



Аннотация. На основе симметрии в ДНК построены оптимальные генетические коды, помехоустойчивость которых относительно полярности аминокислот при мутациях в нуклеотидах значительно выше, чем у стандартного кода. Исследована помехоустойчивость оптимальных симметричных и несимметричных генетических кодов. С учетом баз данных генетических заболеваний показано, что оптимальный код, для которого симметрия выполняется в половине случаев, сохраняет полярность аминокислот при мутациях первого и второго нуклеотидов в кодоне по сравнению со стандартным кодом.

Ключевые слова: генетический код, симметрия в ДНК, мутации в нуклеотидах, помехоустойчивость, аминокислота, кодон.

ВВЕДЕНИЕ

Стандартный генетический код, согласно которому в живых организмах происходит синтез белков на основе информации в ДНК, открыт в 60-х годах XX века. Одним из основных факторов, возможно влияющим на эволюцию кода, является повышение его устойчивости по отношению к мутациям — искажениям генетической информации, возникающим в клетках любого организма. Мутации происходят вследствие репликации в ДНК или разрушительных воздействий мутагенов, таких как химические агенты и радиация, которые изменяют структуру отдельных нуклеотидов в генах.

В работах [1, 2] исследована устойчивость генетического кода относительно полярности аминокислот. Описан механизм мутаций и вводится математический аппарат для оценки помехоустойчивости генетического кода. Проведено сравнение стандартного кода со случайно сгенерированными кодами, на которые налагаются определенные ограничения. Компьютерными расчетами получено, что помехоустойчивость кода — важный фактор для его эволюции: устойчивость к помехам стандартного кода значительно выше, чем у случайных кодов, что свидетельствует о целесообразности выбора стандартного кода. Исследована помехоустойчивость генетического кода относительно таких характеристик аминокислот, как полярность, гидрофобность и склонность к образованию спиралей. Описан генетический алгоритм для оптимизации помехоустойчивости кода, приведены результаты его работы, которые проанализированы с точки зрения различных групп мутаций.

СТАНДАРТНЫЙ ГЕНЕТИЧЕСКИЙ КОД

Известно, что замена аминокислоты и нарушение ее полярности в белках вредна, поскольку чередованием полярных и неполярных аминокислот определяется его пространственная структура. В работах [3, 4] предполагалось, что

Таблица 1. Стандартный генетический код

| Первое основание | Второе основание | | | | Третье основание |
|------------------|--------------------|---------------|-----------------------|------------------|------------------|
| | Т | С | А | Г | |
| Т | фенилаланин | серин | тирозин | цистеин | Т |
| | фенилаланин | серин | тирозин | цистеин | С |
| | лейцин | серин | стоп | стоп | А |
| | лейцин | серин | стоп | триптофан | Г |
| С | лейцин | пролин | гистидин | аргинин | Т |
| | лейцин | пролин | гистидин | аргинин | С |
| | лейцин | пролин | глутамин | аргинин | А |
| | лейцин | пролин | глутамин | аргинин | Г |
| А | изолейцин | треонин | аспарагин | серин | Т |
| | изолейцин | треонин | аспарагин | серин | С |
| | изолейцин | треонин | лизин | аргинин | А |
| | метионин | треонин | лизин | аргинин | Г |
| Г | валин | аланин | аспарагиновая кислота | глицин | Т |
| | валин | аланин | аспарагиновая кислота | глицин | С |
| | валин | аланин | глутаминовая кислота | глицин | А |
| | валин | аланин | глутаминовая кислота | глицин | Г |

в результате эволюции отобран наиболее помехоустойчивый вариант, т.е. стандартный генетический код живых организмов (табл. 1) сводит к минимуму замены между полярными и неполярными аминокислотами.

Аминокислоты отличаются по взаимодействию с молекулами воды. Неполярные аминокислоты более гидрофобны, они отталкивают молекулы воды и слипаются одна с другой. Полярные аминокислоты гидрофильны и взаимодействуют с водой, в результате чего образуется сложная форма, которая соответствует выполняемой белком функции. Если вследствие мутации полярный остаток в белке заменится неполярным (или наоборот), то форма молекулы может измениться настолько, что белок не сможет выполнять свою функцию. Неполярные аминокислоты в табл. 1 и в приведенных далее таблицах выделены жирным шрифтом.

Каждый кодон допускает девять однократных замен, число кодирующих аминокислоты триплетов равно 61. Поэтому количество возможных замен нуклеотидов для всех кодонов составляет 549, из них 23 замены нуклеотидов приводят к появлению стоп-кодонов, 134 — не изменяют кодируемой аминокислоты, 230 — не изменяют 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 изменяют полярности, т.е. его помехоустойчивость составляет $364/526 = 69.20\%$.

Компьютерные расчеты показали, что стандартный генетический код не является оптимальным относительно помехоустойчивости даже локально, поскольку существуют парные перестановки кодонов, а также изменения в типе аминокислоты, которые увеличивают его помехоустойчивость [1, 2]. Существует 40 единичных изменений в типе аминокислоты, увеличивающих помехоустойчивость кода. Из них 38 случаев составляют замены аминокислоты стоп-кодонами. В двух случаях стоп-кодон заменяется полярной аминокислотой. Таким образом, не существует единичных изменений, улучшающих помехоустойчивость, которые бы сохраняли при этом три стоп-кодона. Существует 51 парная перестановка аминокислот со стоп-кодоном, увеличивающая помехоустойчивость.

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

$$n(ijk) = n(\bar{k}\bar{j}\bar{i}),$$

где $n(ijk)$ — число троек оснований (ijk) , подсчитанных по одной нити, $i, j, k \in \{A, C, G, T\}$, $(\bar{k}\bar{j}\bar{i})$ — антикодон кодона (ijk) . В симметричном коде относительно полярности кодон (ijk) кодирует полярную аминокислоту, а антикодон $(\bar{k}\bar{j}\bar{i})$ — неполярную либо наоборот (стоп-кодона не учитываются). При таком определении симметричного кода в каждом гене количество полярных и неполярных аминокислот примерно совпадает.

Проанализировав табл. 1, отметим, что в столбце Т содержатся неполярные аминокислоты, в столбце А — полярные аминокислоты, в столбцах С и G — полярные и неполярные аминокислоты. Поэтому можно считать, что стандартный код на 3/4 соответствует симметричному коду.

В табл. 2 представлен оптимальный помехоустойчивый симметричный код [1, 2], у которого три стоп-кодона и число триплетов, определяющих полярные и неполярные аминокислоты, такое же, как у стандартного кода: 30 триплетов задают полярные аминокислоты, 31 — неполярные. Этот код построен на основе восьми парных перестановок из стандартного кода. Помехоустойчивость кода составляет $408/524 = 77.86\%$. Как видим, симметричный код имеет достаточно простую структуру и его можно получить без выполнения компьютерных расчетов. Помехоустойчивость симметричного кода немного превышает 7/9 (за счет наличия стоп-кодонов), поскольку не более двух мутаций из девяти нарушают полярность аминокислот.

Таблица 2. Симметричный код

| Первое основание | Второе основание | | | | Третье основание |
|------------------|------------------|-----------|-----------------------|---------|------------------|
| | Т | С | А | G | |
| Т | фенилаланин | стоп | тирозин | серин | Т |
| | фенилаланин | триптофан | тирозин | серин | С |
| | лейцин | цистеин | стоп | серин | А |
| | лейцин | цистеин | стоп | серин | G |
| С | лейцин | пролин | гистидин | аргинин | Т |
| | лейцин | пролин | гистидин | аргинин | С |
| | лейцин | пролин | глутамин | аргинин | А |
| | лейцин | пролин | глутамин | аргинин | G |
| А | изолейцин | глицин | аспарагин | серин | Т |
| | изолейцин | глицин | аспарагин | серин | С |
| | изолейцин | глицин | лизин | аргинин | А |
| | метионин | глицин | лизин | аргинин | G |
| G | валин | аланин | аспарагиновая кислота | треонин | Т |
| | валин | аланин | аспарагиновая кислота | треонин | С |
| | валин | аланин | глутаминовая кислота | треонин | А |
| | валин | аланин | глутаминовая кислота | треонин | G |

Первые два столбца (см. табл. 2) полученного кода содержат неполярные аминокислоты, третий и четвертый — полярные аминокислоты, т.е. если кодон (ijk) определяет полярную аминокислоту, то кодон ($\bar{k}\bar{j}\bar{i}$) — неполярную и наоборот. Как отмечалось ранее, код, представленный в табл. 2, оптимален, поскольку с помощью парных перестановок аминокислот нельзя увеличить помехоустойчивость симметричного кода, а других способов ее повышения не существует, если сохранять характеристики стандартного кода.

Как отмечалось ранее, помехоустойчивость кода — важный фактор для его эволюции: устойчивость к помехам стандартного кода значительно выше, чем у случайных кодов, что свидетельствует о целесообразности его выбора.

С помощью генетических алгоритмов построены различные варианты наиболее помехоустойчивых кодов с другими характеристиками, чем у стандартного кода, а также подтвержден оптимальный симметричный код (см. табл. 2) [2].

ГЕНЕТИЧЕСКИЕ МУТАЦИИ

Возможность быстро расшифровывать индивидуальные геномы человека позволила накопить огромные массивы данных о заболеваниях и связанных с ними мутациях в ДНК. Наиболее распространенным типом мутаций, вызывающих заболевания, являются точечные, в результате которых единственный нуклеотид гена заменяется другим нуклеотидом. Исследовались мутации, обусловленные аутоиммунными заболеваниями: аутоиммунные, онкологические, седечно-сосудистые, генетические, нейродегенеративные, психические расстройства, пагубные привычки. Кроме того, имеются данные о мутациях, характеризующих связь физиологических и нейрофизиологических состояний (склонность к лишнему весу, долголетие, группа крови, повышенное давление, различные зависимости), не являющихся болезнями.

Использовались данные из интернет-ресурсов, в которых заболеваниям ставились в соответствие связанные с ними мутации в ДНК, т.е. были получены пары исходных и мутированных триплетов нуклеотидов и соответственно кодируемых ими аминокислот.

Стандартным кодом проверено 300 мутаций для приведенных типов заболеваний: в 100 случаях нарушалась полярность на парах исходного и мутированного кодонов. Количество мутаций третьего нуклеотида составило 43.

Согласно структуре стандартного кода мутации третьего нуклеотида чаще всего не изменяют аминокислоты, если возникают в кодирующих участках геномов, поэтому они не связаны с заболеваниями.

Сложнее обобщить данные о воздействиях молчащих мутаций (соответствующий кодон продолжает кодировать ту же аминокислоту), возникающих вне кодирующих областей генов. Наиболее исследованы области, в которых мутации происходят в интронах или на экзон-интронных границах. В этих областях точечные мутации важны, если они изменяют нуклеотиды, вырезающие интроны. Для стандартного кода 37 мутаций в третьем нуклеотиде не изменили аминокислоты, поэтому они не связаны с кодируемыми областями генов.

Согласно палеонтологическим данным бактерии существовали три миллиарда лет тому назад. Таким образом, весьма вероятно, что и генетический код тоже функционировал. Поскольку у бактерий отсутствуют интроны, в стандартном коде не имелось способов защиты от мутаций в не кодирующих областях гена. Тот факт, что аминокислоты определяются обычно уже первыми двумя нуклеотидами кодона, подтверждает, что генетический код был первоначально «дуплетным» и содержал информацию примерно о 15 «первичных» аминокислотах. Возможно, в таком примитивном коде третий нуклеотид использовался в качестве «запятой». Позднее могла появиться группа «новых» аминокислот: аспарагин, глутамин, метионин, тирозин и триптофан, и для их кодирования потребо-

вались триплеты. Наиболее частыми мутациями являются транзиции. Поэтому замены А на G или С на Т, если они произошли в третьем нуклеotide, часто дают триплеты-синонимы, поэтому вырожденность генетического кода делает геном в целом более стабильным [6]. В рассматриваемом случае 15 замен А на G (или наоборот) и 13 замен С на Т.

В [7] описаны случаи, когда мутации происходят в интронах или на экзон-интронных границах. Два из этих типов: интроны GU-AG и AU-AC, встречаются в кодирующих белок генах эукариотов. В большинстве интронов РНК первые два нуклеотида последовательности интрона 5'-GU-3', а последние два 5'-AG-3'. Поэтому их называют интронами «GU-AG» и все члены этого класса вырезаются одинаковым образом. Эти консервативные мотивы были определены после открытия интронов и предполагалось, что они важны в процессе сплайсинга.

Например, мутация G или T в ДНК-копии 5'-сайта вырезания интрона GU-AG либо мутация A или G в 3'-сайте вырезания прервет процесс сплайсинга, потому что правильная экзон-интронная граница не будет определяться.

Симметричный код (см. табл. 2) в 79 случаях из 100 исправил нарушение полярности при мутациях кодона и в семи случаях нарушил полярность, где она имела место для стандартного кода.

ОПТИМАЛЬНЫЕ НЕСИММЕТРИЧНЫЕ КОДЫ

Отметим, что симметрия кода, понимаемая как переход от полярной аминокислоты к неполярной при переходе от кодона к антикодону, не является необходимым условием оптимальности генетического кода.

Если в коде (см. табл. 2) поменять местами аминокислоты столбцов С и А, получим код (табл. 3), для которого симметрия не выполняется вообще: кодон и антикодон кодируют аминокислоту одного типа, поскольку неполярные аминокислоты содержатся в столбцах Т и А, а полярные — в столбцах С и G. Помехоустойчивость кода в табл. 3 такая же, как у симметричного кода $408 / 524 = 77.86 \%$.

Таблица 3. Оптимальный код, для которого симметрия не выполняется

| Первое основание | Второе основание | | | | Третье основание |
|------------------|--------------------|-----------------------|------------------|---------|------------------|
| | Т | С | А | G | |
| Т | фенилаланин | тирозин | стоп | серин | Т |
| | фенилаланин | тирозин | триптофан | серин | С |
| | лейцин | стоп | цистеин | серин | А |
| | лейцин | стоп | цистеин | серин | G |
| С | лейцин | гистидин | пролин | аргинин | Т |
| | лейцин | гистидин | пролин | аргинин | С |
| | лейцин | глутамин | пролин | аргинин | А |
| | лейцин | глутамин | пролин | аргинин | G |
| А | изолейцин | аспарагин | глицин | серин | Т |
| | изолейцин | аспарагин | глицин | серин | С |
| | изолейцин | лизин | глицин | аргинин | А |
| | метионин | лизин | глицин | аргинин | G |
| G | валин | аспарагиновая кислота | аланин | треонин | Т |
| | валин | аспарагиновая кислота | аланин | треонин | С |
| | валин | глутаминовая кислота | аланин | треонин | А |
| | валин | глутаминовая кислота | аланин | треонин | G |

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

| Первое основание | Второе основание | | | | Третье основание |
|------------------|---|---|--|--|------------------|
| | Т | С | А | Г | |
| Т | фенилаланин стоп тирозин серин | фенилаланин триптофан тирозин серин | лейцин цистеин стоп серин | лейцин цистеин стоп серин | Т С А Г |
| С | лейцин пролин гистидин аргинин | лейцин пролин гистидин аргинин | лейцин пролин глутамин аргинин | лейцин пролин глутамин аргинин | Т С А Г |
| А | изолейцин глицин аспарагин серин | изолейцин глицин аспарагин серин | изолейцин глицин лизин аргинин | метионин глицин лизин аргинин | Т С А Г |
| Г | валин аланин аспарагиновая кислота треонин | валин аланин аспарагиновая кислота треонин | валин аланин глутаминовая кислота треонин | валин аланин глутаминовая кислота треонин | Т С А Г |

Оказалось, что несимметричный код по сравнению со стандартным дает плохие результаты на исследуемых мутациях, поскольку в 62 случаях он исправляет нарушение полярности на паре кодон–мутант, однако в 43 случаях нарушает полярность там, где она имела место для стандартного кода. Это не удивительно, поскольку сравнение проводилось со стандартным кодом, который на 3/4 соответствует симметричному коду.

Если во всех кодонах кода в табл. 2 переставить местами второй и третий нуклеотиды, т.е. в каждой строке кодоны «повернуть на 90°», получим код (табл. 4) с такой же помехоустойчивостью, что и у симметричного кода, в котором симметрия выполняется примерно в половине случаев. В этом коде неполярные аминокислоты определяются триплетами с третьим нуклеотидом Т и С, а у полярных аминокислот третий нуклеотид — А и Г.

Отметим, что этот код во всех случаях сохраняет полярность при мутациях первого и второго нуклеотидов в кодоне по сравнению со стандартным кодом, а нарушение полярности и замена аминокислот происходит в третьем нуклеотиде. Поэтому мутации в третьем нуклеотиде необходимо учитывать на основе стандартного кода.

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 (см. табл. 2) образован из стандартного кода за счет минимального числа из восьми парных перестановок. Из анализа современных баз данных генетических заболеваний следует, что многие заболевания связаны с точечными мутациями нуклеотидов. Нарушение полярности аминокислот при мутациях приводит к изменению структуры белка и, следовательно, возникновению серьезных заболеваний у человека. При исследовании генетических болезней оказалось, что построенный симметричный код во многих случаях исправляет нарушение полярности при мутациях в нуклеотидах. Наполовину симметричный генетический код во всех случаях

сохраняет полярность аминокислот при мутациях первого и второго нуклеотидов в кодоне по сравнению со стандартным кодом.

Стандартный и оптимальные коды определяют дополнительные признаки мутаций, связанные с сохранением и нарушением полярности при мутациях. Поэтому эту информацию можно использовать для качественной диагностики генетических заболеваний с помощью байесовских процедур распознавания, построенных на основе обучающих выборок прецедентов.

СПИСОК ЛИТЕРАТУРЫ

1. Сергиенко И.В., Гупал А.М., Островский А.В. Устойчивость генетического кода к точечным мутациям. *Кибернетика и системный анализ*. 2014. Т. 50, № 5. С. 17–24.
2. Гупал А.М., Сергиенко И.В. Симметрия в ДНК. Методы распознавания дискретных последовательностей. Киев: Наук. думка, 2016. 228 с.
3. Ичас М. Биологический код. Москва: Мир, 1971. 352 с.
4. Медников Б.М. Аксиомы биологии. Москва: Знание, 1982. 136 с.
5. Гупал А.М., Гупал Н.А., Островский А.В. Симметрия и свойства записи генетической информации в ДНК. *Проблемы управления и информатики*. 2011. № 3. С. 88–94.
6. Ленинджер А. Основы биохимии: в 3-х т. Т. 3. Москва: Мир, 1985. 320 с.
7. Браун Т.А. Геномы. Москва; Ижевск: Ин-т компьютерных исследований, 2011. 922 с.

Надійшла до редакції 12.02.2018

І.В. Сергієнко, Б.О. Білецький, А.М. Гупал, М.А. Гупал **ОПТИМАЛЬНІ ЗАВАДОСТІЙКИ ГЕНЕТИЧНІ КОДИ**

Анотація. На основі симетрії в ДНК побудовано оптимальні генетичні коди, завадостійкість яких відносно полярності амінокислот для мутацій у нуклеотидах значно вища, ніж у стандартного коду. Досліджено завадостійкість оптимальних симетричних та несиметричних кодів. На основі баз даних генетичних захворювань показано, що оптимальний код, для якого симетрія виконується у половині випадків, зберігає полярність амінокислот для мутацій першого та другого нуклеотидів у кодоні в порівнянні зі стандартним кодом.

Ключові слова: генетичний код, симетрія в ДНК, мутації в нуклеотидах, завадостійкість, амінокислота, кодон.

I.V. Sergienko, B.A. Biletskyy, A.M. Gupal, M.A. Gupal **OPTIMAL NOISE-IMMUNE GENETIC CODES**

Abstract. Optimal genetic codes are constructed using DNA symmetry, whose noise immunity with respect to polarity of amino acids in case of mutations in nucleotides is much greater than immunity of standard codes. Noise immunity of optimal symmetric and non-symmetric genetic codes is analyzed. Databases of genetic diseases are used to show that optimal symmetric code for which symmetry holds in half of cases keeps polarity of amino acids at mutations in first and second nucleotides of codon as compared with standard code.

Keywords: genetic code, symmetry in DNA, mutations in nucleotides, noise immunity, amino acid, codon.

Сергиенко Иван Васильевич,

академик НАН Украины, доктор физ.-мат. наук, директор Института кибернетики им. В.М. Глушкова НАН Украины, Киев, e-mail: ikcab806@gmail.com.

Білецький Борис Александрович,

кандидат физ.-мат. наук, старший научный сотрудник Института кибернетики им. В.М. Глушкова НАН Украины, Киев, e-mail: borys.biletskyy@gmail.com.

Гупал Анатолий Михайлович,

чл.-кор. НАН Украины, доктор физ.-мат. наук, заведующий отделом Института кибернетики им. В.М. Глушкова НАН Украины, Киев, e-mail: gupalanatol@gmail.com.

Гупал Никита Анатольевич,

кандидат физ.-мат. наук, научный сотрудник Института кибернетики им. В.М. Глушкова НАН Украины, Киев, e-mail: kitgupal@gmail.com.