



Math-Net.Ru

All Russian mathematical portal

N. N. Kozlov, M. G. Bakharev, The mathematical analysis of one biological structure,  
*Keldysh Institute preprints*, 2006, 093

<https://www.mathnet.ru/eng/ipmp641>

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

<https://www.mathnet.ru/eng/agreement>

Download details:

IP: 18.97.9.171

April 26, 2025, 15:16:55





**Ордена Ленина**  
**ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ**  
**имени М.В. Келдыша**  
**Российской академии наук**

**Н.Н. Козлов, М.Г. Бахарев**

Математический анализ одной  
биологической структуры

**Препринт №**

**Москва**

Институт прикладной математики  
имени М.В. Келдыша  
Российской Академии Наук

Н.Н. Козлов, М.Г. Бахарев

Математический анализ одной биологической структуры

Москва 2006

Н.Н. Козлов, М.Г. Бахарев. МАТЕМАТИЧЕСКИЙ АНАЛИЗ ОДНОЙ БИОЛОГИЧЕСКОЙ СТРУКТУРЫ. Препринт ИПМ им. М.В. Келдыша РАН, М., 2006, 28 с

Работа посвящена применению математических и компьютерных методов для исследования одной биологической структуры. Постановка задачи связана с Теоремой для структурных генов или генов кодирующих белки (Козлов Н.Н. ДАН (Математика). 2007. Т. 412. № 5. (в печати)). Из Теоремы следует, что полный набор пар для 20 аминокислот – 400 пар – образует два почти равных по величине множества. Использование кодировок в одном из них может создавать для каждого гена множественные блокировки во всех пяти последовательностях, альтернативных гену. Другое множество не может иметь никакого отношения к созданию блокировок, а используется природой для решения других задач. Одна из них связана с формированием вторичной структуры матричной РНК – мРНК – это копия гена, считанная с ДНК. В работе приводятся результаты исследования указанных задач для одной из наиболее известных таких структур – мРНК MS2. Элементы вторичной структуры исследовались на основе подхода, разработанного ранее в ходе математического анализа перекрывающихся генов. Проведенный анализ показал, что множество пар аминокислот, которые использует природа для построения вторичной структуры указанной мРНК, содержит пары первого множества, что получило свое объяснение на основе указанной Теоремы.

**Ключевые слова:** вторичная структура, перекрывающиеся гены, генетический код, вырожденность кода, блокировка генов, мРНК MS2, 16S-рРНК.

N.N. Kozlov, M.G. Bakharev. The mathematical analysis of one biological structure. Preprint KIAM RAN, Moscow, 2006, 28 p.

Work is devoted to application of mathematical and computer methods for research of one biological structure. Statement of a problem is connected with the Theorem for structural genes or genes coding fibers (Kozlov N.N. DAN (Mathematics). 2007. V. 412. N 5. (in press)). From the Theorem follows that the full set of pairs for 20 amino acids - 400 pairs - forms almost equal two size of set. Use of codings in one of them can create for each gene plural blocking in all five sequences alternative to a gene. Other set cannot have any attitude to creation of blocking, and is used by the nature for the decision of other problems. One of them is connected with formation of secondary structure matrix РНК - mRNA is a copy of a gene, copied from DNA. In work results of research of the specified problems for one of the most known such structures - mRNA MS2. Elements of secondary structure were investigated on the basis of the approach developed earlier during the mathematical analysis of overlapped genes. The lead analysis has shown, that set of pairs amino acids which are used with the nature for construction of secondary structure specified mRNA, contains pairs the first set that has received the explanation on the basis of the specified Theorem.

**Key words:** secondary structure, overlapping genes, genetic code, degenerated code, blocking of genes, mRNA MS2, 16S-rRNA.

Работа выполнена при финансовой поддержке Программы фундаментальных исследований Президиума РАН «Фундаментальные проблемы информатики и информационных технологий», Российского фонда фундаментальных исследований (проекты 04-01-00320, 05-01-00113 и 05-07-90026), а также гранта ведущих научных школ (НШ-2448.2006.1)

### Введение.

Генетическая информация в молекуле ДНК записана в виде определенной последовательности элементов, каждый из которых является одним из четырех нуклеотидов: аденин – А, тимин – Т, цитозин – С, гуанин – G. Эта последовательность нуклеотидов располагается на одной из двух цепей ДНК. Геном является участок одной цепи ДНК. Гены, кодирующие белки, могут соответствовать как одной, называемой плюс- цепью, так и другой минус- цепи двухцепочной ДНК. Каждая из цепей записывается четырехбуквенным алфавитом А,Т,С,С. Устойчивость ДНК поддерживается Уотсон-Криковскими или комплементарными связями, когда А из одной цепи всегда связывается только с Т из другой цепи и аналогично С только с G. Заметим, что связь G-С соответствует трем водородным связям, а А-Т двум (см рис. 1 из [9], см. также [5]).

Белки можно представить как последовательность элементов-аминокислот. В 1966 году было экспериментально установлено, что гены содержат информацию о белках в закодированном виде. Оказалось, что каждая из 20-ти аминокислот кодируется определенными тройками нуклеотидов – кодонами или триплетами. Все эти 64 кодона были экспериментально установлены и представлены в таблице стандартного генетического кода (таблица 1, см. [1]). Одним из свойств генетического кода является его вырожденность. Этим термином называют неоднозначность: одной и той же аминокислоте могут соответствовать несколько кодонов. Дело в том, что лишь две аминокислоты - метионин (Met) и триптофан (Trp) кодируются однозначными кодонами. Все остальные кодируются более чем одним кодоном, но не более чем шестью. Шестью кодонами можно записать только три аминокислоты: серин (Ser), лейцин (Leu), аргинин (Arg). В генетическом коде присутствует 3 кодона - TGA, TAA и TAG, которые не соответствуют ни одной аминокислоте. Их называют терминаторными и обозначают *ter*. Эти кодона играют специальную роль при белковом синтезе, который останавливается, как только одна из таких троек оказывается в фабрике белков (рибосоме).

Ф. Крик сформулировал свою знаменитую центральную догму молекулярной биологии, согласно которой генетическая информация от ДНК к белкам передается через РНК по схеме ДНК – РНК – белок [13]. В основе воспроизводства белка лежат два процесса: транскрипция и трансляция. На первом этапе (транскрипция) происходит копирование информации, записанной на каком-либо участке ДНК. В результате порождается молекула РНК, которую называют матричной РНК - мРНК. Информация в матричной РНК записана в виде последовательности элементов, каждый из которых является одним из четырех нуклеотидов: аденин – А, цитозин – С, гуанин – G, урацил – U (в отличие от Т в ДНК). На втором этапе (трансляция) происходит синтез белковой молекулы по информации, записанной в мРНК.

Помимо первичной структуры или последовательности нуклеотидов у мРНК существует пространственная структура, важнейшей составляющей которой является вторичная структура. Вторичная структура РНК – это представление всех спаренных (т.е. образующие Уотсон-Криковскую связь с каким-либо другим нуклеотидом) и свободных (неспаренных) оснований в молекулярной цепи. Отрезок  $[n_1, n_2]$  свободных оснований в молекулярной цепи называется одонитевым. Отрезки  $[n_1, n_2]$  и  $[n_3, n_4]$  спаренных оснований так, что  $n_1$  спарен с  $n_4$ ,  $n_1 + 1$  с  $n_4 - 1, \dots, n_2$  с  $n_3$ , образуют двухнитевый, или двуспиральный, участок во вторичной структуре РНК. Одонитевые участки (и семейства таких участков) называются петлями, а двухнитевые – стеблями. Стебель можно представить себе как участок винтовой лестницы, где ступеньки – это поперечные, Уотсон-Криковские связи. Длиной стебля называется число пар оснований в нём:  $n_2 - n_1 + 1 = n_4 - n_3 + 1$ . Таким образом, вторичная структура РНК – это совокупность стеблей и петель (см. рис. 1,2). Важным свойством РНК является наличие помимо связей А-У и Г-С неканонической связи Г-У, которая была установлена в экспериментах (на рисунке 2 таких связей 2).

Вторичная структура мРНК важна потому, что при её изменении синтез белка может не идти [13]. Логично предположить, что природа сделала это не случайно. Если принять за 4 среднее число кодонов для аминокислоты, то число генов, соответствующих белку равно  $4^N$ , где N – число аминокислот белка. В нашем случае (при N=544) имеем порядка  $10^{300}$  вариантов. Выбор данного гена объясняется выбором ряда факторов, одним из которых является вторичная структура.

### **Постановка задачи.**

Постановка задачи связана с Теоремой [7], устанавливающая потенциал, который использует природа для решения так называемых задач блокировки. На рисунке 3 приводится фрагмент одного гена (17 смысловых кодонов) и 5 последовательностей кодонов или 5 рамок считывания – РС – сдвинутые относительно исходного гена на -1 (РС1), +1 (РС2) нуклеотид, либо РС3 - РС5, читаемые с минус цепи ДНК и соответствующие сдвигам -1, 0, +1 нуклеотид соответственно. Сама блокировка связана с возникновением в одной или нескольких РС из РС1-РС5, какого-либо из трех кодонов терминации *ter* [14]. Появление такого кодона останавливает синтез белка по альтернативной последовательности. Биологи говорят о мощной биологической защите, которая состоит в том, что если за счет мутаций от РС0 перейдем на одну из пяти РС (РС1-РС5), то белок (как правило эфемерный) синтезируется лишь частично, что не должно повлиять на нормальное функционирование клетки.

Основная задача состояла в том, чтобы установить потенциал генетического кода, который использует природа для указанной блокировки. В формулировке основного результата – Теоремы из [7] –

ограничились рассмотрением только блокировок, образующихся парами аминокислот. Были установлены все 210 пар аминокислот, создающие блокировки в какой-либо РС. Множество из указанных 210 пар аминокислот было разбито на два подмножества. Первое из них соответствует так называемым неизбежным блокировкам и это 31 пара аминокислот. Эти пары выделены на рисунке 2 из [7]. Оказалось, что для этих пар могут блокироваться все РС от РС1 до РС5. Дадим пример такой блокировки (рис 4). Так, для пары аминокислот ValLys имеют место блокировки для РС1, РС2, РС5. Для любого N: A, C, T, G имеет место блокировка TAA (при N=A), TGA в РС5 (при N=C), TGA в РС2 (при N=G) либо дважды TAA одновременно в РС1 и РС5 (при N=T). Эта ситуация существенно отличается, скажем, от пары ValGlu, где блокировки возникают лишь при N=A и N=T, а при других значениях N блокировки не возникают. Подобные пары отнесем к множеству возможных блокировок.

Итак, было установлено [7], что кодировки 190 пар аминокислот из 400 могут решать другую задачу и не имеют никакого отношения к блокировкам. Ниже рассмотрим одну из них, связанную с формированием вторичной структуры мРНК. При этом будет анализироваться одна из наиболее известных вторичных структур мРНК, относящихся к гену репликации одноцепочного РНК вируса MS2 (рис. 1).

Подход к изучению этой структуры был разработан в ходе математического анализа перекрывающихся генов [9]. В ходе исследования проведен математический анализ пяти возможных случаев перекрытия пар генов, многократно обнаруженных в генетических экспериментах. На рисунке 5 такие перекрытия принадлежат всего двум геномам: mt Bovin[10], IS 5[11,12]. Позже было показано, что стебли вторичной структуры мРНК аналогичны (по виду, но не по существу) перекрытию (сл. 3,4,5) фрагментов генов, взятых из разных цепей ДНК (см. рисунок 2). Далее было указано, что малость запретов на генетические перекрытия, которое было установлено для стандартного кода [1] и всех нестандартных кодов[9], позволяет строить феноменальное разнообразие вторичных структур мРНК. Впервые подобный подход к изучению вторичной структуры был представлен Н.Н. Козловым в обзорном докладе на заседании Президиума РАН (18 октября 2005).

При анализе перекрытий пар генов в [2] было введено понятие элементарного перекрытия. Под элементарным перекрытием [6] понимается перекрытие соответствующее одиночным аминокислотам. В этом определении есть неопределенность: сколько нуклеотидов может перекрываться. Поэтому под элементарным перекрытием понимаем перекрывание максимально возможного числа нуклеотидов: это 3 нуклеотида для случая 4 и 2 нуклеотида для всех остальных случаев перекрытий. Постановка задачи состоит в том, что нам следует изучить стебли, в расширенном их представлении (см. ниже), как суперпозицию элементарных перекрытий, принадлежащих разным цепям ДНК.

Таким образом, в данной работе на основе результатов Теоремы из [7] анализируются тексты гена на предмет построения вторичной структуры мРНК и на предмет множественных блокировок.

#### **Анализ полученных результатов.**

В данной работе рассматривается структура матричной РНК MS2 (рисунок 1) – одна из самых больших по количеству нуклеотидов. Она соответствует гену репликации одноцепочного вируса MS2. Впервые изучаемая вторичная структура была опубликована в журнале Nature [3]. Позже она рассматривалась Кантором и Шимелом в [8] и ими было указано, что нуклеотидная последовательность была свернута во вторичную структуру так, чтобы получить максимальный энергетический выигрыш в локальном масштабе. В последние годы рассмотрение в публикациях в мировой литературе данной вторичной структуры активно продолжается, но методы математического моделирования в работах не используются.

**Анализ вторичной структуры.** На основе вторичной структуры (рисунок 1) были выделены все стебли (в том числе длиной единица) и таким образом получен рисунок 6. В нем, в отличие от рисунка 1, для каждого кодона указана также соответствующая аминокислота, а для аминокислот, для которых число кодонов не равно единице, “переменному” нуклеотиду указана одна из букв: Y (отвечающая U или C), X (A или G), Z (A, C), M (A, U, C), N (U, C, A, G), согласно таблице 1.

В результате анализа рисунка 6 было получено, что из 1632 нуклеотидов данной вторичной структуры (рисунок 1) 1088 соответствуют связям (2/3). Остальные позиции (1/3 от общего числа) в связях не участвуют и принадлежат петлям. Возможно, они созданы природой для других функций гена. Целесообразно сравнить указанное число связей с подобным числом у близкой по размеру молекулы РНК – около 1600 нуклеотидов. Имеется ввиду молекула 16S-рРНК E.coli. Вторичная структура такой молекулы согласно [14] содержит около 50 % связей. Полученные различия в 16 % связей для сравниваемых молекул требуют дополнительного исследования. Предварительно можно сказать, что эти различия возможно объясняются тем, что линейный текст мРНК MS2 представляет собой нуклеотидную последовательность, составленную из кодонов, в отличие от текста 16S-рРНК.

Общее число связей данной структуры равно 544. Число пар нуклеотидов, отвечающих разным связям: для связи A-U 203, G-C - 320, G-U - 21. То есть для данной структуры оказалось, что главную роль играют энергетически более устойчивые связи G-C (58 % от общего числа) чем A-U (37 %).

Число связей, относящихся к разным случаям перекрытий: для сл. 3 - 153, сл. 4 – 206, сл 5 - 185.

Все связи в стеблях поделим на зависимые (зависящие от выбора кодировки аминокислоты) и фиксированные (не зависящие от выбора кодировки). Число фиксированных связей равно 149. Из них тройных 0,



парных 46 (23 пары, сл. 3,4) и 103 одиночных. Получим разбиение по случаям перекрытий.

Для парных фиксированных связей в 3-ем случае перекрываемости имеем 7 пар аминокислот, обуславливающих 40 связей (Pro,Gly-4; Ile,Ile-3; Ala,Ala-6; His,Trp-2; Phe,Lys-2; Ile,Met-1; Val,Thr-2). В 4-ом случае получили 2 пары, соответствующие 6 связям (Met,His-2; Pro,Trp-1). Назовём такие связи характерными. Для одиночных фиксированных связей: сл.3 – 20, сл.4 – 49, сл.5 -34.

Зависимых связей всего 395. Из них 210 зависят от выбора кодировки аминокислот Ser, Leu, Arg, каждая из которых кодирует 6 кодонов. Из них 59 связей образуют нерегулярные кодоны этих аминокислот, это AGY(Ser), TTX(Leu), AGX(Arg). Ранее в [15] изучалась влияние этих нерегулярностей на протяженность перекрытий генов в одной цепи ДНК. При этом рассматривались как двойные, так и тройные перекрытия генов.

Анализ вторичной структуры показал, что большую роль в ней играют перекрытия по одному основанию аминокислоты для случаев перекрытий 3 и 5 и по двум или одному основанию для случая перекрытия 4. Назовем такие перекрытия неполными элементарными перекрытиями. Элементарные перекрытия, введенные ранее в [6], назовем полными элементарными перекрытиями. Оказалось, что данная структура состоит на 50 % (272 случая) из неполных элементарных перекрытий и на 50 % из полных элементарных перекрытий. Предварительный анализ показал, что участки соответствующие неполным перекрытиям, по-видимому, не являются функционально значимыми, так как линейный текст мРНК – последовательность кодонов (см. выше).

**Анализ блокировок.** Перейдем к задаче изучения множества, соответствующего блокировкам. Поясним что такое блокировка. В универсальном генетическом коде присутствует 3 кодона (TGA, TAA, TAG), которые не соответствуют ни одной аминокислоте. Их называют терминаторными и обозначают *ter*. Эти кодоны останавливают белковый синтез, который происходит. То есть синтез белка по тексту мРНК начинается с кодона AUG (Met) и продолжается до тех пор, пока не встретится один из терминаторных кодонов (возникает блокировка синтеза белка).

Используя программный комплекс из [4] получили интегральную характеристику цепи РНК (отношение числа терминаторных кодонов по пяти рамкам считывания к численности кодонов в гене). Она равна 0,22 и фактически совпадает со средней интегральной характеристикой для 6-ти из 10-ти геномов [4].

При анализе данных, полученных на выходе программного комплекса, можем найти количество и номера позиций терминаторных кодонов по 5 известным рамкам считывания. В результате общее число терминаторных кодонов равно 121, из них 28 соответствуют сдвигу «-1»

для одной цепи, 25 сдвигу «+1» для одной цепи, 22 сдвигу «-1» для двух цепей, 19 сдвигу «0» для двух цепей, 27 сдвигу «+1» для двух цепей. То есть каждая из PC1-PC5 заблокированы практически одним и тем же количеством кодонов, а цепи ДНК заблокированы 53 и 68 кодонами ter.

Для подробного анализа множества блокировок в тексте мРНК цветом выделены блокировки и стебли (рисунок 7). На основе рисунка получаем, что 46 терминаторных кодона содержится в стеблях (38% от общего числа) и 75 (62 %) в петлях. Напомним, что в нашем понимании блокировка находится в стебле, если все 3 нуклеотида терминаторного кодона соединены связью с другим нуклеотидом. Так как некоторые нуклеотиды из блокировок в петлях (в нашем случае под «блокировкой в петле» понимается блокировка, которая не находится в стебле) могут находиться в стебле, необходимо найти общее число «терминаторных» нуклеотидов (нуклеотиды, принадлежащие терминаторному кодону) в стеблях и петлях. В результате получаем, что в стеблях находится 155 терминаторных нуклеотидов (53 % от общего числа) и 137 в петлях. Таким образом, множество, соответствующее стеблям, и множество, соответствующее блокировкам, имеют пересечение. Из 1088 нуклеотидов, которые участвуют в 544 связях, 155 одновременно входят в блокировки. Из 544 нуклеотидов в петлях одновременно к блокировкам относятся 137 нуклеотида. Таким образом, 407 нуклеотидов в петле не принадлежат стеблям и не участвуют в блокировках. Пока остается неясным, по каким критериям выбирались кодоны-синонимы соответствующие этим нуклеотидам.

Ранее [7] были установлены пары аминокислот, которые обуславливают блокировку по какой-либо из 5 альтернативных рамок считывания независимо от выбора кодирующих аминокислоты кодонов. Полученные блокировки называются неизменными. На рисунке 7 были выделены все пары аминокислот, которые обуславливают неизменную блокировку. Всего таких блокировок 32, из них 15 находятся в стеблях.

Рассмотрим отдельно все терминаторные позиции (Рисунок 8). На рисунке указывается участок текста мРНК, относящийся к терминаторному кодону. Для каждого кодона приписана аминокислота, которую он кодирует и переменные нуклеотиды кодировки данной аминокислоты (если имеется неоднозначная кодировка). Из рисунка видно, что часть блокировочных позиций зависят от выбора кодировки (на рисунке эти позиции выделяются наличием над блокировочным нуклеотидом в строке 2 одной из букв Y,X,M,N или Z). Будем считать, что терминаторный кодон зависит от вырожденности кода, если хотя бы над одним блокировочным нуклеотидом в строке 2 рисунка стоит одна из букв Y,X,M,N или Z. Если же ни над одним блокировочным нуклеотидом нет буквы, то терминаторный кодон не зависит от вырожденности генетического кода.

При подробном анализе рисунка 8 оказалось, что только 3 терминаторных кодона из 121 не зависят от выбора кодировки

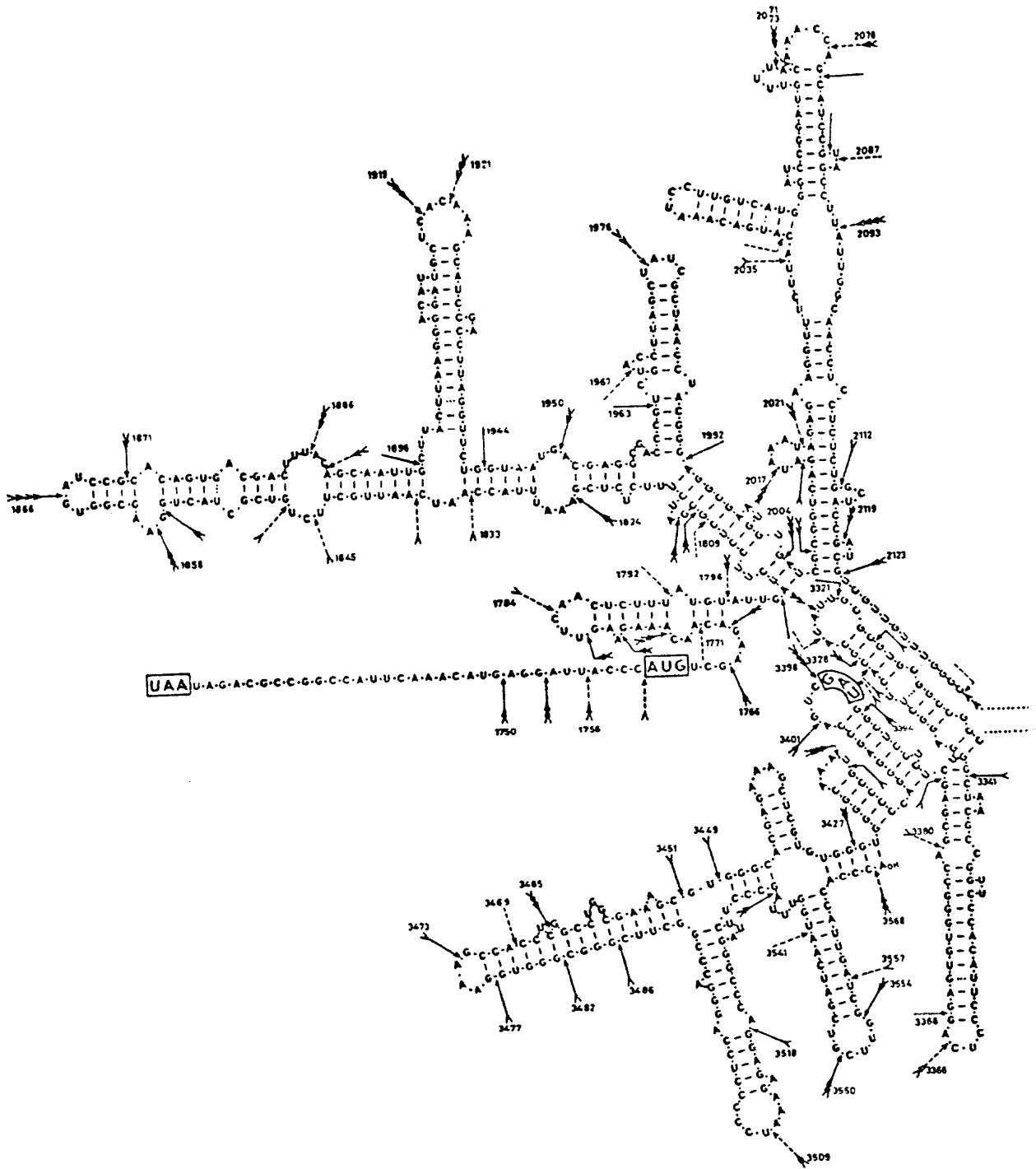
аминокислот, то есть не зависят от вырожденности генетического кода (на рисунке это случаи 45, 65 и 95). Если перейти от терминаторных кодонов к блокировочным нуклеотидам, то получим что из 292 терминаторных нуклеотидов в тексте мРНК, из которых 130 (45 %) зависят от вырожденности генетического кода.

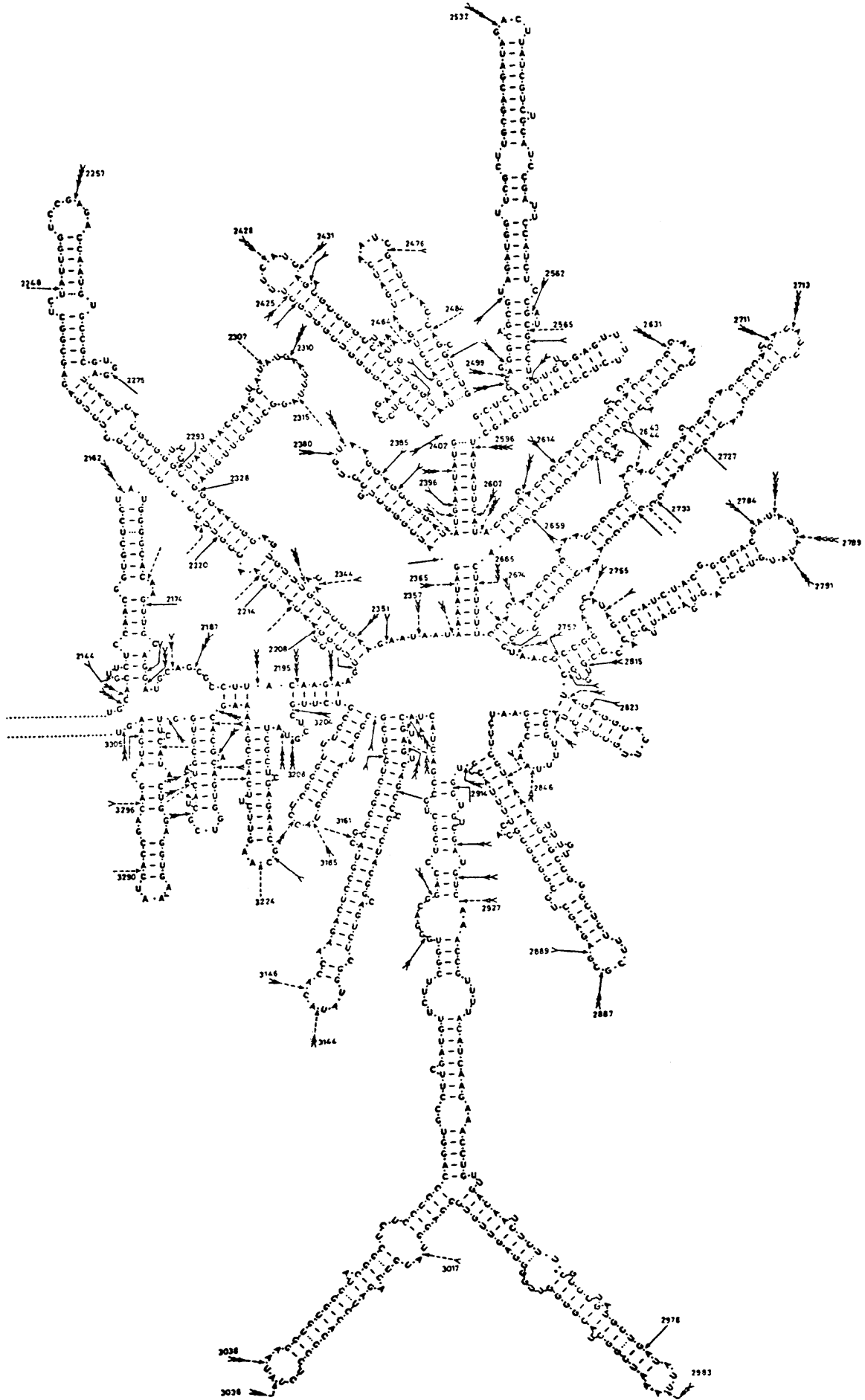
Рассмотрим интересный пример (рисунок 9). Этот пример является случаем неперменной блокировки, потому что при любой кодировке *Pe*, пара *Pe Thr* будут давать блокировку: для кодировки *AUA* это блокировка *UAA*; для кодировок *AUC* и *AUU* это блокировки *UGA* и *UAA* соответственно. Интерес к этому примеру вызван тем, что кодировка *Pe* выбрана таким образом, чтобы была связь *G-C*. При выборе другой кодировки, связь была бы *G-U*, либо ее не было вообще (*G* и *A*). Таким образом, их всех возможных вариантов выбор кодировки в случае пересечения множеств, соответствующих стеблям и блокировкам, в данном случае делается выбор в пользу образования энергетически более выгодные связи.

### Список литературы.

1. Н.Н.Козлов. Теорема для генетического кода. ДАН. 2002. Т. 382. № 5. С. 593-597.
2. Н.Н.Козлов. Применение теоремы для генетического кода. ДАН. 2004. Т. 396. № 6. С. 740-745.
3. W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert & M. Ysebaert. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, **260**, 500-507 (1976).
4. Козлов Н.Н., Грязнов С.С. Некоторые численные характеристики больших геномов. Препринт ИПМ им М.В.Келдыша РАН, 2006, 28 с.
5. Kozlov N.N., Kugushev E.I. Computer simulation of tRNA secondary structure folding. *CABIOS*, 1993, v. 9, p. 253 - 258.
6. Н.Н.Козлов. Элементарные генетические перекрытия. Препринт ИПМ им М.В.Келдыша РАН. №64(2004).
7. Н.Н.Козлов. Математический анализ структурных генов. ДАН (Математика). 2007. Т. 412. № 5. (в печати).
8. Кантор Ч., Шимелл П., Биофизическая химия: В 3-ех т., т.3, М: Мир, 1984
9. Козлов Н.Н. Математический анализ генетических кодов. Математическая биология и биоинформатика. 2005. Т. 1. № 1. С. 70-96, [http://www.matbio.org/downloads\\_en/Kozlov2006\(1\\_70\).pdf](http://www.matbio.org/downloads_en/Kozlov2006(1_70).pdf)
10. Anderson S., de Bruijn M., Coulson A. R., Eperos I. C., Sanger F., Young G. Complete Sequence of Bovine Mitochondrial DNA. *J. Mol. Biol.* **156**. 683-717 (1982).
11. Kröger M., Hobom G. Structural analysis of insertion sequence IS5. *Nature*. 297. 159-162 (1982).
12. Rak B., von Reutern M. Insertion element IS5 contains a third gene. *The EMBO Journal*. 3(4). 807-811 (1984).
13. Ф. Айала, Дж. Кайгер. Современная генетика: В 3-ех т., т. 2, М: Мир, 1988.
14. Льюин Б. Гены. М.: Мир, 1987. 544с
15. Н.Н.Козлов. Перекрывающиеся гены и генетический код. ДАН. 1997. Т. 355. № 6. С. 830-833.

Рисунок 1. Вторичная структура матричной РНК MS2 [3].





А.

	Asp	Pro	Arg	Leu	Tyr	Lys	Val	Trp	Val	
	<u>G·A·U·C·C·A·C·G·C·C·U·C·U·A·U·A·A·G·G·U·G·U·G·G·G·U·A</u>									
3022										3048

Б.

	Asp	Pro	Arg	Leu							
3022	G·A·U·C·C·A·C·G·C·C·U·C·U·									.	
		:					:			.	A
3048	A·U·G·G·G·U·G·U·G·G·A·A·U									.	
	Val Trp Val Lys										

**Рисунок 2.** Линейный текст (А)– фрагмент нуклеотидной и аминокислотных последовательностей, который соответствует стеблю (Б) вторичной структуры матричной РНК MS2 из [3]. Стебель представлен в нижней части этой вторичной структуры (левый стебель). Направление чтения на стебле Б (стрелки) становится различным и этот фрагмент вторичной структуры аналогичен (по виду, а не по сути) перекрытию участков генов AUCCACGCC и GGUGUGGGU (подчеркнуты в А), взятых из разных цепей ДНК (случай перекрытия 4). Шпилечная петля UCUAUAА сделана в формате рисунка вторичной структуры матричной РНК MS2 из [3]. В линейном тексте в области петли выделены терминаторные кодоны, один из которых соответствует плюс–цепи, а другой минус–цепи.

OPC0 MetSerIleLysLeuSerTyrArgGluSerPheSerIleLeuGluGluVal...

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

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

→ TATGAGCATTAAACTTAGTTATAGAGAGTCATTTAGTATATTAGAGGAGGTTTA...

← ATACTCGTAATTTGAATCAATATCTCTCAGTAAATCATATAATCTCCTCCAAAT...

BPC3 (-1) IleLeuMetLeuSerLeu \* LeuSerAspAsnLeuIleAsnSerSerThr ...

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

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

**Рисунок 3.** Шесть РС для фрагмента гена (начало с кодона ATG(Met), направление чтения указано стрелкой →) из которых одна РС является открытой – OPC0 (в ней 17 смысловых кодонов или кодонов кодирующих только аминокислоты), 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 OPC0. Символом \* обозначены каждый из трех кодонов ter. Такие кодоны, как показывает анализ природных генов, многократно присутствуют в каждой из 5-и BPC1-BPC5.

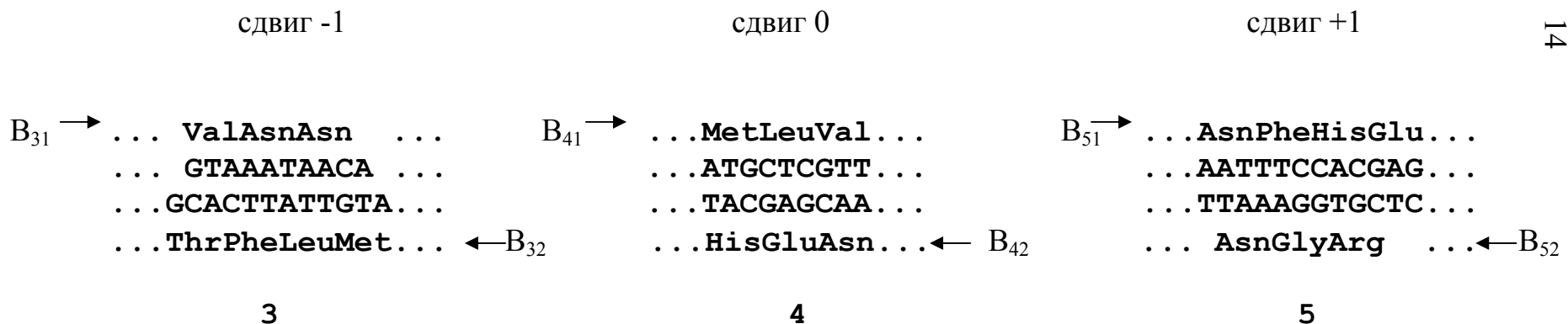
V a l L y s G T N A A X		V a l G l u G T N G A X
N: A - G T A A A X →		N: A - G T A G A X →
N: C - G T C A A X A G T ←		N: C - G T C G A X C A G C T Y
N: G - G T G A A X →		N: G - G T G G A X C A C C T Y
N: T - G T T A A X → C A A T T Y ←		N: T - G T T G A X → C A A C T Y

**Рисунок 4.** Неизбежные блокировки для пары аминокислот ValLys (слева) и возможные блокировки для пары аминокислот ValGlu (справа). Кодоны ter из (1) заштрихованы. Использованы кодировки для стандартного кода Val: GTN, Lys: AAX, Glu: GAX, где N: A, C, T, G; X: A, G; Y: T, C. стрелки указывают направление чтения.

Перекрытия генов из одной цепи ДНК.



Перекрытия генов из разных цепей ДНК.



**Рисунок 5.** Пять возможных случаев перекрываемости генов, соответствующих одной (1,2) либо двум (3,5) цепям ДНК. Чтение текстов при этом осуществляется в разных направлениях (указано стрелкой) : слева направо для  $B_{11}$ ,  $B_{12}$ ,  $B_{21}$ ,  $B_{22}$ ,  $B_{31}$ ,  $B_{41}$ ,  $B_{51}$  и справа налево для  $B_{32}$ ,  $B_{42}$ ,  $B_{52}$ . Здесь все случаи перекрытий соответствуют только двум геномам. Случаи 1, 2, 5 присутствуют в кодировке мтДНК *Bovini*[10], а случаи 3, 4 – IS 5[11,12]. Подробности см. на рисунке 6 из [9].



## 15

1. Thr N <u>ACA</u> UGU Y Cys	2. ThrLys N X <u>ACAAAG</u> AUUUCU X YN LeuSer	3. Lys X AAG UCUCA N Y SerAsn	4. AspLeu YY N <u>GAUCUU</u> UAG Y Asp	5. Pro N <u>CCU</u> UGG N Gly	6. Pro N <u>CCU</u> UAA Y Asn	7. Arg Z N <u>CGC</u> GCG N Ala
8. Leu Y N <u>CUU</u> GAG X Glu	9. Ser N <u>UCU</u> CCG N Pro	10. LeuGlu Y N X <u>CUCGAA</u> GAGCAG X Y GluAsp	11. IleTyrGln M Y X <u>AUUUACCAA</u> UAAUGGUCU Y N N AsnGlySer	12. SerIleAla N M N <u>UCAAUUGCU</u> UCGUU AACGACA N M N N AlaIleAlaThr		
13. ValAlaThrGly N N N N <u>GUCGCUACUGGA</u> CAGCAGUGACAC Y Y Y Y AspAspSerHis	14. SerGly Y N <u>AGCGGU</u> CACGCC Y N HisPro	15. TyrLeuArg YY XZ X <u>UACUUAAGG</u> UGGAU UCCA N X YN GlyLeuThr	16. Asp Y <u>GAC</u> GCC N Pro	17. Glu X <u>GAA</u> GCCUAC N Y ProHis		
18. LeuLeu Y XY N <u>UUGCUC</u> UACGAA Y X HisLys	19. ThrArgArg NZ NZ N <u>ACCCGUCGU</u> GAGGGCAUC X N ZN Y GluArgLeu	20. ThrLeuAla NY X N <u>ACCUUAGCU</u> GAAUCG X N LysAla	21. Arg Z N <u>CGC</u> GCU N Ser			
22. ArgGly Z N N <u>CGCGGU</u> AGCCA N ZY ArgTyr	23. Gln X <u>CAG</u> CGGUCU N N GlySer	24. ArgGluGlyLeu Z X X NY X <u>AGAGAAGGUUUG</u> UCUCUCCUCCAA N N YN YY SerLeuLeuAsn	25. HisAspLys Y Y X <u>CAUGACAAA</u> GGUACUGUU N X Y TrpSerLeu			
26. TrpAsp Y UGGGAU UUCCGA N YY LeuSer	27. Pro N <u>CCG</u> CGA Y Ser	28. ProAspVal N Y N <u>CCGGAUGUU</u> UGCCUAGGA N ZM X Z ArgIleArg	29. Gln X <u>CAA</u> CGA Y Ser	30. Thr N <u>ACG</u> UAG Y Asp		
31. PheSerAsn Y N Y <u>UUCUCCAAC</u> GACGUUGAA X X YX GlnLeuLys	32. GlyAlaPro N N N <u>GGUGCUCU</u> CACGGGGUA Y N HisGlyMet	33. ProTyr N Y <u>CCUUAC</u> GAA X Lys	34. TyrLysLys Y X X <u>UACAAGAAG</u> UCGUUCUGC N N YN Z AlaLeuArg			

## 16

35. PheAla Y N <u>UUCGCU</u> GAAGCC X N LysPro	36. GluGln X X <u>GAACAA</u> GCCUUG N N ProVal	37. GlnAla X N <u>CAAGCA</u> UUUGUG Y N PheVal	38. ThrVal N N <u>ACCGUU</u> AGGCAA N Y GlyAsn	39. Thr N <u>ACC</u> GGG N Gly
40. Pro N <u>CCC</u> CGC N Z Arg	41. ProArgAla NZ N N <u>CCCCGCGCU</u> CUGGCGCAC N N Y ValAlaHis	42. ValLeuArg NY NZ X <u>GCUCUGAGA</u> AGACUA X ZM ArgAla	43. AlaAla N N <u>GCGGCU</u> GCGCCG N N AlaAla	
44. LeuLeuVal Y NY X N <u>CUAUUGGUC</u> UGUAACCAG Y X Y CysGlnAsp	45. ArgTyrAsnGlu Z N Y Y X <u>CGCUAUAACGAG</u> GGGAUGUUGCUC N N N N Y GlyValValLeu	46. Ser N <u>UCA</u> GGA X Z Arg	47. LysIleAsp X M Y <u>AAAAUAGAU</u> CCUUUAUC N Y N Y SerPheLeu	
48. AspArgAla YZ N N <u>GAUCGGGCU</u> UAGUCCGAG Y N X AspProGlu	49. Ala N <u>GCC</u> GAGGAA X X GluLys	50. MetAsnMet Y <u>AUGAAUAUG</u> UAUACUUAUAUC Y N Y N Y TyrSerTyrLeu	51. TyrLeu YY N <u>UACCUC</u> AUAUGG M N IleGly	
52. Lys X <u>AAA</u> UUG N Val	53. Gly N <u>GGG</u> CCU N Ser	54. Gly N <u>GGG</u> CUC N Y Leu	55. ValGlyAla N N N <u>GUCGGUGCU</u> CUCGGCCGC N YN ZN Z LeuArgArg	56. Ala N <u>GCU</u> AGA X Z Arg
57. IleAsp M Y <u>AUAGAC</u> GUCUGC N YN Z LeuArg	58. Leu Y N <u>CUG</u> GAC X Gln	59. AsnAspGln Y Y X AAUGAUCAA CAACUAGCU Y M N AsnIleSer	60. AlaGln N X <u>GCUCAG</u> AUCGAGUCC N YX N LeuGluPro	61. Gln X <u>CAG</u> GUC N Y Leu
62. Gly N <u>GGC</u> GUCCGC N YN Z LeuArg	63. Ser Y <u>AGC</u> CGC N Z Arg	64. SerValHisGly Y N Y N <u>AGCGUAGAUGGU</u> UAGCCUCUACCU Y N M N AspSerIleSer	65. SerLeuAla NY N N <u>UCGCUUGCG</u> UAGCCUACG Y N N AspSerAla	

66. AlaThrIle N N M <u>GCGACGAUA</u> UCUGCUAAU N N X Y SerSerLeu	67. LeuVal Y N N <u>CUGGUG</u> UCCACC N N ProPro	68. TrpSer Y <u>UGGAGU</u> ACCCUC N N Y ProLeu	69. TyrLeu YY N <u>UAUCUC</u> AAGGGU X GluTrp
70. AspArgIle YY N M <u>GAUCGUAUC</u> GGUAGCAUAGCA N ZM N TrpArgIleThr	71. Arg Z N <u>CGC</u> GAGCGG X N GluGly	72. Ser N <u>UCA</u> CGGUAG N Y GlyAsp	73. HisTyrGly Y Y N <u>CACUACGGA</u> UAGAUGC Y N M AspValIle
74. SerThrMetGly N N N <u>UCCACAAUGGGA</u> UGGUUUUACCUA N Y Y M GlyPheHisIle	75. AsnGlyPhe Y N Y <u>AAUGGGUUC</u> AACCCAGCG X N N GlnThrAla	76. PheGlu Y X <u>UUUGAG</u> AAACUG X N LysVal	77. Leu Y N <u>CUA</u> CUGAUA N M ValIle
78. GluSerMet X N <u>GAGUCCAUG</u> ACGGGU N AlaTrp	79. Met <u>AUG</u> CUU Y Phe	80. Ala N <u>GCC</u> AUCGUG N YN LeuVal	81. AlaGly N N <u>GCCGGA</u> UGCCCC N ZN ArgPro
82. Gly N <u>GGC</u> CCC N Pro	83. Gly N <u>GGC</u> ACG N Ala	84. IleTyr M Y <u>AUCUAC</u> UUAGAG M X IleGlu	85. TyrGlyAsp Y N Y <u>UACGGGGAC</u> UGACCCUGU Y N Y SerProCys
86. LeuGlu Y N X <u>CUAGAG</u> CAUCAU Y Y TyrTyr	87. GluAla X N <u>GAGGCA</u> CAUCCG Y N TyrAla	88. TyrGly Y N <u>UACGGU</u> GCC N Pro	89. Arg Z N <u>CGU</u> UGC N Z Arg
90. ArgLysThr Z N X N <u>CGUAAAACG</u> CAUUUU Y Y TyrPhe	91. Thr N <u>ACG</u> GCC N Pro	92. PheVal Y N <u>UUCGUG</u> GCGCGG N N AlaGly	93. SerGly N N <u>UCCGGG</u> CGGCGU N Y GlyCys
94. GlyLeuPhe NY N Y <u>GGGCUCUUU</u> CGAGAG Y X SerGlu	95. Gly N <u>GGU</u> CCG N Ala	96. Val N <u>GUC</u> UCG N Ala	97. ValAspVal N Y N <u>GUCGAUGUC</u> UCGCUCCAG N N YY AlaLeuAsp

## 18

98. LysPro X N <u>AAACCG</u> UGGCUU N Y GlyPhe	99. TyrIle Y M <u>UACAUC</u> CUUGUAGCU Y N PheMetSer	100. Lys X <u>AAG</u> GCUUCC N N SerPro	101. LysProVal X N M <u>AAACCUUU</u> GUGGAC N X ValGln
102. AspAsn Y Y <u>GACAAU</u> CUGUUG N N ValVal	103. LeuPhe Y N Y <u>CUCUUC</u> UUGAGG N N ValGly	104. Phe Y <u>UUC</u> UGG N Gly	105. AlaLeu NY N <u>GCCUCG</u> UGGGG N N Z GlyArg
107. Met <u>AUG</u> AUC N Y Leu	108. MetLeuIle Y N M <u>AUGCUGAUA</u> AUCGGCUAA N YN ZY LeuArgAsn	109. GlyGly N N <u>GGAGGU</u> CCUCCU N N SerSer	110. MetSer N <u>AUGUCA</u> CUCGGC N YN Z LeuArg
112. AspProArgLeu Y NZ N N <u>GAUCCACGCCUC</u> AUGGGUGUGGAA N N X ValTrpValLys	113. Tyr Y <u>UAC</u> GCG N Ala	114. Val N <u>GUA</u> CUC N Y Leu	115. ValVal N N <u>GUAGUC</u> CUC N Y Leu
116. ValSerPro N Y N <u>GUCAGCCCG</u> GUCGGC N YN Z LeuArg	117. Pro N <u>CCG</u> GGC N Z Arg	118. Pro N <u>CCU</u> GGG N Gly	119. ProThrAla N N N <u>CCUACGGCA</u> CAUGCC Y N TyrPro
120. AlaValSerVal N N N N <u>GCAGUCUCGGUA</u> UCAGAACCA N X N ThrLysThr	121. AlaAspThrArg N Y NZ N <u>GCGGAUACCCGU</u> UGCCUUUGGGCU N ZY N N ArgPheGlySer	122. IleAla M N <u>AUCGCU</u> AAGCGA X Y GluSer	123. Arg Z N <u>CGA</u> CUU Y Phe
124. GluArg XZ N <u>GAACGC</u> CUUGAA Y X PheLys	125. HisAsp Y Y <u>CACGAC</u> GGUGCG N TrpAla	126. Asp Y <u>GAC</u> GCG N Ala	127. Ser Y <u>AGU</u> AUA M Ile
			128. SerGly Y N <u>AGUGGU</u> CAU Y Tyr

129. Gly	130. Phe	131. HisThrGlyGlyGlu	132. Ser
N	Y	Y N N N X	N
<u>GGU</u>	<u>UUC</u>	<u>CAUACUGGAGGUGAA</u>	<u>UCC</u>
CGC	GAA	GUACGACAGCCACUA	GGU
N Z	X	Y Y N M	
Arg	Lys	MetSerAspThrIle	Trp
133. SerAlaGlyValArg	134. TrpLeu	135. ThrProVal	
N N N NZ N	Y N	N N N	
<u>UCCGCCGGCGUGCGC</u>	<u>UGGCUA</u>	<u>ACGCCGGUU</u>	
<u>GAGGCUUCACGC</u>	CGA	GCGACC	
X N N N Z	Y	N N	
GluSerThrArg	Ser	AlaPro	
136. ProThrPhePro			
N N Y N			
<u>CCCACAUUCCCU</u>			
<u>GGGUGUGAGGAC</u>			
N Y X X			
GlyCysGluGln			

**Рисунок 6.** Все стебли вторичной структуры, включая стебли, длиной единица. В каждом из случаев чтение текста в верхней последовательности нуклеотидов осуществляется слева направо, для нижней справа налево. На рисунке каждой тройке нуклеотидов (кодону), используя универсальный генетический код (таблица 1), указана соответствующая аминокислота. Если число кодонов для аминокислоты не равно единице, то “переменному” нуклеотиду указана одна из букв: Y (отвечающая U или C), X (A или G), Z (A, C), M (A, U, C), N (U, C, A, G).

## 20

- |  |  |   |   |  |  |
|--|--|---|---|--|--|
| 1. PheAsn<br>Y Y<br>ttcaac<br><b>tca</b><br><b>agt</b>   | 2. Leu<br>Y X<br>tta<br><b>tta</b><br><b>aat</b>           | 3. IleAsp<br>M Y<br>attgat<br><b>tga</b>                  | 4. IleTyr<br>M Y<br>atttac<br><b>tta</b><br><b>aat</b>                    | 5. Ser<br>N<br>tca<br><b>tca</b><br><b>agt</b>                     | 6. AlaThr<br>N N<br>gctact<br><b>cta</b><br><b>gat</b> |
| 7. GlyAsp<br>N Y<br>ggtgat<br><b>tga</b>                 | 8. SerAsp<br>Y Y<br>agtgac<br><b>tga</b>                   | 9. PheThr<br>Y N<br>tttaca<br><b>tta</b><br><b>aat</b>    | 10. AlaTyr<br>N Y<br>gcttac<br><b>tta</b><br><b>aat</b>                   | 11. LeuArg<br>Y XZ X<br>ttaagg<br><b>ttaa</b><br><b>aat</b>        |  |
| 12. LeuThr<br>N N<br>ctcaca<br><b>tca</b><br><b>agt</b>  | 13. LeuGly<br>Y X N<br>ttaggt<br><b>ttag</b><br><b>aat</b> | 14. GlyAsnAsp<br>N Y Y<br>ggtaatgac<br><b>taatga</b>      | 15. LeuAlaIle<br>Y X N M<br>ttagctatc<br><b>ttagcta</b><br><b>aat gat</b> |  |  |
| 16. AlaLys<br>N X<br>gctaag<br><b>ctaa</b><br><b>gat</b> | 17. Leu<br>Y N<br>cta<br><b>cta</b><br><b>gat</b>          | 18. GlyAsp<br>N Y<br>ggtgat<br><b>tga</b>                 | 19. GlyGln<br>N X<br>ggtcag<br><b>tca</b><br><b>agt</b>                   | 20. IleAsnArg<br>M YZ X<br>ataaataga<br><b>taa tag</b>             |  |
| 21. Leu<br>Y X<br>tta<br><b>tta</b><br><b>aat</b>        | 22. HisAsp<br>Y Y<br>catgac<br><b>tga</b>                  | 23. Ser<br>N<br>tca<br><b>tca</b><br><b>agt</b>           | 24. Leu<br>Y X<br>tta<br><b>tta</b><br><b>aat</b>                         | 25. ArgSer<br>N Y<br>cgtagc<br><b>tag</b>                          |  |
| 26. LeuIle<br>N M<br>cttatt<br><b>tta</b><br><b>aat</b>  | 27. GlyTyr<br>N Y<br>ggctac<br><b>cta</b><br><b>gat</b>    | 28. ProMet<br>N<br>cctatg<br><b>cta</b><br><b>gat</b>     | 29. ProTyr<br>N Y<br>ccttac<br><b>tta</b><br><b>aat</b>                   | 30. AlaGlu<br>N X<br>gctgaa<br><b>tga</b>                          |  |
| 31. ValThr<br>N N<br>gttacc<br><b>tta</b><br><b>aat</b>  | 32. LeuArg<br>Y NZ X<br>ctgaga<br><b>tga</b>               | 33. Leu<br>Y N<br>cta<br><b>cta</b><br><b>gat</b>         | 34. IleArg<br>MZ X<br>atcaga<br><b>tca</b><br><b>agt</b>                  | 35. ArgTyrAsn<br>N Y Y<br>cgctataac<br><b>ctataa</b><br><b>gat</b> |  |
| 36. Ser<br>N<br>tca<br><b>tca</b><br><b>agt</b>          | 37. TyrGlu<br>Y X<br>tatgaa<br><b>tga</b>                  | 38. PheArg<br>YZ X<br>tttagg<br><b>ttag</b><br><b>aat</b> | 39. ValGly<br>N N<br>gtaggg<br><b>tag</b>                                 | 40. PheThr<br>Y N<br>tttaca<br><b>tta</b><br><b>aat</b>            |  |
| 41. AsnAsnLys<br>Y Y X<br>aataataaa<br><b>taataa</b>     | 42. IleAsp<br>M Y<br>atagat<br><b>tag</b>                  | 43. CysLys<br>Y X<br>tgtaag<br><b>taa</b>                 | 44. ProAsp<br>N Y<br>cctgat<br><b>tga</b>                                 |  |  |

45. MetAsn Y atgaat <b>tga</b>	46. PheIleArg Y MZ X ttcatcaga <b>tcatca</b> <b>agtagt</b>	47. LeuLys N X ctcaaa <b>tca</b> <b>agt</b>	48. IleAsp M Y atagac <b>tag</b>		
49. LeuAsnAspGln Y N Y Y X ctgaatgatcaa <b>tga tgatca</b> <b>agt</b>	50. IleAsn M Y atcaac <b>tca</b> <b>agt</b>	51. AlaGln N X gctcag <b>tca</b> <b>agt</b>	52. ValAsp N Y gtagat <b>tag</b>		
53. IleAsp M Y atagac <b>tag</b>	54. Leu Y X tta <b>tta</b> <b>aat</b>	55. ProGlu N X cctgag <b>tga</b>	56. Leu Y N cta <b>cta</b> <b>gat</b>	57. Ser N tca <b>tca</b> <b>agt</b>	58. Ser N tca <b>tca</b> <b>agt</b>
59. HisTyr Y Y cactac <b>cta</b> <b>gat</b>	60. ValAsp N Y gtagat <b>tag</b>	61. Leu Y N cta <b>cta</b> <b>gat</b>	62. PheThr Y N ttcaca <b>tca</b> <b>agt</b>	63. PheGlu Y X tttgag <b>tga</b>	
64. LeuGlu Y N X ctagag <b>ctag</b> <b>gat</b>	65. MetIle M atgata <b>tga</b>	66. IleValLys M N X atagtcaaa <b>tagtca</b> <b>agt</b>	67. GlyAsn N Y ggtaac <b>taa</b>		
68. IleGly M N ataggc <b>tag</b>	69. IleTyr M Y atctac <b>cta</b> <b>gat</b>	70. IleIle M M attata <b>tta</b> <b>aat</b>	71. SerGlu Y X agtgag <b>tga</b>	72. ValLeuGlu NY N X gtgctagag <b>ctag</b> <b>gat</b>	
73. AlaTyrTyr N Y Y gcctactac <b>ctacta</b> <b>gatgat</b>	74. PheLys Y X tttaaa <b>ttaa</b> <b>aat</b>	75. ArgLys N X cgtaaa <b>taa</b>	76. PheTyr Y Y ttttac <b>tta</b> <b>aat</b>		
77. ValLys N X gtcaaa <b>tca</b> <b>agt</b>	78. PheTyr Y Y ttttac <b>tta</b> <b>aat</b>	79. IleLys M X atcaag <b>tca</b> <b>agt</b>	80. ValAsp N Y gttgac <b>tga</b>	81. LeuMet Y N ctgatg <b>tga</b>	

82. LeuIle Y N M ctgata <b>tga</b>	83. LeuAsn Y X Y ttaa <b>ttaa</b> aat	84. Leu Y N cta <b>cta</b> gat	85. Ser N tca <b>tca</b> agt	86. LeuTyrLys N Y X ctctataag <b>ctataa</b> gat
87. AspTyrTyr Y Y Y gactactac <b>ctacta</b> gatgat	88. ValValSer N N Y gtagtcagc <b>tagtca</b> agt	89. ProThr N N cctacg <b>cta</b> gat	90. PheSer Y Y ttcagc <b>tca</b> agt	
91. ArgTyr N Y cgctac <b>cta</b> gat	92. IleAla M N atagcg <b>tag</b>	93. GlyGlu N X ggtgaa <b>tga</b>	94. IleThr M N atcacc <b>tca</b> agt	95. MetLys X atgaag <b>tga</b>
96. ValIle N M ggtata <b>tta</b> aat	97. LeuThr Y N N ctaacg <b>ctaa</b> gat	98. ProGln N X cctcag <b>tca</b> agt		

**Рисунок 8.** Случаи терминаторных позиций во вторичной структуре. В каждом из случаев указана последовательность нуклеотидов. Над последовательностью нуклеотидов каждой тройке нуклеотидов указана соответствующая аминокислота, если число кодонов для аминокислоты не равно единице, то “переменному” нуклеотиду указана одна из букв: Y (отвечающая U или C), X (A или G), Z (A, C), M (A, U, C), N (U, C, A, G). Под текстом указаны позиции, отвечающие блокировкам; для блокировок, соответствующих двум цепям, под позицией указывается терминаторный кодон.



А.

His	Thr	Gly	Gly	Glu	Ile	Thr	Asp	Ser	Met
C·A·U·A·C·U·G·G·A·G·G· <b>U·G·A·A</b> ·A·A·U·C·A·C·C·G·A·C·A·G·C·A·U·G									
3274									3303

Б.

			→						
	His	Thr	Gly	Gly	Glu				
3274	C·A·U·A·C·U·G·G·A·G·G· <b>U·G·A·A</b>								
3303	G·U·A·C·G·A·C·A·G·C·C·A·C·U·A								
	Met	Ser	Asp	Thr	Ile				
				←					

**Рисунок 9.** Случай неперменной блокировки в стебле вторичной структуры, когда при выборе кодировки аминокислоты с сохранением связи всегда возникает блокировка. Блокировка на рисунке выделена жирным, пара аминокислот, обуславливающая блокировку, выделена серым фоном. Линейный текст (А)– фрагмент нуклеотидной и аминокислотных последовательностей из Рисунка 1, который соответствует стеблю (Б) вторичной структуры матричной РНК MS2 [3]. Направление чтения на стебле Б (стрелки) становится различным и этот фрагмент вторичной структуры эквивалентен перекрытию участков генов CAU с AUG, CUG с CAG и GGUG с CACC.

**Таблица 1.** Стандартный генетический код  $K_0$ . Генетический код можно рассматривать как оператор преобразования между последовательностью аминокислот (соответствующих белку) и геном. Единицей генетического кода является тройка нуклеотидов - кодон. Для каждой из 20-и аминокислот приводятся: 1- стандартные трехбуквенные сокращения, 2 – число кодонов, 3 – трехбуквенные представления кодонов. Обозначение: X: A, G; Y: T,C; N: A,G,T,C. В последней строке приводятся три терминаторных кодона – ter, каждый из которых обозначает останова синтеза белка.

	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	AAH
11	Glu	2	GAX
12	Ile	3	ATY, ATA
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