

I 21770

АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ОТДЕЛЕНИЕ ОБЩЕЙ БИОЛОГИИ И ОТДЕЛЕНИЕ БИОХИМИИ, ФИЗИОЛОГИИ И ТЕОРЕТИЧЕСКОЙ МЕДИЦИНЫ

ЦИТОЛОГИЯ И ГЕНЕТИКА

CYTOLOGY AND GENETICS

НАУЧНЫЙ ЖУРНАЛ

ТОМ VI, № 6

НОЯБРЬ — ДЕКАБРЬ

897561
197768

ВЫХОДИТ 6 РАЗ В ГОД

ГОД ИЗДАНИЯ 6-Я

ИЗДАТЕЛЬСТВО «НАУКОВА ДУМКА»

КИЕВ 1972

УДК 575 + [016.3]

О ФОРМАЛЬНОЙ СТРУКТУРЕ ГЕНЕТИЧЕСКОГО КОДА

А. Г. ВОЛОХОНСКИЙ

Известная в настоящее время форма таблицы генетического кода до сих пор не получила исчерпывающего объяснения. Вместе с тем, факт универсальности кода заставляет искать какой-то общий принцип, на основании которого могли бы существовать именно те группы триплетов

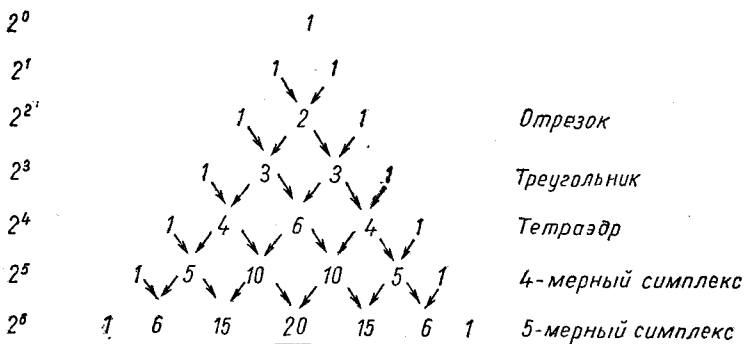


Рис. 1. Треугольник Паскаля и симплексы, соответствующие рядам биномиального распределения. Стрелкой обозначен знак суммирования.

РНК, объединенные соответствием определенным аминокислотам, которые мы видим в кодовой таблице. После остроумных гипотез Гамова [1] и Крика [2], которые оказались несовместимыми с картиной реального кода, не делалось даже попыток ответить на вопрос, почему именно 20 аминокислот — не больше и не меньше — кодируются 64 триплетами РНК. Но как раз этот факт и оказался ключевым для разрешения многих проблем, связанных со структурой кода.

Мы обратили внимание на то, что число возможных триплетов РНК, полученных из четырех оснований У, Ц, А и Г, представляет собой шестую степень числа 2.

$$4^3 = 2^6 = 64.$$

Это означает, что все множество из 64 триплетов можно разложить как 6-й ряд биномиального распределения в треугольнике Паскаля (рис. 1). В середине ряда обнаруживается число 20, что указывает на глубокую онтологическую связь между числами 64 и 20. Далее можно рассмотреть, из каких групп образуется вся совокупность кодонов РНК при построении треугольника Паскаля и как соотносятся эти группы с числом 20 — т. е. построить «формальный код», учитывающий связь числа 20 с другими группами распределения только формальным образом — через правила построения треугольника Паскаля.

Группы биномиального распределения ряда 2^6 построены в треугольнике Паскаля следующим образом:

$$20 = 10 + 10 = (4 + 6) + (6 + 4)$$

$$15 \cdot 2 = (10 + 5) \cdot 2 = [(4 + 6) + (4 + 1)] \cdot 2$$

$$6 \cdot 2 = (5 + 1) \cdot 2 = [(4 + 1) + 1] \cdot 2$$

$$1 \cdot 2 = 2 \text{ (см. также рис. 2).}$$

Числа 4, 6, 6, 4, из которых построено число 20, входят по правилу построения треугольника Паскаля во все группы разложения 2^6 , кроме крайних двух единиц. При этом обе четверки участвуют в сложениях по четыре раза, а обе шестерки — по два раза. Таким образом, в совокупности из 64 триплетов образуется 20 «групп общности», из которых $4 + 4 = 8$ групп общности (I)* содержит каждая по 4 триплета, а $6 + 6 = 12$ групп общности (II) — по

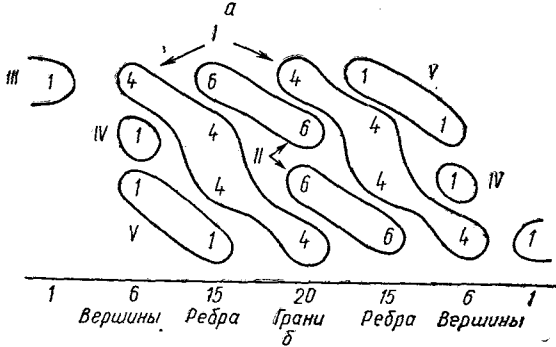
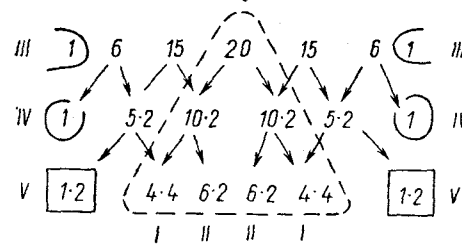


Рис. 2. Два способа образования групп формального кода из ряда биномиального распределения 2^6 и соответствие групп формального кода поверхностным элементам икосаэдра.

1 2	У	Ц	А	Г	З
У	ФЕН(II)	СЕР(I)	ТИР(II)	ЦИС(II)	У
	ЛЕЙ(IV)		осће(III)	umber(II)	Ц
			amber(III)	ТРИ(II)	А
Ц	ЛЕЙ(I)	ПРО(I)	ГИС(II)	АРГ(I)	У
			ГЛУН(II)		А
А	ИЛЕ(II)	ТРЕ(I)	АСПН(II)	СЕР(V)	У
	МЕТ(II)		ЛИЗ(II)	АРГ(V)	Ц
Г	ВАЛ(I)	АЛА(I)	АСП(II)	ГЛИ(I)	У
			ГЛУ(II)		Ц
					А

Рис. 3. Соответствие групп формального кода реальному генетическому коду.

2 триплета — и, кроме того, 8 триплетов, не вошедших в группы общности, остаются по краям рядов разложения (рис. 2). Соответственно положению в рядах треугольника Паскаля эти 8 триплетов образуют еще три группы формального кода:

$$8 = 2 + 2 + 2 \cdot 2,$$

$$\begin{matrix} \text{(III)} & \text{(IV)} & \text{(V)} \\ \text{6-й ряд} & \text{5-й ряд} & \text{4-й ряд} \end{matrix}$$

причем пара триплетов группы III никак не связана с 20 группами общности (I и II), а остальные $4 + 2 = 6$ триплетов (IV и V) связаны с ними косвенно — через группы биномиального распределения 6-го ряда. Если термин «общность» заменить теперь термином «кодирование определенной аминокислоты», получается, что из 20 аминокислот 8 кодируется четырьмя и 12 — двумя кодонами каждая. Кроме того, должны существовать 2 «некодирующих» триплета и 6 триплетов, играющих какую-то особую роль.

Полученный таким образом формальный код состоит из следующих пяти групп:

I — 32 кодона, составляющие 8 групп общности (по 4 кодона в каждой) и соответствующие 8 аминокислотам;

II — 24 кодона, составляющие 12 групп общности (по 2 кодона в каждой) и соответствующие 12 аминокислотам;

* Римскими цифрами мы обозначаем группы формального кода.

- III — 2 кодона, никак не соотносящиеся с 20 группами общности;
- IV — 2 кодона, соотносящиеся с 20 группами общности через группы биномиального распределения;
- V — 4 кодона, соотносящиеся с 20 группами общности через группы биномиального распределения, но иначе, чем группа IV.

Обнаруженные группы формального кода вполне изоморфны реальному генетическому коду (рис. 3). 32 (8·4) кодона группы I для 8 аминокислот, кодируемых 4 и более триплетами, соответствуют 8 группам триплетов, начинающихся с УЦ, ЦУ, ЦГ, ЦЦ, ГУ, ГЦ, ГГ и АЦ; двум дополнительным к кодируемому 6 триплетами лейцину — кодонам УУА и УУГ — соответствует пара IV, а двум парам дополнительных к серину и аргинину кодонов, начинающихся в АГ — группа V (серин и аргинин также кодируются 6 кодонами); двум универсальным консенс-кодонам УАГ и УАА отвечает группа III, никак не связанная с 20 группами общности I и II. Наконец, оставшиеся 24 триплета (для аминокислот, кодируемых 2, а также 3 и 1 триплетами) должны быть сопоставлены с 12 парами группы II. В триплетях, начинающихся с УГ и АУ, принцип распределения по парам нарушен. Характерно, что триплет УГА не является универсальным нонсенс-кодоном, так как у некоторых организмов он кодирует цистеин [3]. Поэтому ситуацию в обеих группах, отклоняющихся от формального распределения, можно признать весьма сходной. Во всем прочем формальный код вполне подобен реальному.

Остановимся на операциях, приводящих к образованию групп формального кода. Рис. 2, по существу, изображает операцию, обратную построению соответствующих рядов биномиального распределения. Ввиду того, что каждое число «P» ($n-1$)-го ряда входит в группы n -го ряда дважды (рис. 1), устанавливается общность (или близость свойств) «P» пар элементов из двух соседних групп n -го ряда. Построение треугольника Паскаля можно представить как развернутую операцию удвоения. Ряд 2^6 получается удвоением 2^5 , а ряд 2^5 — удвоением ряда 2^4 (см. рис. 2, б). Применение изложенного здесь принципа и объединение полученных групп с учетом билатеральной симметрии треугольника Паскаля приводит к образованию пяти групп формального кода.

Смысл обнаруженных соответствий становится понятным при их геометрической интерпретации. Каждый ряд биномиального распределения 2^n представляет набор элементов симплекса в $(n-1)$ -мерном пространстве (рис. 1). Ряд 2^6 изображает набор элементов симплекса для пятимерного пространства (1 центр, 6 вершин, 15 ребер, 20 граней, 15 трехмерных, 6 четырехмерных и 1 пятимерная грань). Проекцией этого тела в трехмерном пространстве является икосаэдр, обладающий аналогичным набором поверхностных элементов, т. е. вершин ($6+6=12$), ребер ($15+15=30$) и граней (20), 2 элемента из 64 не относятся к числу поверхностных элементов икосаэдра и могут быть сопоставлены с универсальными нонсенс-кодонами, а 62 элемента — с группами кодирующих триплетов. Тем самым каждому кодону задается термин пространственной определенности — направление от центра икосаэдра к центру его грани, к центру ребра или к вершине. Каждой из 20 аминокислот также приписывается соответствие одной определенной грани икосаэдра.

Для того чтобы более ясно показать, какие именно элементы икосаэдра соответствуют группам формального кода, можно воспользоваться описанным выше способом построения рядов биномиального распределения. Здесь видно, что аминокислоты группы I должны кодироваться одной гранью, двумя ребрами и одной вершиной, а группы II — одной гранью и одним ребром. Из дополнительных триплетов УУА и УУГ соответствуют вершинам, а каждая из пар, начинающихся с АГ, — вершине и ребру. «Третьему нонсенсу» УГА должно соответствовать ребро, кодон УГГ и АУГ — каждому грань, а прочим кодонам, начинающимся с

Принципы компактности дают возможность точно указать взаиморасположение далеко не всех компонентов кода. Однако соображения, связанные опять-таки с симметрией, позволили нам с большой долей вероятности определить положение всех 20 аминокислот (рис. 5). При таком расположении, в частности, поворотам на 180° (обозначаемым: 180°) соответствуют определенные изменения свойств аминокислотных остатков. Выделяются три парных перехода, причем к каждой паре при-
 мыкает еще по одному непарному переходу: лей $\xrightarrow{180^\circ}$ тре, вал $\xrightarrow{180^\circ}$ сер

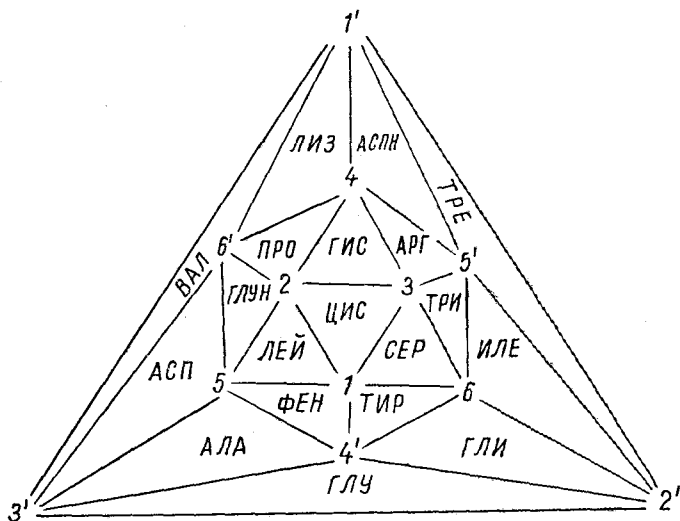


Рис. 5. Наиболее вероятное расположение аминокислот на икосаэдре.

(пара) и мет 180° цис (непарный); иле 180° глун, фен 180° аспн и ала 180° арг; три 180° асп, гис 180° глу и лиз 180° тир. Единственный в своем роде переход гли 180° про соответствует наименьшему изменению свойств аминокислот, так что ось «глицин—пролин» должна играть какую-то особую роль во всей структуре. Сообразно с полученным взаиморасположением в белковой молекуле каждому из аминокислотных остатков также должна быть приписана некоторая область пространства; получающиеся углы должны соответствовать направлениям электромагнитных полей тех или иных радикалов, а значит, определять третичную и четвертичную структуру белковой молекулы, ее биохимическую активность и т. п. Ввиду того, что эти углы отличаются постоянством, они — при большом количестве аминокислот — должны обнаруживаться и в видимых формах, которые поэтому могут сохранять симметрию, свойственную икосаэдру. Эта поворотная симметрия с осями L_5 характерна для живых существ, но вовсе неизвестна среди кристаллов [4]. Установлено также, что вирусы, состоящие из РНК и белка, представляют собой правильные икосаэдры. Можно утверждать, что икосаэдральная форма и L_5 -симметрия являются фундаментальными в организации живого вещества. При таком рассмотрении и сам генетический код представляется не случайным продуктом эволюционных блужданий [5, 6], а закономерным и необходимым следствием исходных принципов, выбранных природой для его осуществления.

В развиваемой концепции тип симметрии является детерминирующим началом для организации кода в его нынешнем виде, а структура кода, в свою очередь, определяет L_5 -симметрию в формах живых существ. Вполне естественно задаться вопросом, почему именно икосаэдр

выбран природой в качестве морфологической основы жизни. Ответ на этот вопрос прост. Икосаэдр, в отличие от тетраэдра, октаэдра и куба, не допускает плотной и равномерной упаковки тождественных симметричных элементов, а это означает, что данная структура не связана с энергетическим минимумом. При температурах выше абсолютного нуля такая система неустойчива; для того чтобы она могла существовать, требуется приток энергии. Рост икосаэдра не может выражаться в бессмысленном повторении одной и той же совокупности атомов пространственной ячейки в различных направлениях (как это происходит при росте неживых кристаллов), а должен сочетаться с изменениями размеров самих ячеек, или должно происходить образование новых, отделенных от первой, ячеек, что дает в зачаточном виде морфологическую дифференциацию или размножение. Нетрудно заметить, что перечисленные достоинства икосаэдра обычно фигурируют в попытках описаний или определений жизни.

В связи с тем, что существующий генетический код однозначно связан с икосаэдральной структурой, нам довольно трудно представить себе возможность его эволюции или его постепенное, а не одноактное возникновение. Столь же трудно представить себе подобную структуру, возникшую на какой-то иной, неуглеродной, химической основе.

* * *

Существенное отличие живых форм от неживых проявляется в собственных им типах симметрии. Если для неживых форм — это L_3 , L_4 и L_6 -симметрия (тетраэдр, октаэдр и куб как фундаментальные формы), то для живых форм — это L_5 -симметрия и икосаэдр как фундаментальная форма.

Структура генетического кода, — т. е. наличие 20 аминокислот, кодируемых определенными группами триплетов РНК, вполне объясняется необходимостью получить икосаэдр как фундаментальную форму для организации живой материи.

Основные положения этой статьи кратко изложены в работе [7].

ЛИТЕРАТУРА

1. G. Gamow, Kgl. Danske Videnskab Selskab. Biol. Modd., 22, 3, 1—13, 1954.
2. F. C. H. Grick, J. S. Griffith, L. K. Orgel, Proc. Natl. Acad. Sci., 43, 416—421, 1957.
3. R. E. Marshall, C. T. Caskey, M. Nirenberg, Science, 155, 3764, 1967.
4. Г. Вейль, Симметрия, М., «Наука», 1968.
5. G. Woese, J. Mol. Biol., 43, 1, 235—240, 1969.
6. F. C. H. Grick, J. Mol. Biol., 38, 3, 367—379, 1968.
7. А. Г. Волохонский, Генетический код и симметрия. Тез. к совещанию «Симметрия в природе», Л., 1971.

Лаборатория Невского
рыбкомбината, Ленинград

Поступила
13.IV 1971 г.

ON FORMAL STRUCTURE OF GENETIC CODE

A. G. VOLOKHONSKY

Summary

The connection between the groups of RNA triplets corresponding to definite amino acids and the rules of constructing the binomial distribution 2^6 groups in the Pascale triangle is postulated basing on the analysis of the formal structure of the genetic code. The geometrical interpretation of the revealed relations shows that the icosahedron is an adequate model of the genetic code.