

СКРЫТАЯ СИММЕТРИЯ ГЕНЕТИЧЕСКОГО КОДА

В.В. Щербик, Л.П. Бучацкий

Киевский национальный университет имени Тараса Шевченко

Информационная связь между нуклеотидными последовательностями мРНК и аминокислотными последовательностями белков осуществляется с помощью генетического кода. За исключением незначительных вариаций генетический код универсален, то есть, одинаков для всех живых организмов.

В генетическом коде каждому триплету нуклеотидов мРНК (кодону) соответствует только одна аминокислота. Но не наоборот – генетический код вырожден. Кроме простого вырождения U/C в третьем положении кодона других явных свойств симметрии в генетическом коде не наблюдается.

Поиск закономерностей в генетическом коде ведется весьма интенсивно. Была предложена модель коэволюции [1–5] аминокислот и их соответствия кодонам мРНК. Математические исследования на основе теории групп [6–12] лишь частично объясняют распределение аминокислот по кодонам мРНК. Ключевой идеей таких подходов является спонтанное нарушение исходной симметрии с одновременным расширением числа аминокислот. Причина такого спонтанного нарушения симметрии осталась неизвестной – не ясен источник хаоса.

Много внимания уделяется топологическим основаниям генетического кода [13–16]. Смысл таких построений состоит в том, что топология триплетов нуклеотидов мРНК должна определять алгебру вырожденности аминокислот. Очевидно, что 64 триплета мРНК можно представить как шестимерный гиперкуб [17]. Но получить числа вырождения аминокислот таким способом не удастся. Несколько лучше ситуация с использованием цветных графов [18–19]. Если учесть вырождение U/C в третьем положении кодона, то можно считать, что кодовых комбинаций всего 48. Генетический код возник как фазовый переход в информационном канале преобразования нуклеотидов в аминокислоты. Матрица связей между кодонами – это объемный граф, который можно вложить в поверх-

ность рода = 25 и раскрасить 20-ю красками, число которых равно числу аминокислот.

В работах [20–21] рассмотрен считывающий или рибосомный оператор, который имеет различные собственные значения для аминокислот. Считывающий оператор близок к проекционным операторам, так как различным кодонам может соответствовать одна и та же аминокислота. Недостатком такого подхода является то, что в нем не учитывается возможность подключения ложной аминокислоты к тРНК.

С целью обнаружения кода взаимодействия аминокислот [22] было предложено разделить множество аминокислот на три группы в зависимости от центральной пары кодон-антикодоновых триплетов. Код взаимодействия аминокислот до сих пор не найден, но разбиение аминокислот на независимые группы представляется нам весьма важным. Мы в явной форме используем таблицу генетического кода, построенную на основе кодон-антикодонного спаривания аминокислот.

Квадратичная форма генетического кода

На рис. 1 приведена таблица генетического кода в виде пар аминокислот, которые соответствуют кодон-антикодонному спариванию нуклеотидов мРНК – тРНК. Все кодоны встречаются по одному разу.

<p>Am5*={F, K, E, *, L, Q}</p> <p>1F(UUU) = K(AAA) 1 1F(UUC) = E(GAA) 1 1L(UUA) = *(UAA) 1 1L(UUG) = Q(CAA) 1 1L(CUU) = K(AAG) 1 1L(CUC) = E(GAG) 1 1L(CUA) = *(UAG) 1 1L(CUG) = Q(CAG) 1</p>	<p>Am8*={S, R, G, *, P, W, T, C, A}</p> <p>1S(UCU) = R(AGA) 1 1S(UCC) = G(GGA) 1 1S(UCA) = *(UGA) 1 1S(UCG) = R(CGA) 1 1P(CCU) = R(AGG) 1 1P(CCC) = G(GGG) 1 1P(CCA) = W(UGG) 1 1P(CCG) = R(CGG) 1</p>
<p>Am7={I, N, D, Y, M, H, V}</p> <p>1I(AUU) = N(AAU) 1 1I(AUC) = D(GAU) 1 1I(AUA) = Y(UAU) 1 1M(AUG) = H(CAU) 1 1V(GUU) = N(AAC) 1 1V(GUC) = D(GAC) 1 1V(GUA) = Y(UAC) 1 1V(GUG) = H(CAC) 1</p>	<p>1T(ACU) = S(AGU) 1 1T(ACC) = G(GGU) 1 1T(ACA) = C(UGU) 1 1T(ACG) = R(CGU) 1 1A(GCU) = S(AGC) 1 1A(GCC) = G(GGC) 1 1A(GCA) = C(UGC) 1 1A(GCG) = R(CGC) 1</p>

Рис. 1. Таблица генетического кода с числом изоформ тРНК, равным 1. Три множества аминокислот Am5*, Am7, Am8* образуют независимые семейства.

Кодоны, соответствующие терминаторам образуют формальные пары аминокислот. Семейства пар аминокислот будем представлять квадратичными формами, коэффициенты в которых равны сумме чисел изоформ тРНК антикодонов. Любая квадратичная форма соответствует некоторому типу симметрии. Для семейств аминокислот – это и есть скрытая симметрия генетического кода.

Квадратичные формы для трех множеств аминокислот имеют вид:

$$Am5^* [2+ 2-]: +2FK +2FE +4L^* +4LQ +2LK +2LE.$$

$$Am7 [2+ 2-]: +2IN +2ID +2IY +2MH +2VN +2VD +2VY +2VH.$$

$$Am8^* [3+ 3-]: +4SR +2SG +2S^* +4PR +2PG +2PW +2TS +2TG +2TC +2TR +2AS +2AG +2AC +2AR.$$

В скобках указаны сигнатуры квадратичных форм. Терминаторы в семействах $Am5^*$ и $Am8^*$ можно считать независимыми, так как сумма этих множеств имеет квадратичную форму с сигнатурой $[5+ 5-]$.

У всех организмов, начиная с простейших, количество изоформ антикодонов больше единицы. Статистика антикодонов не влияет на сигнатуру квадратичной формы каждого семейства аминокислот $Am5^*$, $Am7$, $Am8^*$.

На рис. 2 показана таблица генетического кода в виде пар аминокислот, которые соответствуют изоформам тРНК для организма *Danio rerio (Zebrafish)*.

Квадратичные формы для трех множеств аминокислот (Рис.2) имеют вид:

$$Am5^* [2+ 2-]: +575FK +453FE +68L^* +292LQ +1324LK +242LE +279L^* +611LQ.$$

$$Am7 [3+ 3-]: +310IN +10ID +78IY +623MH +1385VN +171VD +468VY +640VH.$$

$$Am8^* [4+ 4-]: +623SR +42SG +227SU +196SR +447PR +281PG +264PW +268PR + 398TS +18TG +312TC +261TR +526AS +830AG +465AC +137AR.$$

С каждой квадратичной формой сигнатуры $[p+ q-]$ связана алгебра Клиффорда $Cl(p, q)$, имеющая $2^{(p+q)}$ элементов. Квадратичную форму $[1+ 1-]$ будем называть 2-интервалом, двухточечным множеством. Для генетического кода с числом изоформ тРНК, равным 1, количество 2-интервалов в семействах $Am5^*$, $Am7$, $Am8^*$ совпадает с числом водородных связей центральной комплементарной пары нуклеотидов кодона-антикодона. Количество 2-интервалов будем называть состоянием семейства аминокислот.

$Am5^* = \{F, K, E, *, L, Q\}$	$Am8U = \{S, R, G, U, P, W, T, C, A\}$
0565F (UUU) = K (AAA) 0010	0264S (UCU) = R (AGA) 0359
0233F (UUC) = E (GAA) 0220	0040S (UCC) = G (GCA) 0002
0007L (UUA) = * (UAA) 0061	0014S (UCA) = U (UGA) 0213
0110L (UUG) = Q (CAA) 0182	0104S (UCG) = R (CGA) 0092
1018L (CUU) = K (AAG) 0306	0101P (CCU) = R (ACG) 0346
0238L (CUC) = E (GAG) 0004	0279P (CCC) = G (GGG) 0002
0009L (CUA) = * (UAG) 0270	0038P (CCA) = W (UGG) 0226
0297L (CUG) = Q (CAG) 0314	0059P (CCG) = R (CCG) 0209
$Am7 = \{I, N, D, Y, M, H, V\}$	
0025I (AUU) = N (AAU) 0285	0014T (ACU) = S (AGU) 0384
0001I (AUC) = D (GAU) 0009	0008T (ACC) = G (GCU) 0010
0005I (AUA) = Y (UAU) 0073	0009T (ACA) = C (UCU) 0303
0011M (AUG) = H (CAU) 0612	0191T (ACG) = R (CCU) 0070
1138V (GUU) = N (AAC) 0247	0436A (GCU) = S (AGC) 0090
0166V (GUC) = D (GAC) 0005	0826A (GCC) = G (GGC) 0004
0225V (GUA) = Y (UAC) 0243	0098A (GCA) = C (UGC) 0367
0405V (GUG) = H (CAC) 0235	0025A (GCG) = R (CCG) 0112

Рис. 2. Таблица генетического кода с количеством изоформ тРНК для организма *Danio rerio (Zebrafish)*. Числовые данные взяты из базы данных *Genomic tRNA Database*. Аминокислота Sec условно обозначена как U.

Ясно, что наличие изоформ тРНК изменяет состояния семейств аминокислот $Am7$ и $Am8^*$. Но влияет ли это на поведение рибосомы при трансляции мРНК – не известно. Алгебры Клиффорда $Cl(1, 1)$, $Cl(2, 2)$, $Cl(3, 3)$, $Cl(4, 4)$ являются скрытыми симметриями генетического кода. Вопрос о том, сохраняются ли эти алгебры аминокислот при формировании молекулы белка остается без ответа.

Граф состояний генетического кода

На рис. 3 представлен граф генетического кода со скрытыми функциями переходов между состояниями различных семейств аминокислот. Количество функции перехода состояний определяется (по Кантору) как множество всех отображений одного множества в другое. Количество функций перехода состояний должно быть равно 64, то есть количеству кодонов генетического кода. Только в таком случае можно говорить о скрытой симметрии генетического кода. Граф на рис. 3 показывает скрытые, но обратимые переходы между состояниями семейств аминокислот. Переходов внутри семейства аминокислот нет. Исключением является лишь замена для триплета UGA - с кодона терминатора на кодон селеноцистеина.

Отметим, что переход $\textcircled{1} \xrightarrow{-1} \textcircled{1}$ обозначает переход пар аминокислот $SU^* \leftrightarrow S^*$, что означает “удлинение единицы”.

Отметим, что графов с необратимыми переходами между состояниями семейств аминокислот можно построить много. Например,

переходы $1 \leftrightarrow 1$ и $2 \leftrightarrow 2$ можно объединить в один переход $1 \leftrightarrow 4$. Конечно, переход $3 \rightarrow 4$ в семействе $Am8^*$ дает все 64 состояния генетического кода, но является ли он скрытым? Возможно, этот переход связан с wobble-гипотезой: комплементарные пары РНК не имеют 4 водородные связи. Важным свойством обратимых скрытых переходов между семействами аминокислот является удвоение состояний генетического кода.

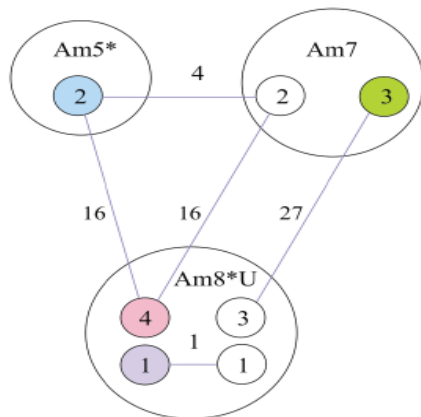


Рис. 3. Граф состояний генетического кода. Состояния множеств аминокислот определяются количеством 2-интервалов. Над линиями связей указано количество функций перехода состояний. Бесцветные состояния соответствуют семействам аминокислот с числом изоформ тРНК, равным 1.

И, все-таки, остается сложной проблема переноса алгебр Клиффорда семейств аминокислот на симметрию аминокислот при образовании молекулы белка.

Интерпретация графа генетического кода

Авторы статей, посвященных структуре генетического кода, основной упор делали на связи триплета нуклеотидов мРНК с аминокислотами. Но ясно, что генетический код является продолжением репликации и транскрипции ДНК. Скрытая симметрия генетического кода связана с этими важными биологическими процессами.

Рассмотрим репликацию ДНК: на графе состояний генетического кода есть треугольный образ $4 \leftrightarrow (2 \leftrightarrow 2)$.

Представим ДНК как комплементарный вектор, состоящий из базисных векторов e_i , занумерованных в направлении 5'3' в (+)-ДНК и в направлении 3'5' в (-)-ДНК следующим образом:

(+)-ДНК:	A(+e1)	T(+e2)	C(+e3)	G(+e4)
(-)-ДНК:	T(-e1)	A(-e2)	G(-e3)	C(-e4)

где (+)-ДНК, (-)-ДНК – комплементарные цепи ДНК.

Если положить $A(+e1) = T(-e1)$, $T(+e2) = A(-e2)$, $C(+e3) = G(-e3)$, $G(+e4) = C(-e4)$, а также для всех нуклеотидов: $N(+e1) = +Ne1$, $N(-e1) = -Ne1$, то комплементарные цепи будут тождественными. Если $e_i e_i = 0$ и $e_i e_j = -e_j e_i$, то комплементарные цепи ортогональны. Векторы e_i подчиняются алгебре Грассмана.

Если в (+)-цепи ДНК выбрать произвольно направления базисных векторов, то это даст 16 вариантов их размещения, то есть переход $4 \leftrightarrow 2$. При этом (-)-цепь ДНК согласуется с (+)-цепью изменением ориентации векторов (переход $2 \leftrightarrow 4$). Аналогично можно рассмотреть (-)-цепь ДНК.

Еще 4 функции перехода можно получить, если считать, что могут быть варианты спариваний нуклеотидов ДНК типа

(+)-ДНК:	A(+e1)	A(-e2)	A(+e3)	A(-e4)
(-)-ДНК:	T(+e1)	T(-e2)	T(-e3)	T(+e4)

Сложнее получить $(27 + 1)$ функций перехода. Эти функции соответствуют чтению нуклеотидов ДНК, то есть транскрипции ДНК. Один нуклеотид является ведущим и определяет способ чтения, например, А. И так $T(-e1) \rightarrow T(-e1') \rightarrow U(+e1')$, где штрих обозначает переход к дуальному базису. Это связано с Лоренц-интервалами чтения нуклеотидов. 27 Функций перехода мы получим, если в дуальном базисе каждому триpletу AGC будет соответствовать любой триplet из этих нуклеотидов.

Результатом транскрипции (-)-цепи ДНК является (+)-цепь РНК с формальной метрикой Лоренца:

(-)-ДНК:	T(-e1)	A(-e2)	G(-e3)	C(-e4)
(+)-РНК:	A(+e1')	T(-e2')	C(+e3')	G(+e4')

Выполняются соотношения:

$$A(+e1') = T(-e1) = A(+e1), \quad C(+e3') = G(-e3) = C(+e3), \\ G(+e4') = C(-e4) = G(+e4), \quad T(-e2') = A(-e2) = T(+e2).$$

Это значит, что для нуклеотида Т в (+)-цепи РНК необходимо инвертировать направление базиса. Экспериментально установлено, что $T(-e_2') = U(+e_2')$ и направление базиса для нуклеотида U совпадает с направлением базисов для других нуклеотидов.

Переход $T \rightarrow U$ и $-e2' \rightarrow +e2'$ соответствует отображению $1 \leftrightarrow 1$ на графе состояний генетического кода.

Уточним алгебру транскрипции ДНК. Дело в том, что замена $T \rightarrow U$ выглядит формальной. В мРНК вводится еще один, локальный базис алгебры Грассмана с образующими a, a^* : $aa^* + a^*a = 0$.

В мРНК устанавливается соответствие между кодом РНК и расположением базисных элементов алгебры Грассмана. Виртуальная РНК с нуклеотидами ДНК:

	+A	+C	+G	+T
(+)-РНК:	+A00	+A10a*	+A01a	+A11a*a
(-)-РНК:	-A11	-A01a	-A10a*	-A00aa*
	-T	-G	-C	-A

Изменение базиса виртуальной мРНК с (+)-цепи на (-)-цепь требует сопряжения в алгебре Грассмана, то есть $a \rightarrow a^*, a^* \rightarrow a, aa^* \rightarrow a^*a$. Этот базис не сохраняется в мРНК.

РНК – прочитанная ДНК – интеграл в алгебре Грассмана:

$$(+) \text{-мРНК:} \quad K_{00} \quad K_{10a^*} \quad K_{01a} \quad K_{11a^*a}$$

Прочитанная РНК: $A(a^*, a) = e^{a^*a} K(a^*, a)$; где $K(a^*, a)$ – код РНК. Поли-А – это постоянная интегрирования, по смыслу не совсем совпадающая с произвольной постоянной интегрирования в интеграле Римана.

В алгебре Грассмана множитель интегрирования e^{a^*a} , возможно, соответствует эпю мРНК. Связь РНК с ДНК определяется формулами интегрирования

$$K_{00} = A_{00} = A, \quad K_{10} = A_{10} = C, \quad K_{01} = A_{01} = G, \quad K_{11} = U = A_{11} - A_{00} = T - A$$

и ассоциируется с формальной метрикой Лоренца.

Формулы интегрирования в алгебре Грассмана приведены в известной книге по теории поля [23].

Базис мРНК до сплайсинга:

$$(+) \text{-мРНК:} \quad +A \quad +Ca^* \quad +Ga \quad +Ua^*a \\ (-) \text{-мРНК:} \quad -Uaa^* \quad -Ga \quad -Ca^* \quad -A$$

Интроны при сплайсинге мРНК имеют, в основном, форму лас-со [24], которая, по нашему мнению, является образом сахарофосфатного базиса мРНК. Значит, при сплайсинге из мРНК вырезается информация о сахарофосфатном базисе.

Цепи мРНК будут ортогональны, если их произведение равно нулю: $\{-AA\} + \{-AG -GA\}a + \{-AC -CA\}a^* + \{+AU -UA +GC -CG\}a^*a = 0$.

Нуклеотиды мРНК, как операторы, удовлетворяют соотношениям: $AA = 0, AG = -GA, AC = -CA, AU = UA, GC = CG$

и образуют некоторую супералгебру. Этот базис можно считать автокаталитическим в мРНК – сохраняется не всегда.

Для мРНК после сплайсинга можно ввести локальный 4-базис алгебры Грассмана, зависящий от пространственных переменных вдоль цепи:

$$(+) \text{-мРНК:} \quad +Ae1 \quad +Ce2 \quad +Ge3 \quad +Ue4 \\ (-) \text{-мРНК:} \quad -Ue1 \quad -Ge2 \quad -Ce3 \quad -Ae4$$

с тождествами $A = -U, C = -G, G = -C, U = -A$. Новый базис алгебры Грассмана является окончательным и сохраняется для мРНК.

Таким образом, мы отобразили все множество функций перехода на графе состояний генетического кода на множество преобразований генетической информации при репликации и транскрипции ДНК.

Выводы

Генетический код подразделяется на три семейства со скрытой симметрией – алгебрами Клиффорда. Структура генетического кода такова, что имеет скрытую связь с процессами репликации и транскрипции ДНК. Базисом нуклеотидов ДНК является линейно-упорядоченный глобальный базис алгебры Грассмана. В РНК для информационных нуклеотидов присутствует локальный базис алгебры Грассмана, причем этот базис изменяется при преобразовании РНК. При сплайсинге из первичного транскрипта РНК вырезаются интроны, возможно кодирующие сахарофосфатный базис РНК.

1. Tze-Fei Wong J. A Co-evolution theory of the origion of the genetic code / J. Tze-Fei Wong // Proc. Nat.Acad. Sci. USA. – 1975. – Vol. 72. – P. 1909-1912.
2. Tze-Fei Wong J. Coevolution theory of the genetic code at age thirty / J. Tze-Fei Wong // BioEssays. – 2005. – Vol. 27. – P. 416-425.
3. Di Giulio M. The coevolution theory of the origin of the genetic code / M. Di Giulio // Physics Life Rev. – 2004. – Vol. 1. – P. 128-137.
4. Di Giulio M. An extension of the coevolution theory of the origin of the genetic code / M. Di Giulio // Biol. Direct. – 2008. – Vol. 3. – P. 37.
5. Di Giulio M. The origin of the genetic code: theories and their relationships, a review / M. Di Giulio // BioSystems. – 2005. – Vol. 80. – P. 175-184.
6. Hornos J.E.M. Algebraic Model for the Evolution of the Genetic Code / J.E.M. Hornos, Y. M. M. Hornos // Phys. Rev. Lett. – 1993. – Vol. 71. – P. 4401- 4404.
7. Bashford J.D. A supersymmetric model for the evolution of the genetic code / J.D. Bashford, I. Tsohantjis, P.D. Jarvis // Proc. Natl. Acad. Sci. USA. – 1998. – Vol. 95. – P. 987-992.
8. Hornos J.E.M. Symmetry and symmetry breaking: an algebraic approach to the genetic code / J.E.M. Hornos, Y.M.M. Hornos, M. Forger // Int. J. Mod. Physics B. – 1999. – Vol. 13. – P. 2795-2885.
9. Antoneli F. On amino acid and codon assignment in algebraic models for the genetic code / F. Antoneli, M. Forger, P.A. Gaviria // Int. J. Mod. Physics B. – 2010. – Vol. 24. – P. 435-463.
10. Négadi T. Symmetry Groups for the Rumer–Konopel’chenko–Shcherbak “Bisections” of the Genetic Code and Applications / T. Négadi // Internet Electron. J. Mol. Des. – 2004. – Vol. 3. – P. 247-270.
11. Forger M. Lie Superalgebras and the multiplet structure of the genetic code i: codon representations / M. Forger, S. Sachse // J. Math. Phys. – 2000. – Vol. 41. – P. 5407-5422.
12. Forger M. Lie Superalgebras and the multiplet structure of the genetic code ii: branching rules / M. Forger, S. Sachse // J. Math. Phys. – 2000. – Vol. 41. – P. 5423-5444.
13. Karasev V. A. Topological nature of the genetic code / V. A. Karasev, V. E. Stefanov // J. Theor. Biol. – 2001. – V209 - P.303-317.
14. Sánchez R. The Genetic Code Boolean Lattice / R. Sánchez, E. Morgado, R. Grau // MATCH Commun. Math. Comput. Chem. – 2004. – Vol. 52. – P. 29-46.
15. Castro-Chavez F. A tetrahedral representation of the genetic code emphasizing aspects of symmetry / F. Castro-Chavez // BIO-Complexity. – 2012. – Vol. 2. – P. 1-6.
16. Yang C.M. The naturally designed spherical symmetry in the genetic code / C.M. Yang // arxiv.org/abs/q-bio.BM/0309014 - 2003.
17. Jimenez-Montano M.A. The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions in vivo and in vitro /

- M.A. Jimenez-Montano, R. de la Mora-Basanez, T. Poschel // BioSystems. – 1996. – Vol. 39. – P. 117-125.
18. Tlusty T. A model for the emergence of the genetic code as a transition in a noisy information channel / T. Tlusty // J. Theor. Biol. – 2007. – Vol. 249 - P. 331-342.
 19. Tlusty T. A relation between the multiplicity of the second eigenvalue of a graph Laplacian, Courant’s nodal line theorem and the substantial dimension of tight polyhedral surfaces / T. Tlusty // Elec. J. Linear Algebra. – 2007. – Vol. 16. – P. 315-324.
 20. Frappat L. A crystal base for the genetic code / L. Frappat, A. Sciarrino, P. Sorba // Phys. Lett. – 1998. – Vol. 250. – P. 214.
 21. Франтам Л. Квантовые группы и генетический код / Л. Франтам, П. Сорба, А. Сиаррино // ТМФ. – 2001. – Т. 128. – С. 27-42.
 22. Чипенс Г.И. Скрытая симметрия генетического кода и законы взаимодействия аминокислот / Г.И. Чипенс // Биоорг. химия. – 1991. – Т. 17. – С. 1335-1346.
 23. Славнов А. А. Введение в квантовую теорию калибровочных полей / А.А. Славнов, Л.Д. Фаддеев. – М.: Наука, 1978. – 240 с.
 24. Сингер М. Гены и геномы. Т. 2 / М. Сингер, П. Берг. – М.: Мир, 1998. – 391 с.

Резюме

Щербик В.В., Бучацкий Л.П. Скрытая симметрия генетического кода.

Пары аминокислот генетического кода разделены на три независимые семейства, которые представлены квадратичными формами. Коэффициенты квадратичных форм равны сумме чисел изоформ тРНК антикодонов спаренных аминокислот. Алгебры Клиффорда $Cl(1, 1)$, $Cl(2, 2)$, $Cl(3, 3)$, $Cl(4, 4)$ являются скрытыми симметриями семейств аминокислот генетического кода. Предложен граф скрытых состояний генетического кода, в котором все множество обратимых функций перехода между семействами аминокислот отображается на множество преобразований генетической информации при репликации и транскрипции ДНК. Для упорядочения нуклеотидов ДНК предложен постоянный линейный глобальный базис алгебры Грассмана. РНК определяется как интеграл в алгебре Грассмана. В РНК для информационных нуклеотидов присутствует локальный базис алгебры Грассмана, причем этот базис изменяется трижды при преобразовании РНК. При сплайсинге из первичного транскрипта РНК вырезаются интроны, вероятно кодирующие сахарофосфатный базис РНК.

Ключевые слова: генетический код, скрытая симметрия, алгебра Клиффорда, алгебра Грассмана.

Резюме

Щербик В.В., Бучацкий Л.П. Прихована симетрія генетичного коду.

Пари амінокислот генетичного коду розділені на три незалежні сімейства, які представлені квадратичними формами. Коефіцієнти ква-

дратичних форм дорівнюють сумі чисел ізоформ тРНК антикодонів спарених амінокислот. Алгебри Кліффорда $Cl(1, 1)$, $Cl(2, 2)$, $Cl(3, 3)$, $Cl(4, 4)$ є прихованими симетріями сімейств амінокислот генетичного коду. Запропоновано граф прихованих станів генетичного коду, в якому вся множина оборотних функцій переходу між сімействами амінокислот відображається на множину перетворень генетичної інформації при реплікації і транскрипції ДНК. Для впорядкування нуклеотидів ДНК запропонований постійний лінійний глобальний базис алгебри Грассмана. РНК визначається як інтеграл в алгебрі Грассмана. У РНК для інформаційних нуклеотидів присутній локальний базис алгебри Грассмана, причому цей базис змінюється тричі при перетворенні РНК. Під час сплайсингу з первинного транскрипту РНК виризаються інтрони, які ймовірно кодуєть сахарофосфатний базис РНК.

Ключові слова: генетичний код, прихована симетрія, алгебра Кліффорда, алгебра Грассмана.

Summary

Stcherbic V.V., Buchatsky L.P. *Hidden symmetry of the genetic code.*

Pairs of amino acids of the genetic code are divided into three separate families, represented by quadratic forms. The coefficients of quadratic forms equal the sum of the numbers of tRNA anticodons isoforms paired amino acids. Clifford algebras $Cl(1, 1)$, $Cl(2, 2)$, $Cl(3, 3)$, $Cl(4, 4)$ are hidden symmetries of the families of amino acids of the genetic code. We propose a graph of hidden states of the genetic code, in which all set of reversible transition functions between the families of amino acids is mapped on the set of transformations of the genetic information in DNA replication and transcription. To order the DNA nucleotides is proposed constant linear global basis of Grassmann algebra. RNA is defined as the integral in the Grassmann algebra. In RNA for information nucleotides is present a local basis of the Grassmann algebra, and this basis is changed three times by the transformation of the RNA. When splicing, introns, are cut from the primary RNA transcript probably encode sugar-phosphate base of RNA.

Key words: the genetic code, hidden symmetry, Clifford algebra, Grassmann algebra.

Рецензент: д.біол.н., проф. С.М. Смірнов

ЕКОЛОГІЧНА І КЛІНІЧНА ІМУНОЛОГІЯ ТА ІМУНОРЕАБІЛІТАЦІЯ