случайный конгломерат соответствий между кодонами м-РНК и аминокислотами белков, а высокоорганизованную систему, проявляющую общие свойства и закономерности: триплетность, однозначность, вырожденность, связность, регулярность, симметричность. Обсуждаются молекулярные механизмы, обеспечивающие многие свойства генетического кода.

www.issep.rssi.ru

“Самым трудным в проблеме кода было понять, что код существует. На это потребовалось целое столетие. Когда это поняли, то для того чтобы разобраться в деталях, хватило каких-нибудь десяти лет” [1].

Проблема генетического кода – это ключевая проблема. В конце 50-х – начале 60-х годов она приковывала к себе внимание, возбуждала активность умов, побуждала веру в величие и мудрость загадок науки. В широком смысле генетический код – это способ записи генетической информации в последовательностях нуклеиновых кислот (ДНК или РНК) о структуре полипептидов (белков). В конкретном смысле генетический код – это соответствие между триплетными кодонами матричной РНК (м-РНК) и аминокислотами кодируемого белка, задаваемое кодовой таблицей (табл. 1).

Развитие проблемы генетического кода прошло несколько этапов. Предтечами этой проблемы можно считать многих выдающихся исследователей. В частности, Н.К. Кольцов (1927, 1935) предложил в общей форме идею молекулы-гена и матричный принцип ее дублирования. Э. Шрёдингер (1944) явно сформулировал необходимость кодирования генетической информации в структуре генов-молекул. П. Колдуэлл и С. Хиншельвуд (1950) предложили идею матричного синтеза белков на ДНК. А. Даунс (1952) сформулировал гипотезу о синтезе белков на РНК.

Научные представления о генетическом коде как о реальной проблеме эксперимента и теории были сформулированы Г.А. Гамовым сразу же после обоснования Дж. Уотсоном и Ф. Криком (1953) модели строения двойной спирали ДНК. Первый этап изучения проблемы (1953–1961) можно назвать гипотетическим. Из модели Уотсона–Крика (см. подробнее [1]) вытекало представление о линейной последовательности ДНК – тексте, построенном из четырех типов нуклеотидов (А, Т, G и С – четыре символа алфавита). Но кодируемые белки тоже имеют линейную первичную структуру – текст, построенный из 20 типов канонических аминокислот (алфавит из 20 символов). Поэтому Г.А. Гамов (1954) сразу же сформулировал идею генетического кода в конкретном смысле – как соответствие двух текстов, записанных при помощи двух разных алфавитов.

Таблица 1. Словарь канонического универсального генетического кода *E. coli* [2]

| Нуклеотид кодона | | | | | | | | | |
|------------------|--|--------------|---|-----------|--|-----------|--|-----------|--------|
| Первый | Второй | | | | | | | | Третий |
| | U | | C | | A | | G | | |
| U | UUU | Б Phe | UCU | М | UAU | Б Tyr | UGU | М Cys | U |
| | UUC | НП | UCC | П Ser | UAC | П | UGC | НП | C |
| | UUA | Leu | UCA | | UAA | Term | UGA | Term | A |
| | UUG | Cp | UCG | | UAG | | UGG | Трп Б | G |
| C | CUU | Cp | CCU | М | CAU | Б His | CGU | Б | U |
| | CUC | НП Leu | CCC | НП Pro | CAC | П | CGC | П Arg | C |
| | CUA | | CCA | | CAA | Gln | CGA | | A |
| | CUG | | CCG | | CAG | Cp | CGG | | G |
| A | AUU | Cp | ACU | М | AAU | Cp Asn | AGU | М Ser | U |
| | AUC | Ile | ACC | П Thr | AAC | П | AGC | П | C |
| | AUA | НП Cp | ACA | | AAA | Lys | AGA | Arg | A |
| | AUG | Met F-Met | ACG | | AAG | Б | AGG | Б | G |
| G | GUU | М | GCU | М | GAU | Cp Asp | GGU | М | U |
| | GUC | НП Val | GCC | НП Ala | GAC | П | GGC | НП Gly | C |
| | GUA | | GCA | | GAA | Glu | GGA | | A |
| | GUG | | GCG | | GAG | Cp | GGG | | G |
| | <i>Все аминокислоты неполярны, не крайних свойств и размеров</i> | | <i>Все аминокислоты малые, а основы сильные</i> | | <i>Все аминокислоты полярные и не малые, а основы слабые</i> | | <i>Крайние варианты аминокислот и аномалии серий</i> | | |

Обозначения аминокислот: Phe – фенилаланин, Leu – лейцин, Ile – изолейцин, Met – метионин, F-Met – формил-метионин, Val – валин, Ser – серин, Pro – пролин, Thr – треонин, Ala – аланин, Tyr – тирозин, His – гистидин, Gln – глутамин, Asn – аспарагин, Lys – лизин, Asp – аспарагиновая кислота, Glu – глутаминовая кислота, Cys – цистеин, Arg – аргинин, Gly – глицин, Term – терминальный нонсенс. Другие обозначения см. в тексте.

Кроме того, он предложил использовать технические средства криптографии (расшифровки неизвестных кодов) для решения центральной проблемы генетики.

Генетический код сразу же приобрел облик великой загадки природы, ребуса для остроумных. Многие сотни математиков, физиков, химиков, биологов, включая Г.А. Гамова, Ф. Крика и др., предложили гипотетические варианты генетического кода, которые представляют теперь лишь исторический интерес. Реальный код оказался совсем иным.

Научными результатами первого этапа можно считать [1]: 1) постановку проблемы генетического кода; 2)

формирование понятий линейного текста, алфавита для нуклеиновых кислот и белков, генетической информации, записанной в этих текстах при помощи символов алфавита; 3) представление о матричной роли РНК в трансляции; 4) понятие о кодонах и доказательство их неперекрывания; 5) предположение о триплетности кодонов и коллинеарности гена и белка, доказанное лишь в дальнейшем, и т.д.

Второй этап (1961–1966) можно назвать экспериментальным, так как в этот период генетический код был расшифрован в прямом эксперименте [1–4]. В 1961 году Ф. Крик с сотрудниками в блестящей работе

показали, что: а) кодоны триплетны; б) между ними нет разделительных знаков (“запятых”); в) гены, кодирующие структуру белков (цистроны), имеют фиксированное начало, ориентированное направление и фиксированный конец; г) существует небольшое число некодирующих триплетов (“нонсенсов”, бессмысленных кодонов), а код в целом сильно вырожден. В 1964 году Ч. Янофски с сотрудниками и С. Бреннер с сотрудниками показали, что ген и кодируемый им белок взаимно коллинеарны, то есть имеется последовательное соответствие между кодонами гена и аминокислотами белка.

Прямая расшифровка генетического кода *in vitro* оказалась возможной благодаря технике белкового синтеза в бесклеточных системах [1–3], то есть в клеточных экстрактах, содержащих все необходимые компоненты аппарата трансляции (т-РНК, рибосомы, аминокислоты, ферменты, источник энергии и т.д.), кроме м-РНК. Вводя в такие системы естественные м-РНК или искусственные небольшие олигорибонуклеотиды, можно было изучать специфичность включения меченых аминокислот в строящиеся полипептиды. М. Ниренберг и Ф. Ледер подавали в бесклеточную систему трансляции *E. coli* различные олигорибонуклеотиды и показали, что индивидуальные фракции тририбонуклеотидов, ассоциированные с рибосомами, связывают определенные фракции т-РНК, заряженные определенными мечеными аминокислотами. С помощью такого метода генетический код был расшифрован полностью. Летом 1966 года на симпозиуме по количественной биологии в Колд-Спринг-Харборе (США) все полученные данные были сведены Ф. Криком воедино [2]. Расшифрованный генетический код *E. coli*, исследованный *in vitro*, полностью согласовывался также с другими независимыми данными, полученными *in vivo* и для других видов. Этот вывод подтверждается также результатами секвенирования последних лет, когда найдено, что тысячи генов и кодируемых ими белков действительно соответствуют друг другу по правилам генетического кода.

Полный словарь генетического кода *E. coli* приведен в табл. 1 [2, 4]. Его анализ позволяет выявить общие свойства генетического кода как целого.

Из 64 возможных триплетов 61 является смысловым кодоном, то есть кодирует аминокислоты. Все кодоны триплетны, неразрывны и не перекрываются в тексте, а также не разделены межкодонными знаками (запятые). Все кодоны однозначны, то есть каждый кодирует единственную аминокислоту. Иначе говоря, в направлении кодон \Rightarrow аминокислота генетический код однозначен.

Обратное соответствие в направлении аминокислота \Rightarrow кодон неоднозначно, и это свойство называется вырожденностью. Отдельные аминокислоты кодируются группами (сериями) кодонов-синонимов. 18 серий из 20 содержат от двух до шести кодонов, две серии (Met и Trp) не вырождены, содержат по одному кодону. Средняя вырожденность генетического кода приблизительно три кодона на серию.

Семнадцать серий из 18 вырожденных имеют свойство связности, то есть образуют тесные группы. На рис. 1 понятие связности иллюстрируется графически. Вершины графов отвечают кодонам, а соединяющие их ребра – заменам одиночных нуклеотидов. Приведены примеры связных серий по четыре (Val – рис. 1, а) и шесть (Leu – рис. 1, б) кодонов. В результате из любого кодона связной серии можно перейти к любому другому синониму путем последовательных замен, не выходя при этом за пределы графа этой серии. Если такой переход можно сделать за один шаг, то серия называется полностью связной (см. рис. 1, а). Связность нарушена только в одной серии (Ser – рис. 1, в), но и тогда она распадается на две связные подсерии.

Вырожденность называется систематической, если кодоны-синонимы различаются в третьей позиции либо пуринами (R = A или G), либо пиримидинами (Y = U или C), либо вообще любыми из четырех нуклеотидов (N = A, G, U или C). Этим принципам удовлетворяют 30 пар кодонов из 32, а также восемь тетрад из 16. Все эти пары связны, а тетрады полностью связны. Остальные

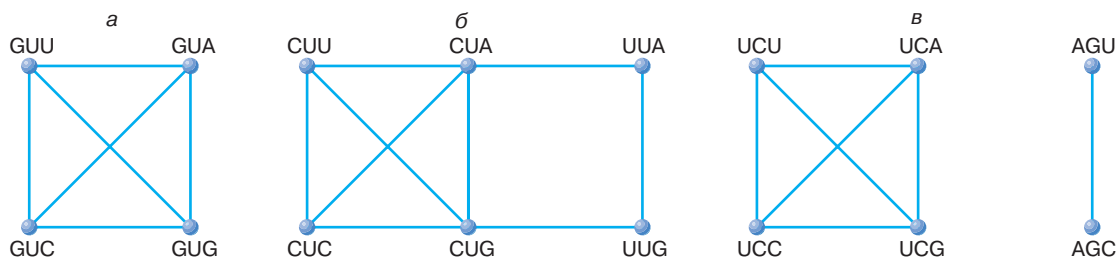


Рис. 1. Графы связности для некоторых кодовых серий. а – серия Val, вырожденность 4 кодона, полностью связная; б – серия Leu, вырожденность 6, связная; в – серия Ser, вырожденность 6, несвязная, но распадается на две полностью связных подсерии

варианты вырожденности называют несистематическими. Они относятся к большим сериям (см. табл. 1): Leu и Arg — связанные серии, Ser — несвязная серия, Pe — три кодона, полносвязная серия.

Генетический код содержит также знаки пунктуации (начала и конца) трансляции. Кодоны AUG, GUG и UUG у прокариот помимо кодирования аминокислот кодируют также инициацию трансляции. Однако однозначность кодирования при этом не нарушается, так как иницирующие знаки локализованы в определенном окружении (контексте), способном образовывать самокомплементарные “шпильки”. У эукариот иницируют триплеты AUG и более слабо — UUG, AUA и ACG. Три “вакантных” триплета у *E. coli* — UAA (*ochre*), UAG (*amber*) и UGA (*opal*) — не кодируют аминокислот, а выполняют роль терминальных знаков трансляции (стоп-кодонов, нонсенс-кодонов или терминальных нонсенов). В норме ими заканчиваются все цистроны, то есть транслируемые гены, единицы трансляции. Мутационное возникновение нонсенов внутри гена приводит к преждевременной термации трансляции и обрыву белка. Нонсены тоже образуют связную серию (см. табл. 1).

Расшифровка генетического кода была одним из самых выдающихся научных открытий XX века.

Третий этап изучения проблемы генетического кода (после 1966 года) связан с углубленным исследованием молекулярных механизмов кодирования, системных свойств генетического кода: симметрии, регулярности, помехоустойчивости, универсальности, а также путей его возникновения и эволюции (см. [4]). Молекулярной системой, обеспечивающей соответствие кодонов м-РНК и аминокислот, является набор адапторных молекул транспортных РНК (т-РНК) и набор кодирующих ферментов аминоксил-т-РНК-синтетаз (АРС-аз). Каждая специфическая молекула т-РНК имеет антикодон, взаимодействующий с кодоном м-РНК, а также специфический сайт взаимодействия с определенной АРС-азой и неспецифический сайт связывания аминокислоты (рис. 2). Каждая АРС-аза опознает все изоакцепторные (переносящие одну аминокислоту) фракции т-РНК, одну определенную аминокислоту и соединяет их макроэргической (энергобогатой) связью. Поэтому соответствие антикодона т-РНК и аминокислоты определяется именно АРС-азой. Фракции т-РНК выполняют функции адапторов (специфических посредников) между кодонами м-РНК и аминокислотами.

Многие свойства генетического кода обеспечиваются свойствами молекул т-РНК и АРС-аз. Триплетный и неразрывный антикодон выделен в антикодонной петле т-РНК специальными модифицированными нуклеотидами (см. рис. 2). Этим обеспечиваются триплетность и неразрывность узнаваемых кодонов мат-

рицы. Все антикодоны одинаково триплетны, поэтому, начиная от иницирующего знака, трансляция осуществляется триплетными шагами, то есть формируется определенная рамка (фаза) трансляции — одна из трех возможных. В этом случае межкодонные знаки (запяты) не нужны, а кодоны не перекрываются. Иницирующие кодоны у *E. coli* опознаются специальной фракцией т-РНК^{F-Met}, переносящей модифицированную аминокислоту формил-метионин. Терминальные нонсены вообще не имеют своих фракций т-РНК, а опознаются специальными белковыми факторами термации.

Однозначность кода в направлении кодон ⇒ аминокислота обеспечивается строгой специфичностью АРС-аз. Каждая АРС-аза узнает единственную аминокислоту, поэтому неоднозначность исключена или маловероятна. В основе систематической вырожденности лежат правила неоднозначности спаривания кодон-антикодон, установленные Ф. Криком [1, 4]. Один антикодон может узнавать один, два или три кодона, различающиеся по третьей позиции (табл. 2). Согласно правилам неоднозначного спаривания, систематическая

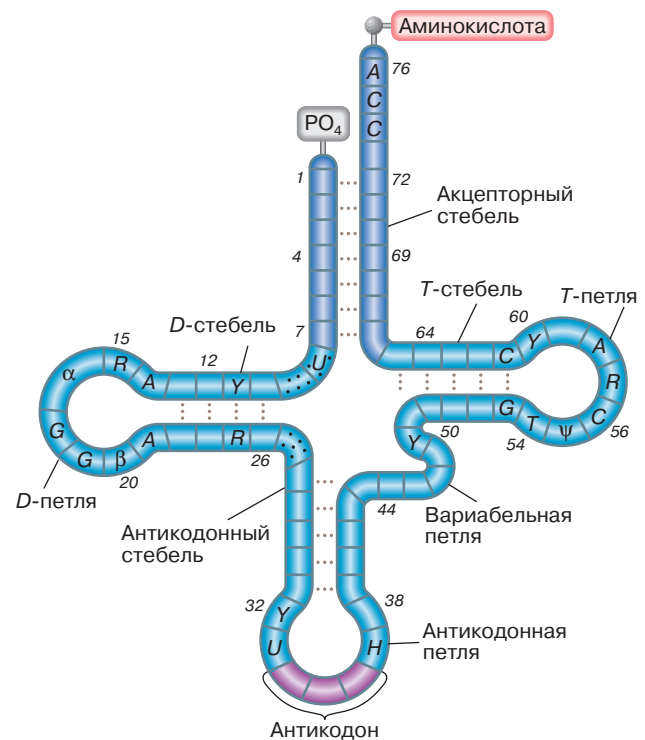


Рис. 2. Обобщенная вторичная структура молекул транспортных РНК (клеверный лист). Указаны основные фрагменты структуры: антикодон, ограниченный модифицированными нуклеотидами, выделяющий триплетную рамку; неспецифический участок АСС, связывающий аминокислоту; четыре двухспиральных участка (стебля) и три неспаренные петли

Таблица 2. Неоднозначность спаривания нуклеотидов в третьем положении кодона и антикодона [2]

| Антикодон | Кодон |
|-----------|---------|
| U | A, G |
| C | G |
| A | U |
| G | U, C |
| I | U, C, A |

Примечание: I – инозин.



вырожденность в парах кодонов обеспечивается отдельными фракциями т-РНК, имеющими U, G или I (инозин) в трех позициях антикодонов. Вырожденность 3 у изолейцина (Ile) требует фракцию т-РНК с I в третьей позиции антикодона. Такой нуклеотид там действительно есть. Вырожденность 4 требует не менее двух фракций т-РНК, вырожденность 6 – не менее трех фракций. Всего генетический код *E. coli* требует не менее 32 фракций т-РНК. Реально у *E. coli* полное число генов т-РНК равно 86 для 79 фракций с различными антикодонами. Следовательно, многие фракции т-РНК частично дублируют друг друга.

Анализ таблицы генетического кода как целого (табл. 1, рис. 3) позволяет выявить удивительные свойства его регулярности и симметрии [4]. Обозначим позиции нуклеотидов в кодонах (5'–1–2–3–3') через $x - y - z$, где x – приставка, y – корень, z – окончание, а xu – основа кодона. Эти термины отражают сходные понятия лингвистики. Корни слов определяют их смысл. Все мутации, затрагивающие корень (y) кодона, также изменяют кодовую серию, то есть нарушают смысл кодона. Приставки слов тоже участвуют в определении смысла, хотя и не так жестко, а многие их изменения меняют смысл слов. Замены в приставках (x) кодонов чаще всего изменяют их смысл, но иногда являются синонимическими. Окончания слов обычно участвуют в словоизменении, то есть в синонимических преобразованиях. Аналогично 70% замен в окончаниях (z) кодонов синонимические. Наконец, приставка и корень слова образуют его основу, несущую полную или доминирующую смысловую нагрузку. Основа кодона (xu) тоже играет ключевую роль в особенностях генетического кода.

Регулярность генетического кода связана с распределением основных свойств кодонов и аминокислот по столбцам (корням) генетического кода (см. табл. 1). Кодоны характеризуются свойствами их основ и корней. Кодоны, имеющие одинаковые основы, образуют 16 тетрад генетического кода. Основа называется силь-

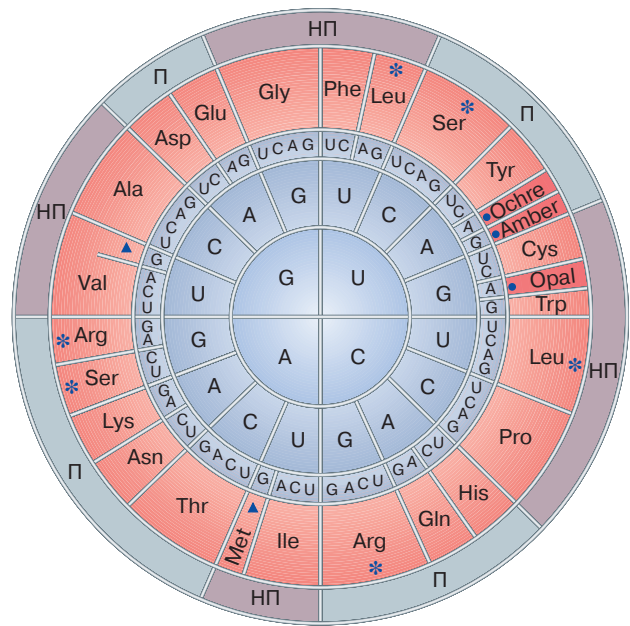


Рис. 3. Генетический код в круговой форме. Для наглядности симметрии важен избранный порядок символов по часовой стрелке: U–C–A–G.

Обозначения аминокислот такие же, что в табл. 1, П – полярные, НП – неполярные. Другие обозначения: ● – терминальные нонсенсы (Term.), ▲ – кодоны, кодирующие аминокислоты Met и Val, но в определенном контексте играющие роль начальных знаков трансляции; * – серии с вырожденностью 6. Правила симметрии см. в тексте

ной, если она полностью определяет смысл (аминокислоту) кодонов тетрады. Нуклеотид в третьей позиции тоже, конечно, необходим, но он может быть любым (N). Такковы основы CU, GU, UC, CC, AC, GC, CG, GG. Легко заметить, что они содержат нуклеотиды в соотношении C : G : U : A = 7 : 5 : 3 : 1. Основа называется слабой, если для однозначного кодирования аминокислоты необходимо также участие определенного третьего нуклеотида (z) в кодоне. Такковы основы UU, AU, UA, CA, AA, GA, UG, AG. Они содержат нуклеотиды в обратном соотношении C : G : U : A = 1 : 3 : 5 : 7. Известно, что участки ДНК с избытком пар G–C более стабильны, чем A–T-богатые участки. Поэтому сильные основы образуют в среднем больше водородных связей с антикодонами т-РНК, чем слабые.

Аминокислоты имеют два основных характерных свойства, существенных в пространственной структуре глобулярных белков: размер (малые – М, средние – Ср, большие – Б) и полярность/неполярность (П/НП).

Легко заметить, что столбцы генетического кода, отвечающие определенным корням кодонов, имеют некоторые групповые свойства (см. табл. 1). Третий столбец (корень А) содержит полярные аминокислоты,

немалые по размеру, а все основы кодонов слабые. Это как бы групповое свойство корня А. Второй столбец (корень С) содержит аминокислоты, малые по размеру, а все основы кодонов сильные. Первый столбец (корень U) включает неполярные аминокислоты разных (но не крайних) размеров, а свойства кодонов неоднозначны. Наконец, четвертый столбец (корень G) содержит все крайние и аномальные варианты аминокислот и кодовых серий [4]: самую реактивную аминокислоту (Cys), самую большую и плоскую (Trp), самую большую и корявую (Arg), самую маленькую (Gly), часть единственной несвязной серии (Ser) и неоднозначный терминальный нонсенс (UGA), который в ряде случаев кодирует 21-ю аминокислоту – селеноцистеин (Sec). Таким образом, если групповые свойства корней (и столбцов) отвечают каким-то общим правилам их возникновения, то четвертый столбец скорее напоминает свалку всего, что не попало в первые три столбца по групповым правилам их формирования.

Теперь рассмотрим не менее впечатляющее свойство симметрии генетического кода. На рис. 3 генетический код изображен в круговой форме [4], где внутренний круг отвечает первым позициям кодонов, среднее кольцо – вторым позициям и внешнее кольцо – третьим позициям. Сильные основы изображены неподделенными секторами внешнего кольца, а слабые – подразделенными. Свойство симметрии состоит в следующем:

1) проведем ось симметрии через центр круга перпендикулярно плоскости листа и повернем круг на 180° в плоскости листа. При этом все сильные и слабые основы сохраняют свои позиции, то есть совмещаются с одноименными;

2) проведем через центр плоскость симметрии, перпендикулярную плоскости листа и строкам текста. При зеркальном отражении круга в этой плоскости все сильные основы меняются местами со слабыми и наоборот;

3) проведем через центр плоскость симметрии, перпендикулярную плоскости листа и параллельную строкам текста. При зеркальном отражении круга в этой плоскости сильные основы меняются на слабые и наоборот.

Генетический код универсален в том смысле, что его основная часть одинакова для всех форм жизни на Земле. Этот вывод обоснован опытом массового секвенирования генов и белков. Почти всегда коллинеарное соответствие генов и белков согласуется с правилами генетического кода. Однако в некоторых экзотических системах трансляции (митохондрии животных, растений и грибов, хлоропласты растений, мельчайшие бактерии – микоплазмы, реснитчатые простейшие и др.) найдены минорные отклонения в генетическом коде, а также изменения правил неоднозначного спаривания и наборов антикодонов и фракций т-РНК. Это своеобразные

разные “диалекты” генетического кода, отражающие специфику их эволюции и функционирования.

Несомненно, что генетический код явился продуктом добиологической молекулярной эволюции и продолжал частично эволюционировать в дальнейшем. В стохастическом процессе молекулярной эволюции свойства генетического кода могли быть: 1) либо предзаданы (преддетерминированы) физико-химическими характеристиками компонент и условий, 2) либо отобраны как адаптивные среди альтернативных вариантов, 3) либо фиксированы случайно. Гипотезы возникновения генетического кода в разной степени учитывают эти возможности [3, 4].

Так, гипотеза “замороженного случая” (Ф. Крик, 1968 год) полагала, что исторически была фиксирована первая случайная, но удовлетворительная система кодирования, которая далее была размножена, подверглась эволюционному усложнению и оптимизации, так как обеспечивала ускоренное воспроизведение. Ясно, что крайний, чисто случайный вариант этой гипотезы нереален, поскольку код обладает очевидными неслучайными системными свойствами. Ясно также, что эти свойства отражают неслучайный, высоко организованный характер генетического кода, связанный с правилами синонимии кодовых серий.

Таким образом, генетический код *E. coli* представляет собой не случайный конгломерат соответствий между кодонами и аминокислотами, а высокоорганизованную систему соответствий, поддерживаемую сложными молекулярными механизмами. По выражению Френсиса Крика, внесшего решающий вклад в открытие и изучение кода, “это ключ к молекулярной биологии, поскольку он показывает, как два великих языка полимеров – язык полинуклеотидов и язык полипептидов связаны между собой” [2].

ЛИТЕРАТУРА

1. Ичас М. Биологический код. М.: Мир, 1971.
2. The Genetic Code. Cold Spring Harbor Symp. Quant. Biol. Cold Spring Harbor; N.Y. 1966. 31.
3. Молекулярная генетика. М.: Мир, 1963.
4. Ратнер В.А. Молекулярная генетика: Принципы и механизмы. Новосибирск: Наука, 1983.

Рецензенты статьи С.Г. Инге-Вечтомов, Л.И. Корочкин

* * *

Вадим Александрович Ратнер, доктор биологических наук, профессор кафедры цитологии и генетики ФЕН НГУ, зав. лабораторией молекулярно-генетических систем Института цитологии и генетики СО РАН, академик РАЕН. Область научных исследований – теоретическая генетика, теория молекулярно-генетических систем, теория молекулярной эволюции. Автор и соавтор 14 монографий на русском, английском и немецком языках, свыше 330 статей и других публикаций.