



И.В. СЕРГИЕНКО, А.М. ГУПАЛ, А.А. ВАГИС

УДК 519.217.2

СИММЕТРИЧНЫЙ КОД И ГЕНЕТИЧЕСКИЕ МУТАЦИИ

Аннотация. С учетом свойств симметрии в ДНК построен код, симметричный относительно полярности аминокислот при мутациях в нуклеотидах. Проведено сравнение стандартного кода со случайно сгенерированными кодами. Исследована помехоустойчивость генетических кодов относительно полярности аминокислот. На основе баз данных генетических заболеваний показано, что симметричный код в большинстве случаев исправляет нарушение полярности при мутациях.

Ключевые слова: генетический код, мутации, помехоустойчивость, аминокислота, нуклеотид, кодон, генетический алгоритм.

ВВЕДЕНИЕ

Универсальный генетический код, согласно которому в живых организмах происходит синтез белков на основе информации, содержащейся в ДНК, был открыт в 60-х годах XX века [1]. Одним из основных факторов, которые могли влиять на эволюцию кода, являлось повышение его устойчивости по отношению к мутациям — искажениям генетической информации, возникающим в клетках любого организма. Мутации происходят из-за ошибок в репликации в ДНК или разрушительных воздействий мутагенов, таких как химические агенты и радиация, которые реагируют с ДНК и изменяют структуру отдельных нуклеотидов (оснований). В многоклеточном организме только события, происходящие в зародышевых клетках, сопряжены с эволюцией генома. Изменения в геномах соматических клеток не влияют на эволюцию, но они приобретают биологическую значимость, если порождают порочный фенотип, который губительно сказывается на здоровье организма.

Определение последовательности нуклеотидов первого расшифрованного генома было выполнено в проекте «Геном Человека», который завершился в 2001 году и стоил 3 млрд долларов. Отмечалось, что геномы людей идентичны на 99,9%, оставшиеся 0,1% генома отвечают за индивидуальные признаки, такие как цвет глаз, способность усваивать лактозу либо склонность к определенным заболеваниям. Следующей целью сообщества, состоящего из академических и коммерческих организаций, стало повышение эффективности методов расшифровки геномов, для того чтобы снизить цену за эту процедуру до тысячи долларов за геном. С тех пор скорость и точность расшифровки геномов постоянно повышается, а стоимость неуклонно снижается. В настоящий период время расшифровки генома человека составляет примерно несколько месяцев, а цена — тысячу долларов.

Известно, что замена аминокислоты и нарушение ее полярности в белках вредны, поскольку чередованием полярных и неполярных аминокислот определяется их пространственная структура. В работах [1, 2] предполагалось, что в результате эволюции отобран наиболее помехоустойчивый вариант, т.е. стандарт-

© И.В. Сергиенко, А.М. Гупал, А.А. Вагис, 2016

ный генетический код живых организмов (табл. 1) сводит к минимуму замены между полярными (п) и неполярными (н) аминокислотами.

Аминокислоты, слагающие белок, делятся на полярные (10 штук) и неполярные (10 штук). Полярные: аспарагиновая кислота (D), глутаминовая кислота (E), гистидин (H), лизин (K), аспарагин (N), глутамин (Q), аргинин (R), серин (S), треонин (T), тирозин (Y). Неполярные: аланин (A), цистеин (C), фенилаланин (F), глицин (G), изолейцин (I), лейцин (L), метионин (M), пролин (P), валин (V), триптофан (W). Аминокислоты отличаются по взаимодействию с молекулами воды. Неполярные аминокислоты более гидрофобны, они отталкивают молекулы воды и слипаются одна с другой, полярные — гидрофильны и взаимодействуют с водой; в результате образуется сложная форма, соответствующая выполняемой белком функции. Если при мутации полярный остаток в белке сменится на неполярный (или наоборот), то форма молекулы может измениться настолько, что белок не сможет выполнять своей функции.

ПОМЕХОУСТОЙЧИВОСТЬ КОДА

Исследуем поведение помехоустойчивости кода при изменении одного из оснований (нуклеотидов) в триплете, который кодирует аминокислоту. Каждый кодон допускает девять однократных замен, число кодирующих аминокислот триплетов равно 61. Поэтому количество возможных замен нуклеотидов для всех кодонов — 549; из них 23 замены нуклеотидов приводят к появлению стоп-кодонов, 134 замены не меняют кодируемой аминокислоты, 230 замен не меняют класса кодируемой аминокислоты. Для стандартного кода возможны 526 замещений, в которых кодон до и после мутации не является одним из трех стоп-кодонов, из них 364 замещения не меняют полярности, т.е. его помехоустойчивость составляет $364/526 = 69,20\%$.

Стандартный генетический код сопоставляет каждой тройке последовательных нуклеотидов (кодону) гена одну из перечисленных аминокислот (табл. 1). При этом трем кодонам (TAA, TAG и TGA — так называемые стоп-кодоны) не соответствует ни одна аминокислота — они сигнализируют об останове процесса синтеза белка. Поскольку оставшийся 61 кодон кодирует 20 аминокислот, стандартный

Таблица 1. Стандартный генетический код живых организмов

Первое основание	Второе основание				Третье основание
	T	C	A	G	
T	Фенилаланин (F), н Фенилаланин (F), н Лейцин (L), н Лейцин (L), н	Серин (S), п Серин (S), п Серин (S), п Серин (S), п	Тирозин (Y), п Тирозин (Y), п Стоп Стоп	Цистеин (C), н Цистеин (C), н Стоп Триптофан (W), н	T C A G
C	Лейцин (L), н Лейцин (L), н Лейцин (L), н Лейцин (L), н	Пролин (P), н Пролин (P), н Пролин (P), н Пролин (P), н	Гистидин (H), п Гистидин (H), п Глутамин (Q), п Глутамин (Q), п	Аргинин (R), п Аргинин (R), п Аргинин (R), п Аргинин (R), п	T C A G
A	Изолейцин (I), н Изолейцин (I), н Изолейцин (I), н Метионин (M), н	Треонин (T), п Треонин (T), п Треонин (T), п Треонин (T), п	Аспарагин (N), п Аспарагин (N), п Лизин (K), п Лизин (K), п	Серин (S), п Серин (S), п Аргинин (R), п Аргинин (R), п	T C A G
G	Валин (V), н Валин (V), н Валин (V), н Валин (V), н	Аланин (A), н Аланин (A), н Аланин (A), н Аланин (A), н	Аспарагиновая кислота (D), п Аспарагиновая кислота (D), п Глутаминовая кислота (E), п Глутаминовая кислота (E), п	Глицин (G), н Глицин (G), н Глицин (G), н Глицин (G), н	T C A G

код характеризуется вырожденностью: всем аминокислотам, кроме триптофана и метионина, соответствует более чем один кодон. Вырожденность кода снижает возможные нарушения, вызванные мутациями. Некоторые аминокислоты, такие как валин, определяются четверкой кодонов, поэтому при любой замене третьего нуклеотида в кодоне изменения аминокислоты в белке не происходит.

Компьютерные расчеты показали, что стандартный генетический код не является оптимальным относительно помехоустойчивости даже локально, поскольку имеются парные перестановки кодонов, а также изменения в типе аминокислоты, повышающие его помехоустойчивость [3, 4]. Существует 40 единичных изменений в типе аминокислоты, увеличивающих помехоустойчивость кода. Из них 38 случаев составляют замены аминокислоты на стоп-кодон. В двух случаях стоп-кодон заменяется полярной аминокислотой. Таким образом, не существует единичных изменений, улучшающих помехоустойчивость, которые бы сохраняли при этом три стоп-кодона. Имеется 51 парная перестановка аминокислот со стоп-кодоном, повышающая помехоустойчивость.

СИММЕТРИЯ В ДНК И СИММЕТРИЧНЫЙ КОД

В работе [5] исследована симметрия и свойства записи генетической информации в ДНК. Соотношения симметрии выводятся по одной нити в ДНК: количество коротких последовательностей нуклеотидов примерно равно количеству обратных комплементарных последовательностей. Для кодонов симметрия по одной нити ДНК записывается в виде

$$n(ijk) = n(\overline{kji}). \quad (1)$$

где $n(ijk)$ — число троек оснований (ijk) , $i, j, k \in \{A, C, G, T\}$, (\overline{kji}) — антикодон кодона (ijk) . В коде, симметричном относительно полярности, кодон (ijk) кодирует полярную аминокислоту, а антикодон (\overline{kji}) — неполярную, либо наоборот (без учета стоп-кодонов).

Для четных кодов симметричный код, понимаемый в таком смысле, построить нельзя. Для пар нуклеотидов выполняются соотношения

$$n(ij) = n(\overline{ji}), \quad (2)$$

где $i, j \in \{A, C, G, T\}$, $\overline{A} = T$, $\overline{C} = G$, $\overline{T} = A$, $\overline{G} = C$. Заметим, что пары AT, TA, CG и GC не содержатся в (2), поскольку они приводят к тавтологии. Поэтому $n(AT) \neq n(TA)$, $n(CG) \neq n(GC)$. Для четырехбуквенного кода таких исключений 16. Таким образом, симметрия в ДНК дает возможность построить код, симметричный относительно полярности аминокислот.

Анализируя табл. 1, отметим, что в столбце T содержатся неполярные аминокислоты, в столбце A — полярные, в столбцах C и G в двух клетках таблицы — полярные и неполярные. Поэтому можно считать, что стандартный код на 3/4 соответствует симметричному коду.

В табл. 2 представлен оптимальный помехоустойчивый симметричный код [3, 4], в котором три стоп-кодона и число триплетов, определяющих полярные и неполярные аминокислоты, такое же, как в стандартном коде: 30 триплетов задают полярные аминокислоты, 31 — неполярные. Этот код построен на основе восьми парных перестановок из стандартного кода, жирным шрифтом выделены изменения по сравнению со стандартным кодом. Помехоустойчивость кода составляет $408/524 = 77,86\%$, при этом 138 замен в основаниях сохраняют ту же аминокислоту. Как видно, симметричный код имеет достаточно простую структуру и его можно получить без выполнения компьютерных расчетов: четыре парные перестановки заменяют треонин на глицин и четыре перестановки — серин на стоп-кодон, триптофан и цистеин, цистеин. Этот код образуется из стандартного кода с помощью минимального числа перестановок.

Помехоустойчивость симметричного кода немного превышает 7/9 (за счет наличия стоп-кодонов), поскольку не более двух мутаций из девяти нарушают полярность аминокислот. Первые два столбца полученного кода содержат неполярные

Таблица 2. Симметричный код

Первое основание	Второе основание				Третье основание
	T	C	A	G	
T	Фенилаланин (F), н	Стоп Триптофан (W), н Цистеин (C), н Цистеин (C), н	Тирозин (Y), п	Серин (S), п	T
	Фенилаланин (F), н		Тирозин (Y), п	Серин (S), п	C
	Лейцин (L), н		Стоп	Серин (S), п	A
	Лейцин (L), н		Стоп	Серин (S), п	G
C	Лейцин (L), н	Пролин (P), н	Гистидин (H), п	Аргинин (R), п	T
	Лейцин (L), н	Пролин (P), н	Гистидин (H), п	Аргинин (R), п	C
	Лейцин (L), н	Пролин (P), н	Глутамин (Q), п	Аргинин (R), п	A
	Лейцин (L), н	Пролин (P), н	Глутамин (Q), п	Аргинин (R), п	G
A	Изолейцин (I), н	Глицин (G), н	Аспарагин (N), п	Серин (S), п	T
	Изолейцин (I), н	Глицин (G), н	Аспарагин (N), п	Серин (S), п	C
	Изолейцин (I), н	Глицин (G), н	Лизин (K), п	Аргинин (R), п	A
	Метионин (M), н	Глицин (G), н	Лизин (K), п	Аргинин (R), п	G
G	Валин (V), н	Аланин (A), н	Аспарагиновая кислота (D), п	Треонин (T), п	T
	Валин (V), н	Аланин (A), н	Аспарагиновая кислота (D), п	Треонин (T), п	C
	Валин (V), н	Аланин (A), н	Глутаминовая кислота (E), п	Треонин (T), п	A
	Валин (V), н	Аланин (A), н	Глутаминовая кислота (E), п	Треонин (T), п	G

аминокислоты, третий и четвертый — полярные, т.е. если кодон (ijk) определяет полярную аминокислоту, то кодон (kji) — неполярную, и наоборот. С помощью парных перестановок аминокислот нельзя повысить помехоустойчивость симметричного кода, а других способов ее повышения не существует.

Например, если в клетке с первыми двумя нуклеотидами TC содержатся аминокислоты цистеин, цистеин, стоп-кодон и триптофан, то получается код, имеющий помехоустойчивость $408/526 = 77,57\%$.

СЛУЧАЙНЫЕ ГЕНЕТИЧЕСКИЕ КОДЫ. ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ

Первым этапом вычислительного эксперимента была оценка помехоустойчивости стандартного генетического кода по сравнению со случайными генетическими кодами, принадлежащими к одному из двух классов: коды, являющиеся «перестановками» стандартного кода, и коды, имеющие примерно одинаковое со стандартным кодом распределение характеристик. Количество стоп-кодонов в кодах этого класса изменялось от двух до четырех, количество кодонов полярных и неполярных аминокислот находилось в диапазоне $27 \leq n \leq 34$.

Результаты усредненных значений и среднеквадратичных отклонений для 10^5 кодов следующие: помехоустойчивость перестановочных кодов — $49,19 \pm 2,91$, помехоустойчивость близких кодов — $50,21 \pm 3,10$. Как видно, стандартный генетический код устойчив к помехам в существенно большей степени, чем генетический код в среднем, что свидетельствует об обоснованности гипотезы о помехоустойчивости как одного из факторов, влияющих на эволюцию кода.

В генетических алгоритмах в качестве «организмов» популяции использовались не сами генетические коды (таблицы, ставящие каждому кодону в соответствие одну из 20 аминокислот или стоп-сигнал), а таблицы полярности, ставящие в соответствие каждому кодону один из трех типов информации: полярную аминокислоту, неполярную аминокислоту или стоп-кодон.

С помощью таблиц полярности можно исключить избыточную информацию, поскольку вариантов генетических кодов существует примерно 21^{64} , а таб-

лиц полярности — всего 3^{64} . Поэтому, используя генетические алгоритмы, удалось достаточно экономно и быстро построить оптимальные варианты помехоустойчивых генетических кодов, которые отличаются от стандартного кода числом стоп-кодона и количеством полярных и неполярных триплетов.

По аналогии со стандартным кодом произвольный генетический код можно рассматривать как функцию $f: C \rightarrow S$, отображающую множество кодонов $C \equiv \{A, C, G, T\}^3$ на множество $S = \{0, 1, stop\}$.

Для оптимизации помехоустойчивости на множествах перестановок стандартного кода и «близких» к нему кодов использовался генетический алгоритм.

Шаг 1. Задать начальное поколение, состоящее из N_0 случайных перестановок стандартного генетического кода.

Шаг 2. Для каждого из T поколений $F_t, t = 1, \dots, T$, выполнить шаги 3–6.

Шаг 3. Добавить в текущее поколение F_t результаты скрещивания каждого кода $f_i \in F_t$ с определенным количеством N_c других кодов из F_t , выбранных случайным образом.

Шаг 4. Добавить в поколение F_t заданное количество N_m мутаций каждого кода $f_i \in F_t$.

Шаг 5. Удалить из поколения коды, не принадлежащие к рассматриваемому классу.

Шаг 6. Перевести в следующее поколение не более L кодов, обладающих наибольшей помехоустойчивостью.

Пусть St — стандартный генетический код, $P_{St}(s), s \in S$, — вероятности фиксации в нем соответствующих состояний:

$$P_{St}(s) = \frac{1}{64} |St^{-1}(s)|.$$

Понятие мутации генетического кода $m(f), f: C \rightarrow S$, определено следующим образом:

$$m(f)(u) = \begin{cases} f(u) & \text{с вероятностью } (1-p), \\ x \in S & \text{с вероятностью } pP_{St}(x)/(1-P_{St}(f(u))). \end{cases}$$

Таким образом, при мутации значение функции кода, соответствующее каждому из 64 кодонов, может с определенной вероятностью p измениться на другое возможное значение, причем распределение вероятностей перехода соответствует распределению значений в стандартном генетическом коде.

Результат скрещивания двух кодов, f_i и f_j , определялся как

$$(f_i \otimes f_j)(u) = \begin{cases} f_i(u) & \text{с вероятностью } 1/2, \\ f_j(u) & \text{с вероятностью } 1/2. \end{cases}$$

Использовались следующие параметры алгоритма: количество поколений $T = 50\,000$; размер начального поколения $N_0 = 50$; максимальный размер поколения $L = 250$; число скрещиваний и мутаций для каждого кода $N_c = 4$ и $N_m = 2$ соответственно; вероятность мутации в отдельном кодоне $p = 0,1$.

С помощью генетических алгоритмов построены различные варианты наиболее помехоустойчивых относительно полярности кодов, в том числе получен оптимальный симметричный код в табл. 2, а также коды, отличающиеся от стандартного кода числом стоп-кодона и количеством триплетов, определяющих полярные и неполярные аминокислоты. Например, построен симметричный код с помехоустойчивостью 78,29%, в котором четыре стоп-кодона расположены в клетке таблицы с первыми двумя нуклеотидами ТА, в клетке ТС содержатся аминокислоты цистеин, цистеин, триптофан, триптофан, а в клетке ТГ — аминокислоты тирозин, тирозин, серин, серин.

ГЕНЕТИЧЕСКИЕ МУТАЦИИ

Возможность быстро расшифровывать индивидуальные геномы людей позволила накопить огромные массивы данных о заболеваниях и связанных с ними мутациях в ДНК. Появились службы планирования семьи, позволяющие супружеским парам определить склонность их будущих детей к определенным заболеваниям. Наиболее распространенным типом мутаций, которые приводят к заболеваниям, являются точечные мутации, в результате которых единичный нуклеотид гена меняется на другой нуклеотид. В настоящей работе исследованы мутации, связанные с четырьмя типами заболеваний: аутоиммунные, генетические, нейродегенеративные, а также другие заболевания, не попавшие в перечисленные группы.

Точечные мутации подразделяют на две категории, различающиеся по характеру изменений в ДНК, вызванных мутагенным агентом. При транзициях происходит замещение нуклеотида типа пурин на пурин ($A \rightarrow G, G \rightarrow A$) или пиримидин на пиримидин ($C \rightarrow T, T \rightarrow C$). Трансверсии представляют собой замены типа пурин на пиримидин или пиримидин на пурин ($A \rightarrow C, G \rightarrow C, G \rightarrow T, C \rightarrow A, C \rightarrow G, T \rightarrow A$ или $T \rightarrow G$). Механизм трансверсий неясен, до сих пор не найдены агенты, которые бы избирательно вызывали такие замены [6].

Поскольку наиболее частыми мутациями являются транзиции, в стандартном и симметричном кодах можно заметить, что при замене третьего нуклеотида A в кодоне на G или T на C аминокислота не меняется. Поэтому такие аминокислоты, как тирозин, цистеин, серин, аргинин и другие, расположены в соответствующих позициях по третьему нуклеотиду.

Среди всех мутаций находились те, которые попадают в белок-кодирующую область гена и потенциально могут повлиять на аминокислотный состав белка. Для этого использовались данные ресурса (http://www.eupedia.com/genetics/medical_dna_test.shtml), где заболеваниям ставятся в соответствие связанные с ними мутации в ДНК. Подробные данные по каждой мутации вычислялись по идентификатору мутации из ресурса (<http://www.ncbi.nlm.nih.gov/snp>). Таким образом, были получены пары исходных и мутировавших триплетов нуклеотидов и соответственно кодируемых ими аминокислот.

Всего было проверено 103 мутации для перечисленных типов заболеваний, в 55 случаях имело место сохранение полярности на паре исходный кодон и его мутация. Наблюдалось 48 транзиций типа пурин на пурин и 27 — пиримидин на пиримидин, трансверсии имели место в 28 случаях. Количество мутаций первого нуклеотида — 57, замен второго — 33 и третьего — 13. Мутации третьего нуклеотида не меняют аминокислоты, однако они наблюдаются в интронах, нарушая вырезание (сплайсинг) интронов из белок-кодирующей последовательности гена, поэтому также приводят к заболеваниям.

Согласно палеонтологическим данным бактерии уже существовали три миллиарда лет назад. Таким образом, весьма вероятно, что генетический код тоже функционировал в то время. Поскольку у бактерий отсутствуют интроны, в стандартном коде не было способов защиты от мутаций в некодирующих областях гена.

В 22 случаях симметричный код (табл. 2) сохраняет полярность исходного кодона за счет исправления полярности мутированного кодона. В одном случае он нарушает полярность на паре кодонов там, где она имела место для стандартного кода. В 21 случае этот код исправил нарушение полярности за счет изменения полярности исходного кодона.

Симметричный код можно рассматривать как надстройку над стандартным кодом, которая исправляет полярность при мутациях в кодонах.

В стандартном генетическом коде сохранение полярности в большей степени зависит от того, какой именно нуклеотид в кодоне подвергается изменению. Например, в стандартном коде при мутации второго нуклеотида сохранение полярности составляет величину 0.38, а при замене первого нуклеотида — 0.59. В симметричном коде сохранение полярности при замене первого нуклеотида значительно выше, чем в стандартном коде, — 0.91, а для второго немного ниже — 0.30. Полученные

Таблица 3. Вариант симметричного кода

Первое основание	Второе основание				Третье основание
	T	C	A	G	
T	Фенилаланин (F), н	Серин (S), п	Тирозин (Y), п	Цистеин (C), н	T
	Фенилаланин (F), н	Серин (S), п	Тирозин (Y), п	Цистеин (C), н	C
	Лейцин (L), н	Серин (S), п	Стоп	Стоп	A
	Лейцин (L), н	Серин (S), п	Стоп	Триптофан (W), н	G
C	Лейцин (L), н	Аргинин (R), п	Гистидин (H), п	Пролин (P), н	T
	Лейцин (L), н	Аргинин (R), п	Гистидин (H), п	Пролин (P), н	C
	Лейцин (L), н	Аргинин (R), п	Глутамин (Q), п	Пролин (P), н	A
	Лейцин (L), н	Аргинин (R), п	Глутамин (Q), п	Пролин (P), н	G
A	Изолейцин (I), н	Треонин (T), п	Аспарагин (N), п	Аланин (A), н	T
	Изолейцин (I), н	Треонин (T), п	Аспарагин (N), п	Аланин (A), н	C
	Изолейцин (I), н	Треонин (T), п	Лизин (K), п	Аланин (A), н	A
	Метионин (M), н	Треонин (T), п	Лизин (K), п	Аланин (A), н	G
G	Валин (V), н	Серин (S), п	Аспарагиновая кислота (D), п	Глицин (G), н	T
	Валин (V), н	Серин (S), п	Аспарагиновая кислота (D), п	Глицин (G), н	C
	Валин (V), н	Аргинин (R), п	Глутаминовая кислота (E), п	Глицин (G), н	A
	Валин (V), н	Аргинин (R), п	Глутаминовая кислота (E), п	Глицин (G), н	G

величины легко получить из табл. 1 и 2. Таким образом, симметричный код примерно на четверть превосходит стандартный код относительно сохранения полярности при мутациях в первых двух нуклеотидах кодона и это соответствует тому, что стандартный код на 3/4 симметричен. Если учесть, что мутации первых двух нуклеотидов наблюдались в 90 случаях, то одна четверть от 90 примерно совпадает с полученным числом 22 исправлений полярности для симметричного кода.

Проверен еще один вариант симметричного кода, который можно построить на основе восьми парных перестановок из стандартного кода. В табл. 3 жирным шрифтом выделены изменения кода по сравнению со стандартным. Помехоустойчивость этого кода составляет величину $408/526 = 77,57\%$, и он уступает симметричному коду в табл. 2. В 19 случаях код в табл. 3 исправляет нарушение полярности на паре исходный кодон–мутированный кодон за счет исправления полярности мутированного кодона, однако в восьми случаях он нарушает полярность там, где она имела место для стандартного кода.

Заметим, что симметрия кода, понимаемая как переход от полярной аминокислоты к неполярной при переходе от кодона к антикодону, не является необходимым условием оптимальности генетического кода. Например, если во всех кодонах кода в табл. 2 переставить местами второй и третий нуклеотиды, т.е. в каждой строке кодоны «повернуть на 90 градусов», получим код в табл. 4 с такой же помехоустойчивостью, в котором симметрия выполняется примерно в половине случаев. Если в коде, приведенном в табл. 2, поменять местами аминокислоты столбцов C и A, получим код, для которого симметрия не выполняется — кодон и антикодон всегда кодируют аминокислоту одного типа (полярную или неполярную).

Оказалось, что описанный несимметричный код дает плохие результаты на исследуемых мутациях, поскольку в 26 случаях он нарушает полярность на паре: исходный кодон–мутация кодона, там, где она сохранялась для стандартного кода. Код, для которого симметрия выполняется в половине случаев (табл. 4), также имеет много таких нарушений по сравнению со стандартным кодом.

В табл. 5 приведены примеры сохранения полярности исходного кодона за счет исправления полярности кодона с мутацией на основе симметричного кода.

Таблица 4. Код, для которого симметрия выполняется в половине случаев

Первое основание	Второе основание				Третье основание
	T	C	A	G	
T	Фенилаланин (F)	Фенилаланин (F)	Лейцин (L)	Лейцин (L)	T
	Стоп	Триптофан (W)	Цистеин (C)	Цистеин (C)	C
	Тирозин (Y)	Тирозин (Y)	Стоп	Стоп	A
	Серин (S)	Серин (S)	Серин (S)	Серин (S)	G
C	Лейцин (L)	Лейцин (L)	Лейцин (L)	Лейцин (L)	T
	Пролин (P)	Пролин (P)	Пролин (P)	Пролин (P)	C
	Гистидин (H)	Гистидин (H)	Глутамин (Q)	Глутамин (Q)	A
	Аргинин (R)	Аргинин (R)	Аргинин (R)	Аргинин (R)	G
A	Изолейцин (I)	Изолейцин (I)	Изолейцин (I)	Метионин (S)	T
	Глицин (G)	Глицин (G)	Глицин (G)	Глицин (G)	C
	Аспарагин (N)	Аспарагин (N)	Лизин (K)	Лизин (K)	A
	Серин (S)	Серин (S)	Аргинин (R)	Аргинин (R)	G
G	Валин (V)	Валин (V)	Валин (V)	Валин (V)	T
	Аланин (A)	Аланин (A)	Аланин (A)	Аланин (A)	C
	Аспарагиновая кислота (D)	Аспарагиновая кислота (D)	Глутаминовая кислота (E)	Глутаминовая кислота (E)	A
	Треонин (T)	Треонин (T)	Треонин (T)	Треонин (T)	G

Таблица 5. Примеры сохранения полярности

Исходный кодон	Мутация кодона	Болезнь
GCA	ACA	Диабет
CGG	TGG	Диабет
ATA	ACA	Астма
ATC	ACC	Сложный склероз
GCC	ACC	Нарушение слуха
CGC	TGC	Альцгеймера

ЗАКЛЮЧЕНИЕ

На основе соотношений симметрии в ДНК построен симметричный код, помехоустойчивость которого при мутациях в нуклеотидах на 9% выше, чем у стандартного кода. Симметричный код в табл. 2 образован из стандартного кода за счет минимального числа из восьми парных перестановок. С помощью генетических алгоритмов построены варианты наиболее помехоустойчивых кодов, отличающихся от стандартного кода числом стоп-кодонов, а также количеством триплетов, определяющих полярные и неполярные аминокислоты. Согласно информации из современных баз данных генетических заболеваний многие болезни связаны с точечными мутациями нуклеотидов. Нарушение полярности аминокислот при мутациях приводит к изменению структуры белка и соответственно возникновению серьезных заболеваний в организме человека. При исследовании генетических болезней оказалось, что построенный симметричный код во многих случаях исправляет нарушение полярности при мутациях в нуклеотидах.

СПИСОК ЛИТЕРАТУРЫ

1. Ичас М. Биологический код. — М.: Мир, 1971. — 352 с.
2. Медников Б.М. Аксиомы биологии. — М.: Знание, 1982. — 136 с.
3. Сергиенко И.В., Гупал А.М., Островский А.В. Помехоустойчивость генетических кодов. Оптимальные коды // Докл. Акад. Наук. — 2013. — 453, № 5. — С. 492–495.
4. Сергиенко И.В., Гупал А.М., Островский А.В. Устойчивость генетического кода к точечным мутациям // Кибернетика и системный анализ. — 2014. — № 5. — С. 17–24.
5. Сергиенко И.В., Гупал А.М., Вагис А.А. Правила симметрии в записи генетической информации в ДНК // Кибернетика и системный анализ. — 2011. — № 3. — С. 88–94.
6. Ленинджер А. Основы биохимии. — М.: Мир, 1985. — Т. 3. — 320 с.

Поступила 27.10.2015