



ИПМ им.М.В.Келдыша РАН • [Электронная библиотека](#)

[Препринты ИПМ](#) • [Препринт № 10 за 2021 г.](#)



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

[Н.Н. Козлов](#), Е.И. Кугушев

**Математический феномен
генетической связи G-U**

Рекомендуемая форма библиографической ссылки: Козлов Н.Н., Кугушев Е.И. Математический феномен генетической связи G-U // Препринты ИПМ им. М.В.Келдыша. 2021. № 10. 15 с. <https://doi.org/10.20948/prepr-2021-10>
<https://library.keldysh.ru/preprint.asp?id=2021-10>

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

Н.Н. Козлов , Е.И. Кугушев

**Математический феномен
генетической связи G-U**

Москва — 2021

Н.Н. Козлов , Е.И. Кугушев

Математический феномен генетической связи G-U

Как известно, вторичная структура молекул РНК в отличие от ДНК содержит помимо канонических или уотсон-криковских пар, неканонические пары G-U , это гуанин – урацил. Последние пары были установлены экспериментально посредством рентгеноструктурного анализа и до сих пор являются загадкой для исследователей. Дело в том, что такие пары в 5-10 раз являются энергетически более слабыми, чем канонические пары A-U, C-G. Возникает вопрос: зачем же нужны такие пары?

Исследование проводилось в рамках установленной ранее авторами строгой математической аналогии некоторых случаев генетических перекрытий и стеблей вторичной структуры матричных РНК. В данной работе показывается, что за счет неканонической пары G-U, с использованием обнаруженных новых неоднозначностей, можно существенно «регулировать» величину свободной энергии стебля (на порядок и более даже для небольших стеблей) – важнейшую и функционально значимую биохимическую характеристику. Оказалось, что обнаруженный эффект для пар G-U значительно превышает подобный эффект для канонических пар A-U, C-G.

Ключевые слова: генетический код, неканонические пары G-U, свободная энергии стебля, новые неоднозначности

Nicolay Nicolayevich Kozlov, , Evgeny .Ivanovich .Kugushev,

Mathematical Phenomenon of Genetic Link G-U

As you know, the secondary structure of RNA molecules, unlike DNA, in addition to canonical or Watson-Crick pairs, non-canonical G-U pairs, this is guanine - uracil. The last pairs were established experimentally by X-ray diffraction analysis and are still a mystery to researchers. The fact is that such pairs are 5-10 times energetically weaker than the canonical pairs A-U, C-G. The question arises: why are such pairs needed?

The study was carried out in the framework of the strict mathematical analogy established earlier by the authors of some cases of genetic overlaps and stems of the secondary structure of messenger RNAs. In this work, it is shown that due to the non-canonical G-U pair, using the new ambiguities found, it is possible to substantially “regulate” the stem free energy (even for small stems) - the most important biochemical characteristic. It turned out that the discovered effect for G-U pairs significantly exceeds a similar effect for canonical pairs A-U, C-G.

Key words: genetic code, non-canonical pairs G-U, free energy of the stem, new ambiguities

Данная работа относится к новому этапу развития математической теории генетического кода [1,2], которая формировалась, начиная с 1993 года. Основным результатом данной работы был получен при анализе новых множеств, порождаемых генетическим кодом и построенных с использованием результатов данной теории. Как известно, структура молекул РНК в отличие от ДНК содержит помимо канонических или уотсон-криковских пар, неканонические пары G-U, это гуанин – урацил. Последние пары были установлены экспериментально посредством рентгеноструктурного анализа [3] и до сих пор являются загадкой для исследователей. Прежде всего, дело в том, что такие пары в 5 - 10 раз являются энергетически более слабыми, чем канонические пары A-U, C-G. Возникает вопрос: зачем же нужны такие пары? Авторы специально не ставили подобную задачу. Однако, в ходе математического анализа генетического кода, с учетом работы [4], нами был обнаружен совершенно неожиданный математический результат, потребовавший своего объяснения, которое дается ниже.

Основу нашего подхода составляет строгая математическая аналогия между стеблями вторичной структуры матричных РНК (копий с ДНК структурных генов) и перекрытиями генов, принадлежащих различным цепям ДНК [1]. Математический анализ перекрывающихся генов основывался на множествах элементарных перекрытий или перекрытий, соответствующих одиночным аминокислотам [5,6]. Для изучения канонических связей было целесообразно ввести редуцированные множества элементарных перекрытий, анализ которых позволил сократить исходные множества элементов практически вдвое: с 283 до 151 [4]. Такое сокращение стало возможным благодаря внутренней симметрии исходных множеств [6]. Уточним главные обозначенные позиции. На рис.1, приводится один из стеблей вторичной структуры матричной РНК MS2, представление которой занимает полные 2 страницы [7]. Это одна из наиболее длинных и наиболее изучаемых в мире структур. Ее фрагмент приводится в конце данного препринта. Математический анализ полной такой структуры проводился нами и некоторые ранее полученные результаты уже опубликованы

[8]. На рис.1А дана первичная, а на рис.1Б – вторичная структура одного из стеблей этой структуры, соответствующего интервалу 3022-3048 из работы [7].

```

    Asp Pro Arg Leu Tyr Lys Val Trp Val
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
|
3022
|
3048

```

А.

```

    ────────────>
                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
    <───────────

```

Б.

Рис.1. Линейный текст (А) и стебель (Б), соответствующий интервалу 3022-3048 вторичной структуры матричной РНК MS2 из [7]. Данный стебель соответствует нижней части полного рисунка этой вторичной структуры (левый стебель). Направление чтения на стебле (фрагмент Б) указывается стрелками; оно становится различным и этот фрагмент вторичной структуры аналогичен (по виду, а не по сути) перекрытию участков генов AUCCACGCC и GGUGUGGGU (подчеркнуты во фрагменте А), взятых из разных цепей ДНК. Шпилечная петля UCUAUAA сделана в формате рисунка из [7].

Нами используются обозначения, принятые в этой работе: вертикальные сплошные черточки соответствуют каноническим парам, а пунктирные – неканоническим G-U. На рис.1Б указаны 9 пар (в полной вторичной структуре РНК MS2 таких пар около 550), относящихся к подчеркнутым фрагментам из рис.

1А. В них лишь 2 являются неканоническими – G-U. Важно подчеркнуть, что вторая ветвь стебля на рис.1Б всего лишь похожа на комплиментарную цепь ДНК (но этой цепью не является!), и обозначенный эффект возникает за счет поворота исходной цепи в области шпилечной петли, которая соответствует цепочке из семи нуклеотидов U·C·U·A·U·A·A из центральной части рис. 1А.

В данной работе изучаются редуцированные множества, аналогичные [4], но в них учитываются также и неканонические пары G-U. Поскольку неканонические пары присутствуют только в стеблях (в перекрытиях их не может быть), то по аналогии с [4] введем в рассмотрение три редуцированных множества элементарных стеблей:

$$\Delta S_{-1}, \Delta S_0, \Delta S_{+1}. \quad (1)$$

Каждое из этих стеблей в этих множествах, отвечает сдвигам между кодонами из разных ветвей стебля: на -1, 0, +1 нуклеотид соответственно; на рис.1Б этот сдвиг равен 0. Предполагается что в каждом из элементарных стеблей каждого из множеств (1) непременно присутствует хотя бы одна неканоническая пара G-U. В этом смысле такие множества являются дополнительными к трем множествам из [4], поскольку именно неканонические пары являются как бы второстепенными, дополнительными по отношению к классическим каноническим парам (см. также ниже (2)). Расчеты множеств (1) на основе генетического кода привели к совершенно неожиданному результату, который в итоге позволил трактовать неизвестную ранее роль неканонической пары G-U. Оказалось, что во множествах (1) присутствуют многократные случаи неоднозначностей. Это такие случаи, когда для одной и той же пары аминокислот имеют место более одного элементарного стебля (в [4] речь шла о генетических перекрытиях и для их множеств стеблей это – одно и то же, если речь идет только о канонических парах). Такая особенность впервые изучалась на редуцированных множествах, содержащих лишь канонические пары [4]. Оказалось, что при этом неоднозначности возникают всего для шести пар аминокислот и всякий раз они были обу-

словлены только нерегулярными кодировками генетического кода. Речь идет о кодировках для трех аминокислот: серина Ser, лейцина Leu и аргинина Arg. Кроме того, обнаруженные неоднозначности были лишь двузначностями и имеют место лишь для двух множеств из трех. Совершенно иная картина наблюдается для множеств (1). Для них неоднозначности могут быть не только двузначными (см.рис.2) и не только связанными с нерегулярностями генетического кода (см. табл.1).

Таблица 1. Сравнительные данные о количестве элементарных стеблей и числе неоднозначностей во множествах S_{-1} , S_0 , S_{+1} и ΔS_{-1} , ΔS_0 , ΔS_{+1} . В скобках представлены величины, полученные с участием нерегулярных кодонных семейств.

| Множество | Количество элементарных стеблей | Число неоднозначностей |
|-----------------|---------------------------------|------------------------|
| S_{-1} | 20 (7) | 0 |
| S_0 | 26 (10) | 3 (3) |
| S_{+1} | 105 (42) | 3 (3) |
| Σ | 151 (59) | 6 (6) |
| | | |
| ΔS_{-1} | 22 (10) | 1 (1) |
| ΔS_0 | 36 (16) | 11 (4) |
| ΔS_{+1} | 81 (38) | 32 (22) |
| Σ | 139 (64) | 44 (27) |

На рис.2 представлены по одному неоднозначному элементарному стеблю из множеств (1), причем, в отличие от рис.1, здесь исключены вертикальные черточки.

Gly - Leu (ΔS_{-1})

| | |
|-----|-----|
| Gly | Gly |
| GGN | GGN |
| NUC | XUU |
| Leu | Leu |

Leu - Arg (ΔS_0)

| | | |
|-----|-----|-----|
| Leu | Leu | Leu |
| UUG | CUU | CUG |
| XGC | GGA | GGC |
| Arg | Arg | Arg |

Ser - Arg (ΔS_{+1})

| | | | |
|-----|-----|-----|-----|
| Ser | Ser | Ser | Ser |
| UCU | AGC | AGU | UCU |
| GGA | UGC | YGC | GGC |
| Arg | Arg | Arg | Arg |

Рис.2. Примеры неоднозначных элементарных стеблей из множеств (1): двузначные для Gly - Leu (ΔS_{-1}), трехзначные для Ala - Gly (ΔS_0) и четырехзначные для Ser - Arg (ΔS_{+1}). В отличие от рис.1, здесь исключены вертикальные черточки.

В таблице 1 это четко подчеркивается: оказалось, что неоднозначности, прежде всего по их числу, в 9 раз превышают подобные случаи для канонических пар. В этой таблице приводятся сравнительные данные с множествами элементарных стеблей S_{-1} , S_0 , S_{+1} , которые адекватны множествам элементарных перекрытий из [4] и отличаются от последних множественной модификацией каждого T (тимин) на U, которая происходит в ходе построения первичной структуры мРНК после считывания с ДНК (см. выше). Таким образом, для канонических пар имеем три множества:

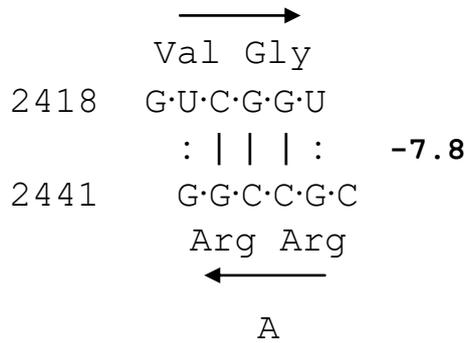
$$S_{-1} = W_{-} (T \rightarrow U), \quad S_0 = W_0 (T \rightarrow U), \quad S_{+1} = W_{+} (T \rightarrow U), \quad (2)$$

где множества W_- , W_0 , W_+ были определены ранее [4]. Описание некоторых элементов множеств S_{-1} , S_0 , S_{+1} , ΔS_{-1} , ΔS_0 , ΔS_{+1} . Представлено на рис 3.

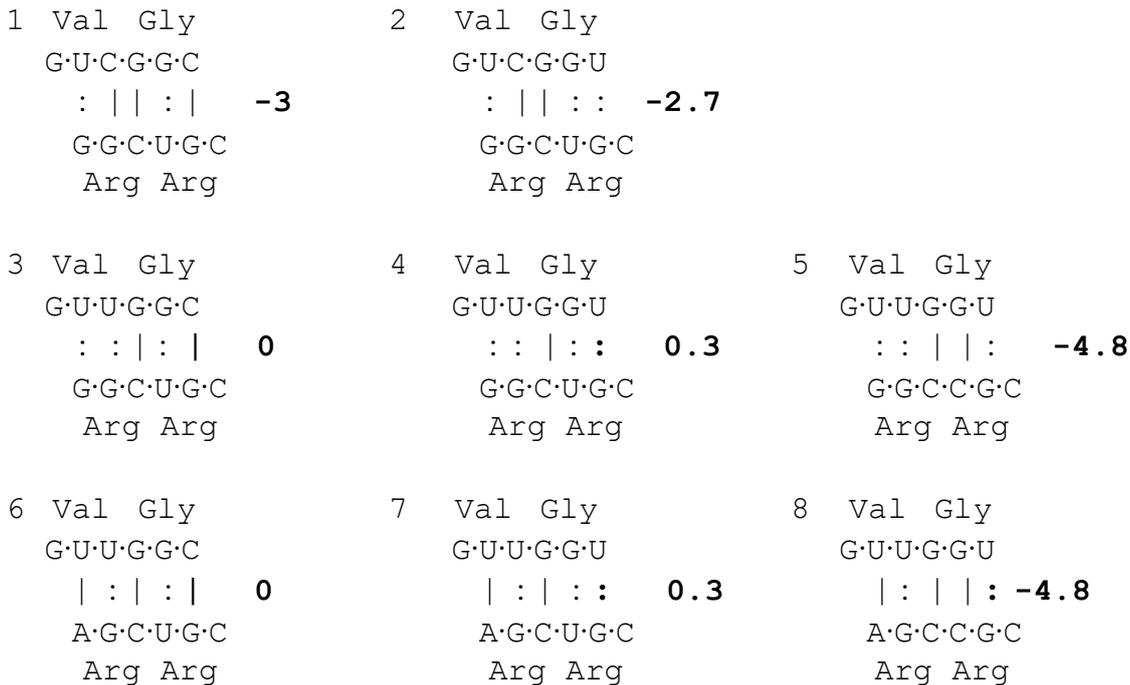
| | | | | | |
|---------------------|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | | 20 |
| | Met | Met | Trp | | Arg |
| $S_{-1}(20)$ | AUG | AUG | UGG | ... | CGN |
| | GUA | MUA | YAC | | NGC |
| | Met | Ile | His | | Arg |
| | 1 | 2 | 3 | | 26 |
| | Met | Trp | Phe | | Ser |
| $S_0(26)$ | AUG | UGG | UUU | ... | UCG |
| | UAC | ACC | AAA | | AGC |
| | His | Pro | Lys | | Arg |
| | 1 | 2 | 3 | | 105 |
| | Met | Met | Met | | Arg |
| $S_{+1}(105)$ | AUG | AUG | AUG | ... | CGC |
| | ACC | ACA | ACG | | CGC |
| | Pro | Thr | Ala | | Arg |
| | 1 | 2 | 3 | | 22 |
| | Met | Trp | Trp | | Leu |
| $\Delta S_{-1}(22)$ | AUG | UGG | UGG | ... | UUX |
| | NUG | GGU | YAU | | XGA |
| | Val | Trp | Tyr | | Arg |
| | 1 | 2 | 3 | | 36 |
| | Met | Met | Met | | Leu |
| $\Delta S_0(36)$ | AUG | AUG | AUG | ... | CUG |
| | UAU | UGU | UGC | | GGC |
| | Tyr | Cys | Arg | | Arg |
| | 1 | 2 | 3 | | 81 |
| | Met | Met | Met | | Arg |
| $\Delta S_{+1}(81)$ | AUG | AUG | AUG | ... | CGU |
| | GUA | AUA | GCC | | CGC |
| | Met | Ile | Pro | | Arg |

Рис.3. Описание некоторых элементов множеств S_{-1} , S_0 , S_{+1} , ΔS_{-1} , ΔS_0 , ΔS_{+1} .

Для иллюстрации главного результата исследования обратимся к рис.4.



Б



В

Рис.4. Представлен стебель вторичной структуры (А), занимающий часть фрагмента 2418-2441 РНК MS2 [7]. Он соответствует (Б) неоднозначным элементарным стеблям из множества ΔS_{+1} для двух пар аминокислот: Val-Arg (3случая) и Gly-Arg (3 случая). Во фрагменте В представлены все возможные стыковки данных элементарных стеблей, за исключением случая А; их оказалось 8.

На рис.4 представлен стебель вторичной структуры (А), занимающий часть фрагмента 2418-2441 РНК MS2 из [7]. Он соответствует элементарным стеблям из множества ΔS_{+1} , а именно элементарным стеблям для двух пар аминокислот: Val-Arg и Gly-Arg. Неоднозначной паре элементарных стеблей Val-Arg как и аналогичной паре Gly-Arg соответствуют 3 случая (фрагмент Б), притом каждый случай отличается от другого суммарной энергетикой связей. Во фрагменте В представлены все возможные стыковки данных элементарных стеблей, за исключением случая А; их оказалось 8. В табл.2 для каждой из версий указывается значение свободной энергии ΔG (ккал/моль), рассчитанной на основе известных правил из [9]. Из всех возможных комбинаций элементарных стеблей, случай, который соответствует стеблю А, имеет наибольший выигрыш свободной энергии. Возможные варианты других восьми версий (из-за неоднозначностей), указанных на рис.4, могут, согласно табл. 2, изменить свободную энергию на порядок и более. Для приведенных восьми версий для стеблей на рис.4 В оказалось, что значение ΔG изменяется в пределах от -4.8 до 0, в то время как подобное значение ΔG для стебля из рис.4 А равно -7.8. Подчеркнем, что при этом пары аминокислот как для верхней, так и для нижней ветвей для всех девяти версий стеблей сохраняются неизменными.

Таким образом, показано, что синхронное изменение кодировок для неоднозначных элементарных стеблей позволяет «регулировать» величину ΔG и даже для относительно небольших стеблей, содержащих всего 5 пар (из них всего две – неканонические пары G-U) эта величина может изменяться на порядок и более. Сравнение с подобными расчетами для неоднозначных элементарных стеблей, содержащих только канонические пары [4], показывает, что эффект для стеблей с неканоническими парами G-U оказался значительно большим. Это связано как с большим числом неоднозначных элементарных стеблей для неканонических пар: 54 против 6, так и с тем, что помимо двузначностей (для канонических пар) имеют место также как трех-, так и четырехзначности, см. рис.2. Итак, на основе строгой математической аналогии

стеблей мРНК и перекрытий генов из различных цепей ДНК, обнаружилась феноменальная роль неканонических пар G-U.

Авторы выражают благодарность Walter Fiers, Professor em. of Molecular Biology руководителю работы[7], за поддержку установленной нами аналогии, что позволило продолжить наши исследования по грантам РФФИ для случая неканонических связей и в итоге привело к полученному результату

Список литературы

1. Козлов Н.Н., Т.М. Энеев. Основы математической теории генетического кода. ДАН 2017 Т.473. № 3 С.268-271.
2. N. N. Kozlov. Mathematical Theory of Genetic Code Global Journal of Science Frontier Research: F Mathematics and Decision Sciences Volume 18 Issue 2 Version 1.0 Year 2018 PP 47-85. DOI:10.17406/GJSER
3. Rich A., Kim S.H. // Sci. Am. 1978. Vol. 238, №1. P. 52-62.
4. Козлов Н.Н. // ДАН 2010. Т. 432, № 2. С. 157-161.
5. Козлов Н.Н. // ДАН 2008. Т.423, № 3. С.295-299.
6. Козлов Н.Н. Элементарные генетические перекрытия. Препринт ИПМ им. М.В.Келдыша, РАН, М., 2004. № 64. 27с.
(http://www.keldysh.ru/papers/2004/rep64/rep2004_64.html (MS Word))
7. Fiers W., Contreras R., Duerinck F., et al. // Nature 1976. V. 260, p. 500-507.
8. Козлов Н.Н., Бахарев М.Г. Математический анализ одной биологической структуры. Препринт ИПМ им. М.В.Келдыша РАН, № 93, М. 2006, 28 с.
http://www.keldysh.ru/papers/2006/source/rep2006_93.doc
9. Кантор Р.Ю., Шиммель П. Биофизическая химия. Т. 3. Мир, М., 1985, 536 с.

Ниже приводится фрагмент вторичной структуры матричной РНК MS2 [7].
Рисунок 1 составлен на основе левого нижнего стебля.

