

УДК 519.272.2

А. М. Гупал, д-р физ.-мат. наук, член-кор. НАН України,

Н. А. Гупал, научный сотрудник

Институт кибернетики имени В. М. Глушкова НАН Украины, г. Киев

## СИММЕТРИЯ В ДНК. СИММЕТРИЧНЫЙ КОД

Из симметрии в ДНК построен симметричный код относительно полярности аминокислот при мутациях в нуклеотидах. Исследована помехоустойчивость генетических кодов относительно полярности аминокислот. На основе баз данных генетических заболеваний показано, что симметричный код во многих случаях исправляет нарушение полярности при мутациях.

**Ключевые слова:** *генетический код, мутации, аминокислота, нуклеотид.*

**Введение.** Первые публикации по симметрии в ДНК относятся к 90-м годам XX в. после появления новой технологии секвенирования, которую использовали в проекте «Геном человека». В работах зарубежных авторов в виде эмпирических таблиц было показано, что в одной цепочке ДНК количество коротких последовательностей нуклеотидов примерно равно количеству обратных комплементарных последовательностей. В результате такого представления компьютерных расчетов в этих работах феномен симметрии не был объяснен.

ДНК имеет форму двойной спирали, информация записана в четырехбуквенном алфавите нуклеотидов (оснований) аденин (А), цитозин (С), гуанин (G), тимин (Т). Известно, что С–G, А–T — комплементарные пары оснований, связывающие две цепи. Запись и считывание оснований у первой нити хромосомы ДНК в модели Уотсона-Крика выполняется слева направо, а у комплементарной второй нити — справа налево.

**Симметрия оснований.** Вычисления на компьютере получили, что количества оснований А и Т, а также С и G, подсчитанные по одной нити ДНК, примерно совпадают и ошибка составляет сотые доли процента.

В работе [1] симметрия в ДНК представлена в виде математических формул, что значительно упрощает восприятие этого феномена, а также является основой построения математического аппарата для получения новых результатов.

**Симметрия пар оснований.** В результате расчетов получили, что для пар оснований выполняются соотношения в виде формулы

$$n(ij) = n(\bar{j}\bar{i}), \quad (1)$$

где  $i, j \in \{A, C, G, T\}$ ,  $\bar{A} = T$ ,  $\bar{C} = G$ ,  $\bar{T} = A$ ,  $\bar{G} = C$ . Отметим, что пары AT, TA, CG и GC не присутствуют в (1), поскольку они приводят к тавтологии.

Формулы симметрии (1) определяют шесть дополнительных связывающих ограничений в записи генетической информации по одной нити ДНК, и поэтому они воспринимаются как совершенно необязательные требования. Возможно поэтому, феномен симметрии не изложен в учебниках и монографиях по биологии и биоинформатике.

Отметим, что длина участков генов составляет несколько тысяч нуклеотидов. Они расположены на двух нитях ДНК и для них симметрия также выполняется. Это означает, что оценки переходных вероятностей для однородных цепей Маркова, подсчитанные на двух противоположных нитях ДНК, совпадают. Поэтому для решения задач распознавания генов и белков можно использовать модели Маркова определенных порядков, в том числе и со скрытыми состояниями. При отсутствии симметрии применять такие модели было бы невозможно. Не удивительно, что модели Маркова рассматривают во многих работах, посвященных распознаванию аминокислотных и нуклеотидных последовательностей.

С помощью формул вида (1) выведено правило понижения симметрии: показано, что из симметрии последовательности оснований определенной длины вытекает симметрия для последовательностей, имеющих меньшую длину, вплоть до отдельных оснований.

Из условия Маркова, которое было сформулировано при анализе литературных текстов, следует правило повышения симметрии: из симметрии пар оснований вытекает симметрия коротких последовательностей оснований.

Отметим, что длина участков генов составляет несколько тысяч нуклеотидов. Они расположены на двух нитях ДНК и для них симметрия также выполняется. Это означает, что оценки переходных вероятностей для однородных цепей Маркова, подсчитанные на двух противоположных нитях ДНК, совпадают. Поэтому для решения задач распознавания генов и белков можно использовать модели Маркова определенных порядков, в том числе и со скрытыми состояниями. При отсутствии симметрии применять такие модели было бы невозможно. В зарубежных публикациях они составляют примерно половину статей в биоинформатике.

Легко видеть, что для коротких слов модель с одинаковой полярностью имеет больше связывающих ограничений, чем модель с противоположной полярностью нитей. Поэтому ДНК с противоположной полярностью нитей имеет больше степеней свободы, чем ДНК с одинаковой полярностью, т. е. с точки зрения теории информации модель Уотсона-Крика более эффективна.

Серьезный аргумент в пользу феномена симметрии — исследование устойчивости стандартного генетического кода при случайных мутациях нуклеотидов в кодоне.

Стандартный генетический код, согласно которому в живых организмах происходит синтез белков на основе информации, содержащейся в ДНК, был открыт в 60-х годах XX в. 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).

Аминокислоты различаются по взаимодействию с молекулами воды. Неполярные аминокислоты более гидрофобны, они отталкивают молекулы воды и слипаются друг с другом. Полярные аминокислоты гидрофильны и взаимодействуют с водой, в результате чего образуется сложная форма, которая соответствует 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 свою функцию.

Стандартный генетический код сопоставляет каждой тройке нуклеотидов одну из двадцати аминокислот. Помехоустойчивость такого кода при точечных мутациях в нуклеотидах относительно полярности аминокислот составляет 69,20 %.

С помощью восьми парных перестановок построен симметричный код в табл. с помехоустойчивостью 77,86 %, в котором кодон и антикодон определяют аминокислоты с разной полярностью.

В стандартном коде в клетке С первой строки содержатся четыре полярные аминокислоты серин; а в клетке G — цистеин, цистеин, стоп и триптофан; в клетке С третьей строки содержатся четыре полярные аминокислоты треонин; а в клетке G четвертой строки — четыре неполярные аминокислоты глицин.

Стандартный код в таком смысле на 3/4 соответствует симметричному коду. Симметричный код оптимален, поскольку не существует способов повышения его помехоустойчивости, сохраняя при этом характеристики стандартного кода [2, 3].

В таблице неполярные аминокислоты выделены жирным шрифтом. Первые два столбца построенного кода содержат неполярные аминокислоты, третий и четвертый — полярные аминокислоты.

*Симметричный код*

Первое основание	Второе основание				Третье основание
	T	C	A	G	
T	фенилаланин фенилаланин лейцин лейцин	стоп триптофан цистеин цистеин	тирозин тирозин стоп стоп	серин серин серин серин	T C A G
C	лейцин лейцин лейцин лейцин	пролин пролин пролин пролин	гистидин гистидин глутамин глутамин	аргинин аргинин аргинин аргинин	T C A G
A	изолейцин изолейцин изолейцин метионин	глицин глицин глицин глицин	аспарагин аспарагин лизин лизин	серин серин аргинин аргинин	T C A G
G	валин валин валин валин	аланин аланин аланин аланин	асп. к-та асп. к-та глут. к-та глут. к-та	треонин треонин треонин треонин	T C A G

Компьютерные расчеты получили, что помехоустойчивость кода — важный фактор для эволюции кода: устойчивость к помехам стандартного кода значительно выше, чем у случайных кодов, что свидетельствует о неслучайном выборе стандартного кода.

На основе баз данных генетических заболеваний стандартным кодом было проверено триста мутаций для различных типов заболеваний. Примерно половина из них привела к нарушению полярности или к мутациям третьего нуклеотида. Симметричный код в 80 % случаев исправил нарушение полярности при мутациях кодона. Рассмотрено два варианта несимметричных кодов, помехоустойчивость которых такая же, как у симметричного оптимального кода. Однако эти коды дали плохие результаты по сравнению со стандартным кодом.

**Выводы.** Таким образом, показано, что помехоустойчивость кода имеет большое влияние на нарушение полярности аминокислот при точечных мутациях нуклеотидов.

**Список использованной литературы:**

1. Гупал А. М., Гупал Н. А., Островский А.В. Симметрия и свойства записи генетической информации в ДНК. *Проблемы управления и информатики*. 2011. № 3. С. 120–127.
2. Сергиенко И. В., Гупал А. М., Вагис А. А. Симметричный код и генетические мутации. *Кибернетика и системный анализ*. 2016. № 2. С. 73–80.

3. Гупал А. М., Сергиенко И. В. Симметрия в ДНК. Методы распознавания дискретных последовательностей. К.: Наукова думка, 2016. 228 с.

The symmetric code concerning polarity of amino acids at mutations in nucleotides is constructed of symmetry in DNA. The noise immunity of genetic code against polarity is analyzed. On the basis of databases of genetic diseases it is shown that the symmetric code in most cases corrects violation of polarity at mutations.

**Key words:** *genetic code, mutations, amino acid, nucleotide.*

Получено 9.03.2017

УДК 517.988

**С. В. Денисов**, ассистент,

**В. В. Семёнов**, д-р физ.-мат. наук, профессор

Киевский национальный университет имени Т. Шевченко, г. Киев

### **МОДИФИЦИРОВАННЫЙ ЭКСТРАГРАДИЕНТНЫЙ МЕТОД ДЛЯ ВАРИАЦИОННЫХ НЕРАВЕНСТВ**

Предложен модифицированный экстраградиентный метод с динамической регулировкой величины шага для решения вариационных неравенств с монотонными операторами, действующими в гильбертовом пространстве. Доказана слабая сходимость метода без предположения о липшицевости операторов.

**Ключевые слова:** *вариационное неравенство, монотонный оператор, экстраградиентный метод, слабая сходимость.*

**Введение.** Многие задачи исследования операций и математической физики могут быть записаны в форме вариационных неравенств вида:

$$\text{найти } x \in C : (Ax, y - x) \geq 0 \quad \forall y \in C,$$

где  $C$  — непустое выпуклое подмножество гильбертового пространства  $H$ ,  $A$  — монотонный оператор, действующий в  $H$  [1, 2]. Решение вариационных неравенств является активно развивающимся направлением прикладного нелинейного анализа. К настоящему времени предложено большое количество методов, в частности, методов проекционного типа (использующих операцию метрического проектирования на допустимое множество  $C$ ).

Известно, что в случае неоптимизационных постановок (поиск седловой точки или равновесия Нэша) для сходимости наиболее простого проекционного метода (аналога метода проекции градиента) необходимо выполнение усиленных условий монотонности.