



GLOBAL JOURNAL OF SCIENCE FRONTIER RESEARCH: F  
MATHEMATICS AND DECISION SCIENCES  
Volume 18 Issue 2 Version 1.0 Year 2018  
Type : Double Blind Peer Reviewed International Research Journal  
Publisher: Global Journals  
Online ISSN: 2249-4626 & Print ISSN: 0975-5896

## Mathematical Theory of Genetic Code

By N. N. Kozlov

*Institute of Applied Mathematics M.V. Keldysh RAS*

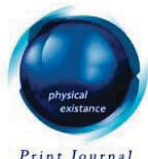
**Abstract-** On the basis of studying of genes new properties of a genetic code were established and its major integrated characteristics are calculated. Two groups of such characteristics were allocated. The first group treats integrated characteristics for the DNA areas where genes are in pairs blocked and all 5 cases of the overlappings allowed by structure of DNA were investigated. The second group of characteristics treats most DNA extended areas in which there are no genetic overlappings. The interrelation of the established integrated characteristics in the called groups is established. Were as a result found a number of unknown before effects. It was succeeded to establish two functions in which all codons reassignment participate in mitochondrial genetic codes (human and other organisms), and also essential distinction in integrated characteristics of such codes by comparison with a standard code. Other properties of structure of the genetic code, following from the received results are established also. On the basis of the constructed theory, the problem of calculating the genetic code was set (in 1992) and the first complete publication was published recently [26].

*GJSFR-F Classification: MSC 2010: 00A71*



*Strictly as per the compliance and regulations of:*





R<sub>ef</sub>

26. Kozlov. N. N. (2017) Computation of the Genetic Code: Full Version. Journal Computer and Communications, 5, 78-94 <https://doi.org/10.4236/jcc.2017.510008>.

# Mathematical Theory of Genetic Code

## Математическая теория генетического кода

N. N. Kozlov

**Abstract-** On the basis of studying of genes new properties of a genetic code were established and its major integrated characteristics are calculated. Two groups of such characteristics were allocated. The first group treats integrated characteristics for the DNA areas where genes are in pairs blocked and all 5 cases of the overlappings allowed by structure of DNA were investigated. The second group of characteristics treats most DNA extended areas in which there are no genetic overlappings. The interrelation of the established integrated characteristics in the called groups is established. Were as a result found a number of unknown before effects. It was succeeded to establish two functions in which all codons reassignment participate in mitochondrial genetic codes (human and other organisms), and also essential distinction in integrated characteristics of such codes by comparison with a standard code. Other properties of structure of the genetic code, following from the received results are established also. On the basis of the constructed theory, the problem of calculating the genetic code was set (in 1992) and the first complete publication was published recently [26].

**Abstract-** На основе изучения генов были установлены новые свойства генетического кода и вычислены важнейшие его интегральные характеристики. Были выделены две группы таких характеристик. Первая группа относится к интегральным характеристикам для областей ДНК, где гены попарно перекрываются и были исследованы все 5 случаев перекрытий, разрешенных структурой ДНК. Вторая группа характеристик относится к наиболее протяженным областям ДНК, в которых нет генетических перекрытий. Устанавливается взаимосвязь установленных интегральных характеристик в названных группах. В результате были обнаружены ряд неизвестных ранее эффектов. Удалось установить две функции, в которых участвуют все переосмысленные кодоны в митохондриальных генетических кодах (человека и других организмов), а также существенное различие в интегральных характеристиках таких кодов по сравнению со стандартным кодом. Установлены также другие свойства структуры генетического кода, вытекающие из полученных результатов. На основе построенной теории была поставлена (в 1992 году) и решена задача расчета генетического кода и первая полная публикация опубликована недавно: [26].

### I. INTRODUCTION

По результатам многолетних исследований автором была разработана математическая теория генетического кода, кратко представленная в статье N.N. Kozlov and T.M. Eneev. Fundamentals of a Mathematical Theory of Genetic Code. Doklady Mathematics 2017, Volume 95, № 2, pp. 144-146. Данная работа представляет собой расширенную версию этой статьи. Основное внимание будет уделяться вопросам, получившим строгое математическое обоснование. Мы пытались дать материал в виде, доступном для широкого круга математиков, поэтому введены два первых вводных пункта. Задача упрощается тем, что расширенные версии некоторые из главных

*Author:* Институт прикладной математики им. М.В. Келдыша РАН, Москва. e-mail: [gencod@keldysh.ru](mailto:gencod@keldysh.ru)

результатов были опубликованы в ряде статей, которые процитированы в списке литературы. Однако принципиально важно представить такие результаты в их взаимосвязи, ведь они касаются загадок одной и той же таинственной структуры - генетического кода. Следует отметить, что ни один из приведенных результатов не был обнаружен автором в других публикациях, причем как в отечественных, так и в зарубежных.

## II. Генетический Код

История открытия генетического кода достаточно подробно описана М. Ичасом [1, 2]- одним из участников пионерских исследований по этой проблеме. Он пишет: «... расшифровка биологического кода действительно революционизирующее событие, ее, быть может, уместно сравнить с другим событием, вызвавшим переворот в науке столет назад с появлением дарвиновского «Происхождения видов», [1]. Самым трудным в проблеме кода было понять, что код существует. На это потребовалось почти целое столетие. Отсчет его ведется от работы Менделя [3], который показал, что наследственные признаки передаются дискретными частицами, которые мы сегодня называем генами. Эта работа, как известно, почти не вызвала интереса. «Из всего того, что нам известно, складывается впечатление, что Менделю были, в общем-то, безразличны отклики на его работу. Опубликовав свой главный труд, он посчитал свой долг исполненным: если на нее не обратили внимания, то тем хуже для читателей, а не для автора», [2, стр. 142]. В 1900 году три независимых исследователя одновременно своими опытами подтвердили результаты, полученные Менделем. Только завершив работу, они узнали, что 34 года назад их опередил Мендель. После 1900 года генетика стала развиваться быстро и непрерывно.

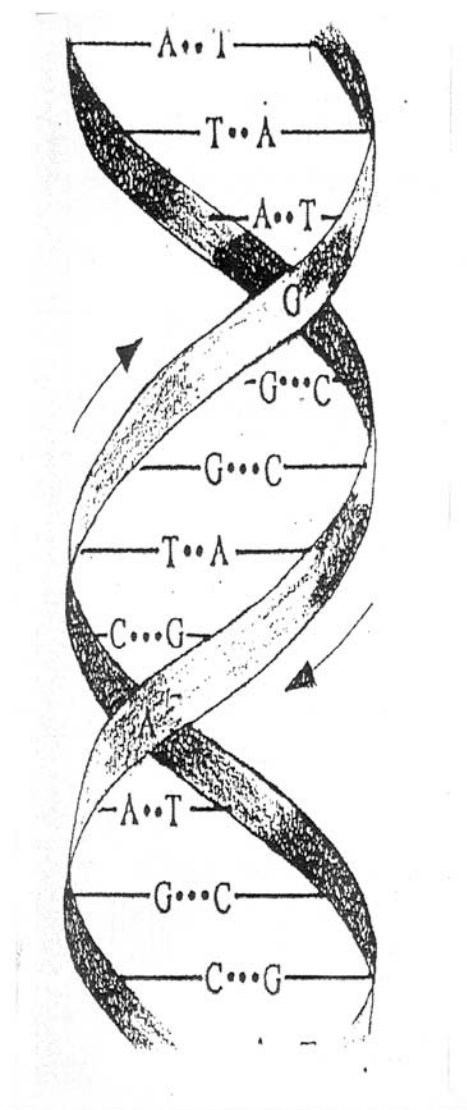
Впервые идея молекулярно-биологического подхода к проблемам генетики была сформулирована известным физиком Э. Шредингером в книге «Что такое жизнь? С точки зрения физика», [4], которая в оригинале увидела свет в 1945 году. На странице 28 читаем представление о коде (за 21 год до его окончательной разгадки!): «Называя структуру хромосомных нитей шифровальным кодом, мы подразумеваем, что всеохватывающий ум, вроде такого, который некогда представлял себе Лаплас и которому каждая причинная связь непосредственно открыта, мог бы, исходя из структуры хромосом, сказать, разовьется ли яйцо при благоприятных условиях в черного петуха или в крапчатую курицу, в муху или растение маиса, в рододендрон, жука, мышь или человека». Помимо этого и других блистательных предвидений следует отметить, что эта книга сыграла решающую роль в судьбе ряда физиков-теоретиков. Назову лишь две фамилии, о которых будет идти речь в дальнейшем. Это Ф. Крик, который в 1946 году оставил теоретическую физику и обратился к задачам биологии после прочтения этой книги. Его Нобелевская лекция была посвящена проблеме кода, а не структуре ДНК, за которую он был удостоен Нобелевской премии (F.Crick - Nobel Lecture, Dec. 11, 1962: On the Genetic Code, Internet). У истоков проблемы кода стоял также физик Г. Гамов, на которого Ф.Крик ссылается на первой странице указанной лекции. Однако на финальном этапе исследований именно

Ref

1. Ичас М. Биологический код/– М.: Мир, 1971, 359 с. англ. пер.: Ycas M. The biological code//Amsterdam – London. – 1969. – 359P.

биохимики экспериментально установили генетический код и указали его роль в биосинтезе белка. За этот результат Х. Корана, М. Ниренберг и Р. Холли получили Нобелевскую премию в 1968г.

Но сначала была решена проблема структуры ДНК. Аспиранту Д. Уотсону понадобилось примерно полтора года, чтобы совместно с руководителем Ф. Криком решить одну из важнейших проблем биологии, которая в настоящее время считается одной из главных фундаментальных проблем, решенных в прошлом столетии. Речь идет о структуре молекул ДНК, которую мир впервые увидел 25 апреля 1953 года: работа [5], объемом в одну (!) страницу журнала «Nature», поставила точку на дискуссии относительно роли ДНК в передаче наследственной информации. Описания, которые дают для ДНК сегодня, различны. Для наших целей достаточно упрощенного описания. Модель двойной спирали ДНК представляет собой две нити, закрученные друг относительно друга (рис. 1.).



*Рис. 1:* Модель двойной спирали ДНК. Чтение текста гена указано стрелками по одной цепи - сверху вниз, по другой - снизу вверх

Ref

5. Watson, J.D., Crick, F.H.C. A structure for Deoxyribose Nucleic Acid//Nature. - 1953. - V.171. - P. 737-738.

По сути дела, это – двойная винтовая линия, а не какая ни спираль. Алфавит ДНК содержит всего 4 буквы: А, С, G, Т. Это четыре нуклеотида: аденин, цитозин, гуанин, и тимин соответственно. Точки между этими буквами на рис. 1. указывают на количество водородных связей: две связи между А и Т и три - между С и G. Именно эта блестящая догадка Уотсона, который ввел эти комплиментарные пары [6], и позволила объяснить важнейшие свойства передачи наследственной информации. (Эти связи существуют между двумя спиралями ДНК). ДНК измеряют по-разному, в том числе, и количеством пар нуклеотидов. Например, для ДНК человека их около 3.2 миллиардов ([ftp://ftp.ncbi.nih.gov/refseq/H\\_sapiens/](ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/)).

Тайна гена была окончательно разгадана в 1966 году (к столетию работы Менделя [3]), когда в ходе экспериментальных исследований было окончательно установлено, что гены есть одонитевые участки ДНК и содержат информацию о белке в закодированном виде. Оказалось, что каждая из 20 аминокислот - элементов, из которых состоят все известные белки, кодируется определенными тройками нуклеотидов - кодонами или триплетами. Для четырех букв: А, С, G, Т, имеем 64 кодона: ААА, ААС, ААG, ..., ТТТ. Смысл всех этих кодонов был экспериментально установлен и представлен в таблице генетического кода, причем кодировка, которую выбрала природа, оказалась своеобразной. В табл. 1. она представлена полностью.

*Таблица 1:* Стандартный генетический код.

|    | 1      | 2 | 3        |
|----|--------|---|----------|
| 1  | Met    | 1 | ATG      |
| 2  | Trp    | 1 | TGG      |
| 3  | Phe    | 2 | TTY      |
| 4  | Tyr    | 2 | TAY      |
| 5  | His    | 2 | CAY      |
| 6  | Asn    | 2 | AAY      |
| 7  | Asp    | 2 | GAY      |
| 8  | Cys    | 2 | TGY      |
| 9  | Gln    | 2 | CAX      |
| 10 | Lys    | 2 | AAX      |
| 11 | Glu    | 2 | GAX      |
| 12 | Ile    | 3 | ATM      |
| 13 | Val    | 4 | GTN      |
| 14 | Pro    | 4 | CCN      |
| 15 | Thr    | 4 | ACN      |
| 16 | Ala    | 4 | GCN      |
| 17 | Gly    | 4 | GGN      |
| 18 | Ser    | 6 | TCN, AGY |
| 19 | Leu    | 6 | CTN, TTX |
| 20 | Arg    | 6 | CGN, AGX |
|    | ter(*) | 3 | TAX, TGA |

Примечание. Для каждой из аминокислот приводятся: 1 – стандартные трехбуквенные сокращения, 2 – число кодонов-синонимов, 3 – трехбуквенные нуклеотидные представления кодонов. Обозначение: X: А, G; Y: Т, С; М: Т, С, А; N: А, G, Т, С. В последней строке приводятся три терминаторных кодона – ter, каждый из которых обозначает останов синтеза белка.

Ref

3. Mendel G. Versuche über Pflanzenhybriden, Verhandl./Naturforsch. Ver. Brün. – 1866. – V.4, – P. 3–47.



Оказалось, что только две аминокислоты - метионин (Met) и триптофан (Trp) кодируются однозначно кодонами ATG и TGG соответственно. Все остальные аминокислоты кодируются более чем одним кодоном (это кодоны-синонимы), но не более чем шестью. Последнее наблюдается только для трех аминокислот: серин (Ser), лейцин (Leu), аргинин (Arg). Такие три кодировки названы нерегулярными, в отличие от 17-и других, регулярных для которых каждые 1-ая и 2-ая позиции одинаковы в соответствующем наборе кодонов-синонимов. Полное число смысловых кодонов или троек, кодирующих какую-либо аминокислоту, равно 61. Кодоны терминации *ter* (\*) не соответствуют никаким аминокислотам, каждый из них останавливает синтез белка. В виду важности этих

Кодонов при дальнейшем анализе выделим их в (1):

$$\text{ter: TAA, TAG, TGA.} \quad (1)$$

Укажем, что помимо вырожденности (т.е., когда одной и той же аминокислоте соответствуют, как правило, несколько кодонов-синонимов) важнейшим свойством кода является его универсальность: код одинаков для почти всех живых организмов. Однако к настоящему времени обнаружены ряд отклонений кода от стандартного, что является одной из наиболее загадочных особенностей кода (см. табл.2).

**Таблица 2:** Стандартный –  $K^0$  и нестандартные генетические коды  $K^1$  -  $K^{14}$  и их характеристики  $p$

| 1           | 2  | 3  | $p$ | $p$ |
|-------------|--|--|-----|-----|
| $K^0$       | The standard code  |  | 16  |     |
| $K^1$       | The Vertebrate Mitochondrial Code  | TGA( <i>ter</i> )→Trp, ATA(Ile)→Met, AGX(Arg)→ <i>ter</i>                | 7   |     |
| $K^2$       | The Invertebrate Mitochondrial Code  | TGA( <i>ter</i> )→Trp, ATA(Ile)→Met, AGX(Arg)→Ser                        | 7   |     |
| $K^3$       | The Echinoderm and Flatworm Mitochondrial Code   | TGA( <i>ter</i> )→Trp, AAA(Lys)→Asn, AGX(Arg)→Ser                        | 5   |     |
| $K^4$       | The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma / Spiroplasma Code | TGA( <i>ter</i> )→Trp  | 6   |     |
| $K^5$       | The Ciliate, Dasycladacean and Hexamita Nuclear Code   | TAX( <i>ter</i> )→Gln  | 5   |     |
| $K^6$       | The Euplotid Nuclear Code  | TGA( <i>ter</i> )→Cys  | 5   |     |
| $K^7$       | The Alternative Yeast Nuclear Code   | CTG(Leu)→Ser   | 16  |     |
| $K^8$       | The Ascidian Mitochondrial Code  | TGA( <i>ter</i> )→Trp, ATA(Ile)→Met, AGX(Arg)→Gly                        | 7   |     |
| $K^9$       | The Alternative Flatworm Mitochondrial Code  | TGA( <i>ter</i> )→Trp, AAA(Lys)→Asn, TAA( <i>ter</i> )→Tyr, AGX(Arg)→Ser | 0   |     |
| $K^{10}$    | Blepharisma Nuclear Code   | TAG( <i>ter</i> )→Gln  | 10  |     |
| $K^{11}$    | Chlorophycean Mitochondrial Code   | TAG( <i>ter</i> )→Leu  | 10  |     |
| $K^{12}$    | Trematode Mitochondrial Code   | TGA( <i>ter</i> )→Trp, AAA(Lys)→Asn, ATA(Ile)→Met, AGX(Arg)→Ser          | 6   |     |
| $K^{13}$    | Scenedesmus Obliquus Mitochondrial Code  | TAG( <i>ter</i> )→Leu, TCA(Ser)→ <i>ter</i>                              | 10  |     |
| $K^{14}$    | Thraustochytrium Mitochondrial Code  | TTA(Leu)→ <i>ter</i>   | 21  |     |
| Примечание. |  |  |     |     |

1- коды  $K^0$  -  $K^{14}$ , 2 – их названия, 3- отклонения от стандартного кода

Колонки 2 и 3 получены на основе: <http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=t>

### III. Перекрывающиеся Гены

Математический анализ структуры генетического кода, прежде всего, опирается на исследование некоторых необычных способов записи генов, кодирующих белки. Необычный способ записи генов состоит в том, что один и тот же участок цепи ДНК, кодирующий белок, может читаться со сдвигом фазы на +1 либо -1 нуклеотид либо с другой цепи ДНК (с подобными сдвигами либо без них). Иными словами один и тот же указанный участок ДНК может кодировать два, и более белков, вплоть до шести. Такие гены были названы перекрывающимися. Отметим, что как показывают эксперименты, такое чтение оказывается разрешенным лишь в некоторых случаях, а в подавляющем большинстве случаев существует запрет на указанные альтернативные чтения. Этот запрет состоит в том, что указанные сдвиги приводят к совершенно иным последовательностям кодонов отличным от исходной последовательности (когда сдвигов нет). Но было установлено, что в подобных альтернативных последовательностях непременно возникают какие-либо кодоны из (1); так устроен ген кодирующий белок, или так выбраны соответствующие кодировки аминокислот вследствие вырожденности кода. Из-за роли названных трех кодонов (они останавливают (блокируют) белковый синтез), белок при альтернативном чтении не синтезируется. Был сделан вывод о мощной биологической защите: природе не нужны эфемерные белки, она не синтезирует белки соответствующие сдвинутым позициям (например, если в ходе мутаций начальная точка сдвинется). Тем самым были введены в рассмотрение две рамки считывания - РС: открытая рамка считывания (ОРС) - последовательность кодонов, не содержащая кодонов терминации и блокированная РС - БРС, когда такие кодоны встречаются [7]. На рис. 2 белок соответствует ОРС, сдвинутые позиции нуклеотидных последовательностей, как +1 так и -1 - БРС.

|          |  |     |
|----------|--|-----|
| сдвиг -1 | AsnGlyGlyLeuLeu * * AlaGlu...  | БРС |
| сдвиг +1 | TrpArgLeuArgIleValGly * ...  | БРС |
| белок →  | MetGlnAlaCysTyrSerArgLeuLys...   | ОРС |
| ген →    | AATGGAGGCTTGCTATAGTAGGCTGAAG...  |     |
|          | <div style="display: flex; justify-content: space-around; align-items: center;"> <div>•</div> <div style="text-align: center;">↑</div> <div style="text-align: center;">↑</div> <div style="text-align: center;">↑</div> </div> <div style="display: flex; justify-content: space-around; align-items: center;"> <div></div> <div>С</div> <div>С</div> <div>С</div> </div> |     |

**Рис. 2:** Участок белковой последовательности (первая аминокислота - Met) закодирован в гене начиная с ATG (первый нуклеотид А в этом триплете помечен жирной точкой) и случаи сдвигов на +1 либо -1 нуклеотид

В сдвинутых состояниях получим другие последовательности кодонов (две другие РС), в каждой из которых будет неоднократно присутствовать кодоны терминации ter, которые на рис.2 помечены одним и тем же символом символом \*. Это будут две РС с блокировками - БРС. Можно видеть, что с помощью трех нуклеотидных замен на один и тот же нуклеотид С (позиции замен указаны под текстом гена), ни один из трех кодонов ter (символ \*) на заданном участке гена

Ref

7. Lewin B. Genes VI // Oxford University Press, - 1997.- 879 P.

не возникнет, причем при таких заменах белковая последовательность не изменится т.к. указанные три замены соответствуют трем заменам кодонов на их синонимы. Однако типичный ген устроен так, чтобы указанные сдвиги давали именно две БРС [7]. Оказалось, что лишь для перекрывающихся генов такого запрета не существует. Впервые этот эффект был установлен экспериментально в 1976 году в ходе исследований по чтению первого целого генома - вируса бактерии ΦX 174 [8]. После этих исследований их руководитель F.Sanger становится единственным в истории двукратным лауреатом Нобелевской премии по химии. F.Sanger проявил интерес к одной из первых наших работ В письме к нему я сформулировал новое свойство первого целого генома. Оно состоит в том, что для записи такого генома требуется использовать все 61 смысловых кодонов- это из-за перекрытий генов, впервые обнаруженные в этом геноме. Я просил представить этот результат в NAURE. Его ответ приводится в файле Sanger. Полная нуклеотидная последовательность кольцевой одноцепочечной ДНК ΦX174 (по-видимому фактор одноцепочности ДНК, установленный ранее экспериментально и явился решающим для чтения первого целого генома) содержит 5386 нуклеотидов [10], однако общее число аминокислотных остатков в последовательностях всех белков умноженное на 3 (с учетом не кодирующих участков) превышает это число нуклеотидов. Было показано, что ген E содержит 273 нуклеотида и локализуется внутри гена D [8]. Это первое экспериментально обнаруженное перекрытие представлено на рис.3.



|   | D   | ALA | CYS | VAL | TYR | GLY     |
|---|-----|-----|-----|-----|-----|---------|
|   | E   |     |     |     |     | MET VAL |
| THR LEU ASP PHE VAL GLY TYR PRO ARG PHE PRO ALA PRO VAL GLU PHE ILE ALA VAL                                     |     |     |     |     |     |         |
| ARG TRP THR LEU TRP ASP THR LEU ALA PHE LEU LEU LEU SER LEU LEU LEU PRO SER                                     |     |     |     |     |     |         |
| T A C G C T G G A C T T T G T G G A T A C C C T C G C T T C C T G C T G T T A T T G C T G C C T                 | 591 | 591 | 591 | 591 | 591 | 591     |
| ILE ALA TYR TYR VAL HIS PRO VAL ASN ILE GLN THR ALA CYS LEU ILE MET GLU GLY ALA                                 |     |     |     |     |     |         |
| LEU LEU ILE MET PHE ILE PRO SER THR PHE LEU LYS ARG PRO VAL SER TRP LYS ALA LEU                                 |     |     |     |     |     |         |
| C A T I G C T A T A T A T G T T C A T C C C G T C A C A C A T T C A A C G G C C T G T C T C A T G A A G G C G C | 641 | 641 | 641 | 641 | 641 | 641     |
| GLU PHE THR GLU ASN ILE ILE ASN GLY VAL GLU ARG PRO VAL LYS ALA ALA GLU LEU PHE                                 |     |     |     |     |     |         |
| ASN LEU ARG LYS THR LEU LEU MET ALA SER SER VAL ARG LEU LYS PRO LEU ASN CYS SER                                 |     |     |     |     |     |         |
| T G A A T T A C G G A A A C A T T A T A T G G C G T C G A G C G T C C G G T A A A G C C G C T G A T T G T T     | 701 | 701 | 711 | 721 | 731 | 741     |
| ALA PHE THR LEU ARG VAL ARG ALA GLY ASN THR ASP VAL LEU THR ASP ALA GLU GLU ASN                                 |     |     |     |     |     |         |
| ARG LEU PRO CYS VAL TYR ALA GLN GLU THR LEU THR PHE LEU LEU LEU THR GLN LYS LYS THR                             |     |     |     |     |     |         |
| C G C G T T A C C T T G C G T G T A C G C G C A G G A A A C A C T G A C G T T A C T G A C G C A G A A A A A     | 761 | 771 | 781 | 791 | 801 | 811     |
| VAL ARG GLN LYS LEU ARG ALA GLU GLY VAL   |     |     |     |     |     |         |
| CYS VAL LYS ASN TYR VAL ARG LYS GLU   |     |     |     |     |     |         |
| C G T G C G T C A A A A T A C T G T G C G G A G G A G T G A   | 821 | 831 | 841 |     |     |         |

*Рис. 3:* Первое генетическое перекрытие, обнаруженное экспериментально [8]. Рисунок представлен в формате публикации полного текста первого целого генома для бактериофага ФХ 174 [10]. Видим, что начиная с позиции 568 и до позиции 840, была установлена кодировка нового белка Е на участке нуклеотидной последовательности другого белка - белка D

 $R_{\text{ef}}$ 

10. Sanger F., Coulson A.R., Friedmann T., Air G.M., Barrell B.G., Brown N.L., Fiddes J.C., Hutchison C.A., III, Slocombe P.M., Smith M. The Nucleotide Sequence of Bacteriophage  $\Phi$ X174//J. Mol. Biol. - 1978. - V.125. - P.225-246.

В настоящее время считается, что перекрывающиеся гены представляют собой хотя и необычный, но все же довольно распространенный элемент организации генома. В расшифрованном человеческом геноме обнаружены множественные генетические перекрытия [11]- их оказалось около 1700. Накопленный обширный материал по генетическим перекрытиям выдвинул задачу их тщательного и всестороннего анализа. Остановимся на некоторых результатах, полученных нами на основе математического анализа.

Можно видеть, что различающихся случаев перекрытий генов, разрешенных структурой ДНК, всего 5 (рис.4), из которых первые два относятся к перекрытиям генов из одной и той же цепи ДНК, а оставшиеся 3- к перекрытиям генов, взятых из разных цепей ДНК.

Перекрытия генов из одной цепи ДНК

сдвиг -1

B<sub>11</sub>→... LysIleTyr...  
B<sub>12</sub>→... GluAsnLeuPhe...  
...GAAAATTTATTT...

сдвиг +1

B<sub>21</sub>→ MetLeu...  
B<sub>22</sub>→... GlnCys...  
...CAATGCTA...

1

2

Перекрытия генов из разных цепей ДНК

сдвиг -1

B<sub>31</sub>→... ValAsnAsn...  
...GTGAATAACA...  
...GCACTTATTGTA...  
...ThrPheLeuMet...←B<sub>32</sub>

сдвиг 0

B<sub>41</sub>→ MetLeuVal  
...ATGCTCGTT...  
...TACGAGCAA...  
...HisGluAsn...←B<sub>42</sub>

сдвиг +1

B<sub>51</sub>→... AsnPheHisGlu...  
...AATTCCACGAG...  
...TTAAAGGTGCTC...  
...AsnGlyArg... ←B<sub>52</sub>

3

4

5

**Рис. 4:** Пять возможных случаев перекрытий генов, соответствующих одной (1,2) либо двум цепям ДНК (3-5). Чтение текстов при этом осуществляется в разных направлениях (указано стрелкой): слева направо для B<sub>11</sub>, B<sub>12</sub>, B<sub>21</sub>, B<sub>22</sub>, B<sub>31</sub>, B<sub>41</sub>, B<sub>51</sub> и справа налево для B<sub>32</sub>, B<sub>42</sub>, B<sub>52</sub>. В этих фрагментах присутствуют только канонические пары ДНК: CG и AT

На рис.4. представлены лишь небольшие фрагменты реальных перекрытий, причем полная протяженность некоторых из них достигает почти 1300 нуклеотидов. Кроме того, суммарная протяженность перекрытий может достигать более половины размера генома (вирус GSHV).

Следует подчеркнуть, что именно анализ множественных взаимосвязей кодонов в генетических перекрытиях и является главным инструментом проведенных исследований.

#### IV. Вырожденность Кода

Одна из задач, которая была поставлена нами, относится к фундаментальной проблеме генетического кода: зачем нужна вырожденность кода, когда для одной и той же аминокислоты имеют место, как правило, более одной кодировки, вплоть до 6-и кодировок. Было изучено участие всех смысловых кодонов в перекрытиях генов. Оказалось, что существует множество геномов с перекрытиями, в которых непременно должны участвовать все 61 смысловых кодонов, и при исключении хотя бы одного кодона обнаруженная в экспериментах запись генетических перекрытий представляется невозможной. Одним из таких геномов является первый целый геном для вируса бактерии ΦX174, содержащий перекрытия для 814 нуклеотидов [10]. Наша статья [9], а также статья, сопутствующая ей, были процитированы не менее чем по 100 раз каждая. (см. файл 300 citation).

#### V. Нерегулярности Кода

Следующая наша задача была связана с анализом кодонов, отклоняющихся генетический код от однородной структуры. Это одна из наиболее загадочных особенностей генетического кода. Как показал анализ добавочные кодонные представления из табл.1, или нерегулярности, для Ser это AGY, для Leu это TTX, для Arg это AGX, позволяют в принципе «организовать» перекрытия в ряде геномов либо существенно расширить диапазоны генетических перекрытий как для двойных (для генома ΦX 174 в 7 раз), так и для тройных (для HIV-2 [12] в 5 раз) перекрытий, если бы они были организованы с использованием однородного кода, или кода без нерегулярностей.

Потенциальные возможности кода для построения множества генетических перекрытий.

Далее нами была поставлена задача о том, каков же потенциал генетического кода для создания всех указанных случаев перекрытий. Ответ оказался следующим - феноменальный потенциал!, Этот результат оценил Лауреат Нобелевской премии de Duve to me (see the file) Оказалось, что только 16 пар аминокислот из возможных 400 могут создавать препятствия для построения всех 5-и случаев перекрытий. Это пары аминокислот всего для трех случаев перекрытий:

в случае 2 это 5 пар:

MetMet, MetAsn, MetLys, MetIle, MetThr, (2)

в случае 3 это 6 пар:

$$\text{PheTyr, TyrTyr, HisTyr, AsnTyr, AspTyr, CysTyr,} \quad (3)$$

в случае 5 это 5 пар:

$$\text{PheMet, PheAsn, PheLys, PheIle, PheThr.} \quad (4)$$

Иными словами создается впечатление, что генетический код как бы подобран под перекрытия. Так ли это? Ответ на этот вопрос будет дан ниже.

Итак, нами установлена первая интегральная характеристика генетического кода, которую обозначим через  $p$  и которая равна 16 для стандартного кода:

$$p = 16 \quad (5)$$

## VI. Общее Свойство Всех Природных Кодов

Полученный результат привел к остановке новых задач. Каково значение  $p$  для нестандартных (девиантных) кодов, число которых равно 14 и продолжает расти? Отметим, что первый нестандартный код, был обнаружен в 1979 году в клетке человека в отдельной органелле – в митохондрии: гены митохондриальной ДНК – mtДНК были записаны таким кодом [13]. Переосмысленными оказались всего 4 кодона. Расчеты показали, что значение  $p$  для всех 14 девиантных кодов не превышает значения 22 или около 5% полного числа пар аминокислот[14], см. Табл.2. При этом запретными оказались все те же самые три случая перекрытий, как и для стандартного кода и кроме того был обнаружен один код с нулевым значением  $p$ . Таким образом все природные генетические коды имеют малое число запретов на построение генетических перекрытий. Это может рассматриваться как общее свойство всех природных кодов, известных к настоящему времени. Возникает вопрос: почему природные коды этому соответствуют, в то время как число записей генов с перекрытиями неизмеримо меньше чем обычных неперекрывающихся генов и какова роль переосмысленных кодонов? Оба эти вопроса были решены.

## VII. Одна Математическая Аналогия

При решении первого из этих вопросов была установлена важная математическая аналогия между перекрытиями генов из разных цепей ДНК и важнейшими структурными единицами. Как известно, ген считывается с ДНК, происходит в его тексте множественная модификация нуклеотидов Т на U (урацил) и образуется матричная РНК, которая в свою очередь структурируется. Важнейшими элементами такой структуры являются стебли – фрагменты, содержащие связи подобные связям в ДНК. На рис. 5 дается один из стеблей самой известной вторичной структуры матричной РНК- RNA MS2 [15], которая содержит более 130 стеблей. Структура была полностью нами проанализирована, и полученные результаты представлены в моей монографии [25].

Ref

13. Batrell B.G., Bankier A.T., Drouin J. A different genetic code in human mitochondria//Nature. – 1979. – V. 282. – P.189-194.



А.

AspProArgLeuTyrLysValTrpVal  
3022 GAUCCACGCCUCUAUAAGGUGUGGGUC 3048

Б.

→  
AspProArgLeu  
3022 GAUCCACGCCUCU  
          |||||A  
3048 CUGGGUGUGGAAU  
          ValTrpValLys ←

*Рис. 5:* Стебель вторичной структуры матричной RNA MS2. Вверху (А) дан линейный текст – это фрагмент в диапазоне 3022-3048. а заштрихованные участки соответствуют стеблям вторичной структуры. Под текстом (Б) указан фрагмент вторичной структуры. Показано присутствие неканонической пары GU (они были обнаружен экспериментально в структурах РНК), помимо канонической CG в ДНК, и аналога канонической -AU

На рис.5 изображен один стебель этой вторичной структуры (Б), представлен также фрагмент А, в диапазоне 3022-3048 первичной структуры в стилистике процитированной статьи. Однако во фрагменте Б, дана запись не только нуклеотидов, но и соответствующих аминокислот согласно А. Стебель из рис.5Б соответствует перекрытию фрагментов как будто бы взятых из разных цепей ДНК. Направление чтения на стебле Б (стрелки) становится различным и этот фрагмент вторичной структуры эквивалентен перекрытию заштрихованных в А участков, взятых из разных цепей ДНК (случай перекрытия 4). Однако здесь нет никаких разных цепей. Этот эффект - поворот исходного направления чтения (→) (нижняя строка рис. 5Б (←)) обусловлен присутствием, так называемой шпильчатой петли UCUAUAA – последовательность не входящую в стебель.

Какова же роль малости первой характеристики  $p$ ? Дело в том, что ее малое значение позволяет строить феноменальное разнообразие генетических перекрытий генов, и тем самым это позволяет строить феноменальное разнообразие вторичных структур, в том числе также функционально значимые участки вторичных структур мРНК для всех генов, записанных как стандартным, так и любым из известных девиантных кодов. Увеличение указанной характеристики для кода, отклоненного от стандартного (например, на порядок, как это было показано для гипотетического кода из монографии), приводит к значительному сокращению такого разнообразия. Тем самым установлено, что малое значение  $p$  выполняет две функции: позволяет строить как феноменальное множество генетических перекрытий, так и феноменальное разнообразие вторичных структур матричных РНК для всех генов.

# VIII. О Роли Переосмысленных Кодонов

Рассмотрим теперь вопрос о роли переосмысленных кодонов. Нами был поставлен вопрос о возможной взаимосвязи ограничений на перекрываемость (2)-(4) с вариабельностью кода, наблюдаемой у ряда организмов. Анализ показал, что такая взаимосвязь существует, и она выражается в том, что для ряда девиантных кодов (примеры для некоторых из них, обнаруженных в митохондриальных ДНК, представлены на рис.6), природные переосмысления кодонов приводят к возможности построения генетических перекрытий, запретных для стандартного кода.

## 1 (Human)

ATФаза6→MetAsn...  
URFA6L→... Trp...  
... ATGAA...

8528

## 2 (Drosophila yakuba)

ATФаза6→MetMet...  
URFA6L→... Trp...  
... ATGATG...

4067

## 3 (Paracentrotus lividus)

ATФаза6→MetThrMetThr...  
ATФаза8→... TrpGlnTrp...  
... ATGACAATGAC...

8680

## 4 (Apis mellifera ligustica)

ATФаза6→MetLys...  
ATФаза8→... Trp...  
... ATGAA...

4585

**Рис. 6:** Фрагменты генетических перекрытий, обнаруженные в митохондриях четырех организмов, гены которых записаны кодами, отклоненными от кода стандартного. Это перекрытия в одной цепи ДНК. Фрагменты и названия белков даны по публикациям[16]-[19]. Число под нуклеотидом указывает его номер в геноме

В каждом из четырех фрагментов перекрытий, приведенных на рис. 3 показана роль одной и той же перестановки: TGA(ter)→Trp. Эта природная перестановка наблюдается для трех девиантных кодов, которым соответствуют приведенные фрагменты соответственно; второй и четвертый фрагмент записаны одним и тем же

Ref

19. Crozier R.H., Crozier Y.C. The Mitochondrial Genome of the Honeybee *Apis mellifera*: Complete Sequence and Genome Organization // Genetics. – 1993. – V.133 – P.97 – 117.

девиантным кодом. Причем указанная перестановка присутствует во всех трех девиантных кодах. Оказалось, что такая перестановка делает возможными перекрытия для пар MetAsn (рис.6.1, этот случай соответствует ДНК митохондрии человека), MetMet (рис.6.2), дважды MetThr (рис.6.3) и MetLys, которые являются запретными для стандартного кода, см. (2) Указанные пары нуклеотидов и переосмысленные кодоны выделены. Тем самым размер геномов сокращается за счет возможности построения перекрытий генов, которые невозможны для стандартного кода. Такое сокращение для живой клетки может быть достаточно большим, т.к. число митохондрий, как правило, больше 1 и может достигать миллиона. Проведенное исследование процитировано не менее 100 раз ( см файл 300 citation).

Полученные результаты позволили обратиться к анализу экспериментальных данных по всем девиантным генетическим кодам, или кодам, отклоненным от кода стандартного. Однако в рамках генетических перекрытий мне не удалось объяснить функциональную значимость всех переосмысленных кодонов во всех девиантных кодах. Требуемое решение было найдено при исследовании областей ДНК, где гены не перекрываются, а таких генов – подавляющее большинство.

### IX. О Двух Интегральных Характеристиках Кода

Речь идет о природной блокировке генов, когда все 5 последовательностей кодонов, альтернативных последовательности гена ли рамок считывания - РС, содержат многократные остановки синтеза белка, или кодоны из набора (1). На рис.7 это показано для участка гена. См.также ранний упрощенный рис.2.

```
OPC0      MetSerIleLysLeuSerTyrArgGluSerPheSerIleLeuGluGluVal...

BPC1 (-1) TyrGluHis * Thr * Leu * ArgValIle * TyrIleArgGlyGly...

BPC2 (+1)      * AlaLeuAsnLeuValIleGluSerHisLeuValTyr * ArgArgPhe...

→          TATGAGCATTAACCTTAGTTATAGAGAGTCATTTAGTATATTAGAGGAGGTTTA...
←          ATACTCGTAATTTGAATCAATATCTCTCAGTAAATCATATAATCTCCTCCAAAT...

BPC3 (-1) IleLeuMetLeuSerLeu * LeuSerAspAsnLeuIleAsnSerSerThr ...

BPC4 (0)      HisAlaAsnPheLysThrIleSerLeu * LysThrTyr * LeuLeuAsn...

BPC5 (+1)      SerCys * Val * AsnTyrLeuThrMet * TyrIleLeuProProLys...
```

**Рис. 7:** Шесть РС для фрагмента гена (начало с кодона ATG(Met), направление чтения указано стрелкой → из которых одна РС является открытой – OPCO (в ней 17 смысловых кодонов), а 5 РС – альтернативные РС являются заблокированными: BPC1- BPC5. При этом BPC3-BPC5 соответствуют другой цепи ДНК и чтение последовательностей кодонов осуществляется в обратном направлении (←). В скобках указан сдвиг в нуклеотидах относительно OPCO. Символом \* были обозначены каждый из трех кодонов ter из (1)

Установлен потенциал такой блокировки для стандартного генетического кода: было показано, что для такого кода в процессе блокировки участвуют лишь не более 210 пар аминокислот из 400 возможных. Показано также, что 31 пара из них дает неизбежные блокировки, которые имеют место при любых кодировках аминокислот, содержащихся в этих парах. Тем самым была введена в рассмотрение вторая интегральная характеристика генетического кода [20] содержащая две компоненты:

$$q: q_{min}=31, \quad q_{max}=210 \quad (6)$$

С учетом этого, помимо неизбежных, были введены в рассмотрение возможные блокировки, возникающие для ограниченного числа кодировок; эти блокировки - главная составляющая 210 блокировок, куда включены также 31 неизбежная блокировка.

| V a l L y s |   |   |   |     |         |   | V a l G l u |   |   |   |       |         |   |
|-------------|---|---|---|-----|---------|---|-------------|---|---|---|-------|---------|---|
| G T N A A X |   |   |   |     |         |   | G T N G A X |   |   |   |       |         |   |
| N:          | A | - | G | T A | AA X    | → | N:          | A | - | G | T A G | A X     | → |
| N:          | C | - | G | T C | A A X   |   | N:          | C | - | G | T C   | G A X   |   |
|             |   |   | A | G T |         | ← |             |   |   | C | A G   | C T Y   |   |
| N:          | G | - | G | T G | AA X    | → | N:          | G | - | G | T G   | G A X   |   |
|             |   |   |   |     |         |   |             |   |   | C | A C   | C T Y   |   |
| N:          | T | - | G | T   | T A A X | → | N:          | T | - | G | T     | T G A X | → |
|             |   |   | C | A A | TT Y    | ← |             |   |   | C | A A   | C T Y   |   |

**Рис. 8:** Неизбежные блокировки для пары аминокислот ValLys (слева) и возможные блокировки для пары аминокислот ValGlu (справа). Кодоны ter из (1) заштрихованы. Стрелки указывают направление чтения. Полный перечень неизбежных блокировок представлен в табл.3

**Таб 3:** Полный перечень пар аминокислот, вызывающих неизбежную блокировку (столбец 1), с указанием номеров РС, которые блокируются (столбец 2)

| №  | 1      | 2   | №  | 1      | 2       |
|----|--------|-----|----|--------|---------|
| 1  | MetMet | 2   | 17 | IleMet | 2,5     |
| 2  | MetAsn | 2   | 18 | ValMet | 2,5     |
| 3  | MetLys | 2   | 19 | LeuMet | 2,4,5   |
| 4  | MetIle | 2   | 20 | IleAsn | 1,2,5   |
| 5  | MetThr | 2   | 21 | ValAsn | 1,2,5   |
| 6  | PheTyr | 3   | 22 | LeuAsn | 1,2,4,5 |
| 7  | TyrTyr | 3   | 23 | IleLys | 1,2,5   |
| 8  | HisTyr | 3   | 24 | ValLys | 1,2,5   |
| 9  | AsnTyr | 3   | 25 | LeuLys | 1,2,4,5 |
| 10 | AspTyr | 3   | 26 | IleIle | 2,5     |
| 11 | CysTyr | 3   | 27 | ValIle | 2,5     |
| 12 | PheMet | 5   | 28 | LeuIle | 2,4,5   |
| 13 | PheAsn | 1,5 | 29 | IleThr | 2,5     |
| 14 | PheLys | 1,5 | 30 | ValThr | 2,5     |
| 15 | PheIle | 5   | 31 | LeuThr | 2,4,5   |
| 16 | PheThr | 5   |    |        |         |

R<sub>ef</sub>

20. Козлов Н.Н. Интегральные характеристики генетического кода. Математическое моделирование, т.22, №9, С.51-66, 2010.; англ.пер.: N.N. Kozlov. Integral characteristics of genetic code. Mathematical Models and Computer Simulations April 2011, Volume 3, Issue 2, pp 123-134.

В столбце 2 этой таблицы указываются номера РС для которых (потенциально) может иметь место блокировка. Число таких РС, в зависимости от пары, меняется от 1 до 4. Представим множество этих пар в виде двух подмножеств. К первому отнесем пары неизбежно блокирующие одну и ту же РС. Таких пар оказалось всего 16- это первые 16 пар из табл.3. Имеем РС2 неизбежно блокируют 5 пар аминокислот, которые совпадают с набором (2), определенным выше, поскольку в РС2 образуется кодон ter – TGA. Для РС3 имеем 6 блокирующих пар, совпадающих с набором (3), а в РС3 образуется один из кодонов ter: TAA или TAG. Особо следует сказать о блокировке для 5-и пар с номерами 12-16, которые совпадают с парами из (4), определенные выше. Отметим лишь, что каждая из этих последних пар неизбежно блокирует РС5, поскольку в РС5 образуется один из кодонов ter: TAA или TGA. Однако в парах PheAsn, PheLys помимо РС5 в табл.3 указывается также РС1. Однако последняя РС не соответствует неизбежной блокировке в отличие от РС5. Таким образом, пары 1-16 из табл.3 образуют набор пар аминокислот, запретных для перекрытий двух генов и были установлены выше. На основе п.2 была введена в рассмотрение числовая характеристика, которая была обозначена буквой  $p$ , и которая соответствует числу различающихся блокировочных пар из (2)-(4), имеем значение  $p$  из (5.) Тем самым использование подмножества неизбежных блокировок позволяет установить связь изучаемых характеристик: имеет место неравенство:

$$0 \leq p \leq q_{\min} \quad (7)$$

Иными словами интегральная характеристика генетического кода  $p$  не является независимой, а определяется выбором характеристики  $q_{\min}$ , которая используется в решении совершенно другой задачи – в блокировке не перекрывающихся генов.

При рассмотрении только одной задачи – перекрытия пар генов можно было бы сделать вывод о том, что генетический код был «выбран» под перекрытия генов, поскольку только 16 пар из 400 возможных запретны для перекрытий. Это справедливо для всех 5 способов парных перекрытий генов, разрешенных структурой ДНК. Однако, при рассмотрении двух задач и двух интегральных характеристик  $p$  и  $q$  оказалось, с учетом связи (7), что генетический код был ориентирован на «выбор» двухкомпонентной интегральной характеристики (6), одна из компонент которой согласно неравенству (7) определяет область значения другой интегральной характеристики  $p$ . Таким образом, малость интегральной характеристики  $p$  является следствием более общего принципа, связанного с выбором всего набора неизбежных пар аминокислот, создающих блокировки. Т.е. пары аминокислот соответствующие характеристике  $p$  могут быть «выбраны» только из этого ограниченного набора соответствующего  $q_{\min}$ , а не из полного набора 400 пар аминокислот. По какому критерию произошел предполагаемый «выбор» генетического кода пока остается неясным.

В связи с анализом задачи о блокировках, нами были исследованы ряд геномов с общим числом генов более 200 000. Рассказ об этой работе требует отдельного рассмотрения, более подробный анализ представлен в [21]. Отметим лишь один



результат, полученный для генома человека. Из 25613 генов в этом геноме три гена вовсе не содержат никаких блокировок: Для каждого из них рисунки аналогичные рис.8 не содержат ни одного кодона терминации, ни в одной из пяти альтернативных РС. Это гены MT1G и MT1M из хромосомы 16 и KLK8 из хромосомы 19 (см. рис.9).

KLK8 M G R P R P R A A K T W M F L L L L G G A W A G R F W R P P G V \*

(-1) G T P P T S C G Q D V D V P A L A G G S L G R A I L E A P W C V

(+1) W D A P D L V R P R R G C S C S C W G E P G Q G D S G G P L V C

→ atgggacgcccccgacctcgtcgggccaagacgtggatgttctctgctcttctgggggagcctgggcagggcgattctggaggccccctggtgtgtg

← tacctctcgggggctggagcacgcgggttctgcacctacaaggacgagaacgacccccctcggacccgtcccgctaagacctccgggggaccacacac

(-1) P V G G V E H P W S T S T G A R A P P L R P L A I R S A G Q H T

(0) H S A G S R T R G L R P H E Q E Q Q P S G P C P S E P P G R T H

(+1) P R G R G R A A L V H I N R S K S P P A Q A P R N Q L G G P T H

**Рис. 9:** Ген KLK8 из человеческого генома и 5 альтернативных РС. Каждая из этих РС не содержит ни одного кодона терминации.(сравни с рис.8)

В связи с полученным результатом было выдвинуто ряд гипотез, из которых простейшая – это не что иное, как случаи перекрытий 6 генов.

#### Х. Математический Анализ Девиантности Кода

Задача исследования блокировок для девиантных кодов позволила завершить анализ одной из важнейших фундаментальных проблем, связанных с ролью всех переосмысленных кодонов. Данные были получены для mtДНК двух организмов: H.Sapiens (код K<sup>1</sup>) и A.Mellifera (код K<sup>2</sup>). Обратимся к Табл.4.

**Таб 4:** Сводная таблица участия переосмысленных кодонов в двух функциях: в блокировке (в столбце 1 знак + соответствует участию, знак– неучастию), либо в генетическом перекрытии (столбец 2). Данные были получены для mtДНК двух организмов: H.Sapiens (код K<sup>1</sup>) и A.Mellifera (код K<sup>2</sup>). Переосмысленные кодоны ука заны в столбце: отклонения от стандартного кода

| Организм                      | Отклонения от стандартного кода | 1 | 2 |
|-------------------------------|---------------------------------|---|---|
| H.Sapiens (K <sup>1</sup> )   | ATA(Ile)→Met                    | + | – |
|                               | TGA(ter)→Trp                    | + | + |
|                               | AGA(Arg)→ter                    | + | – |
|                               | AGG(Arg)→ter                    | + | – |
| A.Mellifera (K <sup>2</sup> ) | ATA(Ile)→Met                    | + | + |
|                               | TGA(ter)→Trp                    | – | + |
|                               | AGA(Arg)→Ser                    | + | – |
|                               | AGG(Arg)→Ser                    | + | – |

Из таблицы следует, что для mtДНК H.Sapiens(K<sup>1</sup>) имеет место участие всех переосмысленных кодонов(ATA(Ile)→Met, TGA(ter)→Trp, AGA(Arg)→ter, AGG(Arg)→ter) в процессе блокировки, а также переосмысленный кодон TGA(ter)→Trp участвует кроме того и в перекрытии генов, запретном для стандартного кода. Для mtДНК A.Mellifera (K<sup>2</sup>) в процессе блокировки участвуют все переосмысленные кодоны (AGA(Arg)→Ser, AGG(Arg)→Ser, ATA(Ile)→Met), кроме того в перекрытиях участвуют два подобных кодона: TGA(ter)→Trp и ATA(Ile)→Met. Тем самым нами показано, что все переосмысленные кодоны в каждом из указанных двух девиантных кодов были использованы либо в процессе блокировки, либо в процессе перекрытий генов (разумеется, запрещенных для стандартного кода), либо и в том и другом процессах.

### XI. Произволен Ли Код?

Полученные результаты приводит к тому выводу, что отклонения кода от стандартного не носит случайного характера, а несут совершенно четкую функциональную нагрузку (ср. «Переосмысление кодонов указывают на то, что в генетическом коде митохондрий могут происходить случайные перемены», см. [22]. В последней монографии также читаем «The code seem to have been selected arbitrarile[22]...», («Код, по-видимому, был «выбран, произвольно...»). Из приведенных результатов следует, возможность для всех смысловых кодонных семейств, практически беспрепятственно записывать две белковые последовательности одним и тем же участком ДНК, причем для этого может быть использован наиболее благоприятный (по сочетанию аминокислот в перекрытии) один из 5-и вариантов такой компактной записи генов (5 случаев перекрытий). Категорический запрет существует не более чем для около 5% пар аминокислот, как для стандартного кода, так и для всех 14-и известных на сегодня нестандартных кодов. Т.е. 15 таблиц кодов удовлетворяют одному и тому же общему свойству. Это не оставляет никаких шансов для какой-либо произвольности, случайности.

### XII. О Множествах Элементарных Перекрытий

Основные рабочие множества в данной теории – это множества элементарных генетических перекрытий, полностью представленные в Приложении. Элементарные перекрытия- это перекрытия по одиночным аминокислотам. Такие множества многократно использовались в ходе построения данной теории. Прежде всего-в доказательстве теоремы для генетического кода и далее эти множества были модифицированы 14 раз (по числу девиантных кодов) для получения первой интегральной характеристики этих 14 кодов, которые представлены на рис.2. Важнейший этап исследований был связан с математическим анализом неоднозначностей в этих множествах. Компонентами исследуемого множества являются элементарные генетические перекрытия- это перекрытия для одиночных аминокислот. Анализ показал, что множество содержат особенности, которые были названы неоднозначностями. Исследуемые неоднозначности, соответствуют случаям, когда для одной и той же пары аминокислот имеет место более одного элементарного

перекрытия. Как и всякие особые случаи в математике этот феномен привлек наше пристальное внимание. Принципиально важно, что полученные результаты применимы для всего многообразия живой природы, белки которой записаны практически одним и тем же генетическим кодом. Анализ показал, что неоднозначности имеют место только для случаев перекрытий генов, принадлежащих различным цепям ДНК. Число неоднозначных элементарных перекрытий оказалось относительно небольшим – всего 6. Проведенное исследование позволило выявить три функции возможного использования указанных неоднозначностей. Этих функций оказалось три [23]. Одна из функций неоднозначностей исследовалась в новой модели, предложенной ранее автором. Она состоит в том, что перекрытия пар генов, принадлежащие различным цепям ДНК, являются математическими аналогами стеблей вторичной структуры матричной РНК. Показывается, что за счет неоднозначности можно «регулировать» величину свободной энергии стебля функционально значимую биохимическую характеристику [24]. Теперь о других двух характеристиках согласно [23]. Первая из них связана с решением задачи о потенциальных позициях молчащих мутаций для случаев перекрытий генов, принадлежащих различным цепям ДНК. Вторая связана с расширением возможности построения множеств генетических перекрытий более двух генов; анализируются структуры возможных перекрытий 6-и генов в геноме человека.

Изучение пространственной структуры ДНК показало, что кроме трех семейств форм двойных спиралей с антипараллельной ориентацией нитей возможно образование двойных спиралей ДНК с параллельной взаимной ориентацией нитей. Математический анализ подобных случаев представлен в п.4.4 из [25], где были проанализированы все три новых случая перекрытий пар генов.

### XIII. Заключение

На основе построенной теории была решена одна из прорывных задач – задача расчета генетического кода. Подобные задачи в мире неизвестны и могли быть поставлены только в 21 веке. Один из подходов к решению этой задачи дается в статье [26]. Впервые представлено подробное описание метода расчета генетического кода, идея которого впервые опубликована ранее, а выбор одного из главных множеств для расчетов опирался на статью. Такое множество соответствует полному набору аминокислотных представлений из множества тройных перекрытий генов, принадлежащих одной и той же цепи ДНК. Отдельный вопрос был связан с начальным приближением, запускающим итерационный процесс поиска всех кодов по представленным начальным данным. Математический анализ показал, что указанное множество содержит некоторые неоднозначности, которые были установлены на основе предложенного нами сжатого представления множества. В итоге разработанный метод расчета сводился к двум главным этапам исследований, где на первом этапе в расчетах были использованы только области однозначности. Предложенный подход позволил значительно сократить объем вычислений на каждом шаге в этой сложнейшей дискретной структуре. Математическая теория генетического кода непрерывно развивается. Укажем на последние две работы в этом направлении. Первая относится к анализу потенциала генетического кода для построения перекрытий шести генов [27].

Ref

23. Н.Н. Козлов. Три функции неоднозначностей, порождаемых генетическим кодом. Математическое моделирование, т. 27, № 2, с. 74 - 84, 2015.

Вторая связана с проведенным математическим анализом, который позволил сформулировать свойство тройки терминаторных кодонов стандартного генетического кода, при сравнении с другими теоретически возможными тройками [28].

Работа выполнена при финансовой поддержке РФФИ- Российского фонда фундаментальных исследований (коды проектов 16-01-00018, 17-01-00053).

### Литература

1. Ичас М. Биологический код/– М.: Мир, 1971, 359 с. англ. пер.: Ycas M. The biological code//Amsterdam – London. – 1969. – 359P.
2. Ичас М. О природе живого: механизмы и смысл//М.: Мир, 1994. – 496с. англ. пер.:Martynas Ychas Meaning and Mechanisms. 1994.
3. Mendel G. Versuche über Pflanzenhybriden, Verhandl//Naturforsch. Ver. Brünn. – 1866. – V.4, – P. 3–47.
4. Шредингер Э. Что такое жизнь? С точки зрения физика//М.: Атом-издат, 1972. – 88 с. англ. пер.: E. Schrödinger. What is Life? The Physical Aspect of the Living Cell. — Cambridge: University Press, 1944.
5. Watson, J.D., Crick, F.H.C. A structure for Deoxyribose Nucleic Acid//Nature. – 1953. – V.171. – P. 737-738.
6. Уотсон Д. Двойная спираль. Воспоминания об открытии структуры ДНК//М.: Мир, – 1969. – 152с. англ. пер.: James D. Watson: The double helix. A personal Account of the Discovery of the Structure of DNA. Atheneum New York, 1968.
7. Lewin B. Genes VI // Oxford University Press, – 1997.- 879 P.
8. Barrell B.G., Air G.M. and Hitchison C.A. III. Overlapping genes in bacteriophage ΦX174//Nature. – 1976. – V. 264. – P. 34-41.
9. Н.Н.Козлов. О востребованности каждого из 64 кодонов в генетических перекрытиях. ДАН. 1999. Т. 367. № 4. С. 544-547. англ. пер.: N.N. Kozlov. Involvement of Each of 64 Codons in Gene Overlappings. Doklady Biochemistry 367, 126-128 (1999).
10. Sanger F., Coulson A.R., Friedmann T., Air G.M., Barrell B.G., Brown N.L., Fiddes J.C., Hutchison C.A., III, Slocombe P.M., Smith M. The Nucleotide Sequence of Bacteriophage ΦX174//J. Mol. Biol. – 1978. – V.125. – P.225-246.
11. Nakayama T., Asai S., Takahashi Y., Nishida Y. Overlapping of Genes in the Human. Genome //Nevill Juvenile Bonfire Society 2007, vol.3, no. 1, p. 14-19.
12. Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L., Alizon M. Genome organization and Transactivation of the Human Immunodeficiency Virus Type 2// Nature. – 1987. – V.326. – P. 662 - 669.
13. Barrell B.G., Bankier A.T., Drouin J. A different genetic code in human mitochondria//Nature. – 1979. – V. 282. – P.189-194.
14. Н.Н. Козлов. Одна интегральная характеристика множества генетических кодов. Свойство всех известных природных кодов. Математическое моделирование, т.26, №5, с.113-125, 2014; англ.пер.: N.N.. Kozlov. One integral characteristic of the set of genetic codes. The property of all known natural codes. Mathematical Models and Computer Simulations November 2014, Volume 6, Issue 6, pp 622-630
15. Fiers W., Contreras R., Duerinck F., Haegeman G., Iserentant D., Merregaert J., W. Min Jou, Molemans F., Raeymaekers A., A. Van den Berghe, Volckaert G. &

Ref

28. Н.Н. Козлов, Е.И. Кугушев, Т.М. Энеев. Математический анализ кодонов, останавливающих синтез белка, ДАН 2017 Т.477. № 3.

- Ysebaert M. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene//Nature. – 1976. – V.260. – P. 500-507.
16. Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H. , Staden R., and Young, I.G. Sequence and organization of the human mitochondrial genome//Nature. – 1981. –V. 290. – P. 457-464.
17. Clary D.O., Wolstenholme D.R. The Mitochondrial DNA Molecule of Drosophila yakuba: Nucleotide Sequence, Gene Organization, and Genetic Code//J. Mol. Evol. – 1985. –V. 22. – P. 252-271.
18. Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N. Saccone C. The Complete Nucleotide Sequence, Gene Organization, and Genetic Code of the Mitochondrial Genome of Paracentrotus lividus//The J. Biological Chemistry. – 1989. – V. 264, No. 19. – P. 10965 - 10975.
19. Crozier R.H., Crozier Y.C. The Mitochondrial Genome of the Honeybee Apis mellifera: Complete Sequence and Genome Organization // Genetics. – 1993. – V.133 – P.97 – 117.
20. Козлов Н.Н. Интегральные характеристики генетического кода. Математическое моделирование, т.22.№9.С.51-66.,2010.; англ.пер.: N.N. Kozlov. Integral characteristics of genetic code.Mathematical Models and Computer Simulations April 2011, Volume 3, Issue 2, pp 123-134.
21. Н.Н. Козлов. О некоторых новых характеристиках больших геномов. Математическое моделирование, т. 24, № 10.С.98-108, 2012.; англ. пер.: N.N. Kozlov. Some New Characteristics of Large Genomes Mathematical Models and Computer Simulations 2013, Volume 5, Issue 3, pp 220-228.
22. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J., Molecular Biology of the Cell. //- New York, London: Garland Publishing, Inc., 1994. 1294p.
23. Н.Н. Козлов. Три функции неоднозначностей, порождаемых генетическим кодом. Математическое моделирование, т. 27, № 2, с. 74 - 84, 2015.
24. Н.Н. Козлов. Одна функция неоднозначностей из множеств, порождаемых генетическим кодом. Математическое моделирование, т. 24, № 6, с. 57-66, 2012.
25. Н.Н. Козлов. Генетический код: взгляд математика. Монография. Palmarium Academic Publishing, 2014, 336 с ISBN: 978-3-639-63268-2.
26. Kozlov. N. N. (2017) Computation of the Genetic Code: Full Version. Journal Computer and Communications, 5, 78-94 <https://doi.org/10.4236/jcc.2017.510008>.
27. Н.Н. Козлов, Е.И. Кугушев, Т.М. Энеев. Потенциал генетического кода для перекрытий шести и трех генов. ДАН 2017 Т.473. № 4 С.404-406.
28. Н.Н. Козлов, Е.И. Кугушев, Т.М. Энеев. Математический анализ кодонов, останавливающих синтез белка, ДАН 2017 Т.477. № 3.



# GLOBAL JOURNALS GUIDELINES HANDBOOK 2018

---

[WWW.GLOBALJOURNALS.ORG](http://WWW.GLOBALJOURNALS.ORG)