

Math-Net.Ru

All Russian mathematical portal

N. N. Kozlov, One model of computation of the genetic code, *Keldysh Institute preprints*, 2020, 052

DOI: 10.20948/prepr-2020-52

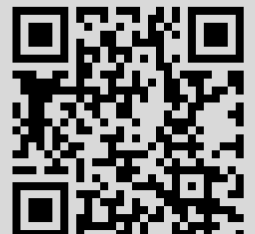
Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

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

Download details:

IP: 18.97.14.87

January 14, 2025, 06:50:24





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

[Препринты ИПМ](#) • [Препринт № 52 за 2020 г.](#)



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

[Козлов Н.Н.](#)

Одна модель расчета
генетического кода

Рекомендуемая форма библиографической ссылки: Козлов Н.Н. Одна модель расчета генетического кода // Препринты ИПМ им. М.В.Келдыша. 2020. № 52. 40 с.
<http://doi.org/10.20948/prepr-2020-52>
URL: <http://library.keldysh.ru/preprint.asp?id=2020-52>

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

Н.Н. Козлов

**Одна модель расчета
генетического кода**

Москва — 2020

Козлов Н.Н.

Одна модель расчета генетического кода

Представлено подробное описание метода расчета генетического кода, идея которого впервые опубликована ранее [3], а выбор одного из главных множеств для расчетов опирался на статью [4]. 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.

Ключевые слова: генетический код, перекрывающиеся гены, однородные перекрытия, расчет кода

Nicolay Nicolayevich Kozlov

One model of computation of the genetic code

For the first time provides a detailed description of the method of calculation of the genetic code, the idea of which was first published earlier [3], and the choice of one of the most important sets for the calculation was based on an article [4]. Such a set of amino acid corresponds to a complete set of representations of the plurality of overlapping triple gene belonging to the same DNA strand. A separate issue was the initial point, triggering an iterative search process all codes submitted by the initial data. Mathematical analysis has shown that the said set contains some ambiguities, which have been founded on the basis of our proposed compressed representation of the set. As a result, the developed method of calculation was limited to the two main stages of research, where the first stage only the of the area were used in the calculations. The proposed approach will significantly reduce the amount of computations at each step in this complex discrete structure.

Key words: genetic code, overlapping genes uniform overlap, account code

Введение

Данная статья содержит расширенную версию на русском языке работы, которая была опубликована на английском в открытом доступе. Ниже даны также данные по числу просмотров (1242) и скачиваний (672) на середину ноября 2019:

Computation of the Genetic Code: Full Version

N. N. Kozlov

Journal of Computer and Communications Vol.5 No.10

Full-Text HTML XML Pub. Date: August 30, 2017

DOI: 10.4236/jcc.2017.510008 **672** Downloads **1 242** Views

Первая оценка этой работы по ряду критериев была дана на конференции в Цюрихе — это были только две оценки: хорошо и очень хорошо (см. Файл оцЦюрихв приложении). Нами был сделан доклад на этой конференции:

Nicolay N. Kozlov, Olga N. Kozlova. Computation of the Genetic Code / Proc. Of the Fifth Intl. Conf. Advances in Bio-Informatics, Bio-Technology and Environmental Engineering — ABBE, 02-03 September, 2017 Zurich, Switzerland. Copyright @Institute of Research Engineers and Doctors, USA. All rights reserved. ISBN: 978-1-63248-130-6 doi: 10. 15224/978-1-63248-130-6-05

Через полгода мне сообщили о новой стоимости моей стартовой книги по математическому анализу генетического кода (см. файл Amazon в приложении). Теперь она в 17 раз дороже, чем в РФ.

Автору неизвестны постановки и решения в мире подобного класса задач

Идея расчета генетического кода возникла у нас после многолетних исследований по математической генетике. Основная идея состояла в том, что код, по-видимому, сегодня — через полвека после его открытия — может быть вычислен на основе экспериментальных данных, уже известных к настоящему времени. Подход, предлагаемый ниже, не является главным или всеобъемлющим, и его можно рассматривать всего лишь как одну из попыток, чтобы первоочередной задачей стала задача поиска тех перекрытий, где участвует одна и та же аминокислота (в перекрытиях она участвует с разными кодировками, согласно генетическому коду). По этому поводу было проведено специальное исследование.

1. Теорема для однородных перекрытий

Нами рассматриваются необычные способы записи генетической информации – перекрывающиеся гены, когда один и тот же участок ДНК соответствует более чем одному белку. Нами были исследованы все 5

возможных случаев перекрытий генов, разрешенных структурой ДНК, которые были изучены ранее [5]. Это изучение основывалось на математическом анализе всех 5 возможных случаев перекрытий и опиралось на множества так называемых элементарных генетических перекрытий — э.п., или перекрытий, соответствующих паре одиночных аминокислот. В [6] представлен краткий анализ таких множеств, а окончательная версия содержится в [2]. На рис. 1 представлено описание структуры множеств W_1 - W_5 , и приведены по 4 э.п. в каждом из этих множеств.

$W_1(80)$	Met Tyr TATG 1	Met His CATG 2	Met Asn AATG 3	...	Arg Ser TCGN 80
$W_2(80)$	Met Trp ATGG 1	Met Cys ATGY 2	Trp Gly TGGN 3	...	Arg Gly ZGGN 80
$W_3(35)$	Met ATG GTA Met 1	Met ATG MTA Ile 2	Trp TGG YAC His 3	...	Arg AGX NTC Leu 35
$W_4(52)$	Met ATG TAC His 1	Trp TGG ACC Pro 2	Phe TTT AAA Lys 3	...	Arg CGC GCG Ala 52
$W_5(196)$	Met ATG ACC Pro 1	Met ATG ACA Thr 2	Met ATG ACG Ala 3	...	Arg AGG CCT Ser 196

Рис. 1. Описание структуры множеств W_1 - W_5 . Приведены по 4 э.п. в каждом из множеств. В скобках указывается полное число э.п. в соответствующем множестве.

Принципиальная позиция данного исследования обозначена в [2], где было показано, что представленный перечень элементарных перекрытий позволяет строить любое (!) допустимое структурой генетического кода перекрытие не

только 2, но и всех допустимых перекрытий от 3 до 6 генов. Актуальность задач обусловлена современной ситуацией: перекрывающиеся гены, общепринятые в вирусах, митохондриях, бактериях и плазидах, были обнаружены в последующие годы в эукариотических больших геномах, в том числе человека, причем число перекрытий, как правило, является значительным, для генома человека это около 1700 [7].

При математическом анализе перекрытий более двух генов нами были исследованы некоторые задачи. Конечно, можно было бы построить множества всех э.п. от 3 до 6 генов. Это сделать нетрудно с помощью современных вычислительных средств. Однако, главное — какие новые выводы это может дать. И поэтому мы идем традиционным путем — от задач. Остановимся сначала кратко лишь на некоторых из них, решения для которых нами уже опубликованы. Первая из них относится к анализу неоднозначностей [8-10], это когда одной и той же паре аминокислот соответствуют две кодировки (см. пример ниже). Другая задача была связана с построением и анализом множества элементарных перекрытий для 3 генов, перекрывающихся в одной и той же цепи ДНК. Установлено, что подобных перекрытий всего 307. На основе таких перекрытий была поставлена новая задача, связанная с расчетом генетического кода математическими методами [11, 12]. Вопрос о том, почему именно такое множество было выбрано для расчета кода, был основан на теореме, которая была опубликована сравнительно недавно [4]. Речь идет о расчете и анализе всех однородных э.п. от 2 до 6 генов — это э.п., соответствующие одной и той же аминокислоте. Решение ее дает нижеследующая теорема.

Назовем э.п.- i элементарным перекрытием для i аминокислот, где $i \in (2,6)$. Тем самым введенные ранее в [2] э.п. будут именоваться как э.п.-2. На рис. 2 показано общее представление для э.п.-6.

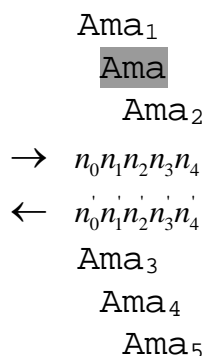
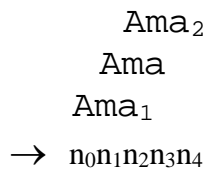


Рис. 2. Общее представление для э.п.-6. (см. текст)

Для аминокислоты $A_m a$ (выделена штриховкой), кодируемой триплетом $n_1 n_2 n_3$, имеют место 5 альтернативных аминокислот $A_m a_1 - A_m a_5$, кодировки которых образованы сдвигами $-1, +1$ в той же цепи ДНК (\rightarrow) и $-1, 0, +1$ в комплементарной цепи ДНК (\leftarrow). Обозначения $n_i, i \in (0,4)$ — нуклеотиды из

набора A, T, C, G ; n'_i $i \in (0,4)$ – комплементарные составляющие: т.е. для $n_i=A$, $n'_i=T$; $n_i=C$, $n'_i=G$ для любых $i \in (0,4)$ и наоборот. Для того чтобы последовательно выделить э.п.-2 для всех 5 случаев парных перекрытий из [2,3], на рис. 3 следует последовательно оставить одну из 5 пар аминокислот: (Ama, Ama_i) , $i \in (1,5)$. Для того чтобы последовательно выделить все случаи перекрытий для э.п.-3, на рис. 3 следует оставить Ama и пару аминокислот из 5 возможных, или имеем 10 случаев тройных перекрытий, и т.д. Обратимся к поиску однородного подмножества для всех возможных э.п.-2 – э.п.-6, или подмножества, в каждом э.п. которого присутствует одна и та же аминокислота. Обозначим его через W_{2-6} . Для представления однородных э.п. приняты следующие обозначения: $Y: T, C$; $X: A, G$; $N: A, G, C, T$.

(А)



(Б)

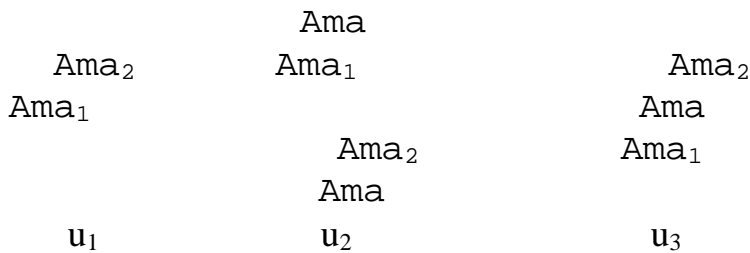


Рис. 3. (А) Для аминокислоты Ama , кодируемой триплетом $n_1n_2n_3$, имеют место 2 альтернативных аминокислоты Ama_1 и Ama_2 . Кодировки их образованы сдвигами $-1, +1$ в той же цепи ДНК (\rightarrow). Кодон $n_0n_1n_2$ для Ama_1 перекрывается с кодоном $n_1n_2n_3$ для Ama – перекрытие содержит два нуклеотида n_1n_2 ; кодон $n_2n_3n_4$ для Ama_2 перекрывается с кодоном $n_1n_2n_3$ для Ama – перекрытие содержит два нуклеотида n_2n_3 . В итоге тройное перекрытие содержит всего одну общую позицию n_2 . (Б) Элементы множеств сочетаний аминокислот, образованные на основе элементарного перекрытия из фрагмента А. Слева – один элемент множества U_1 , в центре – два элемента множества U_2 и справа – один элемент множества U_3 .

Имеет место

Теорема. Однородное подмножество W_{2-6} содержит всего 31 э.п., принадлежащих лишь наборам э.п.-2 – э.п.-3, из которых:

- 5 э.п.-2 для перекрытий из одной цепи ДНК:

Phe	Lys	Pro	Gly	Leu	
Phe	Lys	Pro	Gly	Leu	(1)
TTYT	AAAX	CCCN	GGGN	CTTX	
1	2	3	4	5	

- 19 э.п.-2 для перекрытий из различных цепей ДНК:

Met	Tyr	Ile	Ala	Arg			
ATG	TAY	ATM	GCN	CGN			
GTA	YAT	MTA	NCG	NGC			
Met	Tyr	Ile	Ala	Arg			
6	7	8	9	10			
							(2)
Tyr	His	Asn	Asp	Cys	Ile	Val	
TAT	CAT	AAT	GAT	TGC	ATA	GTA	
TAT	TAC	TAA	TAG	CGT	ATA	ATG	
Tyr	His	Asn	Asp	Cys	Ile	Val	
11	12	13	14	15	16	17	

Pro	Thr	Ala	Gly	Ser	Ser	Leu	Arg
CCG	ACG	GCG	GGC	TCG	AGC	YTA	CGC
GCC	GCA	GCG	CGG	GCT	CGA	ATY	CGC
Pro	Thr	Ala	Gly	Ser	Ser	Leu	Arg
18	19	20	21	22.1	22.2	23	24

- 4 э.п.-3 для перекрытий из одной цепи ДНК:

	Phe	Lys	Pro	Gly	
	Phe	Lys	Pro	Gly	(3)
	Phe	Lys	Pro	Gly	
	TTTTY	AAAAX	CCCCN	GGGN	
	1	2	3	4	

- 3 э.п.-3 для перекрытий из различных цепей ДНК:

	Pro	Gly	Leu	
	Pro	Gly	Leu	
	CCCG	GGGC	CTTA	(4)
	GGGCC	CCCGG	GAATY	
	Pro	Gly	Leu	
	5	6	7	

Дадим комментарий к формулировке теоремы. Случаи 22.1 и 22.2 соответствуют одной и той же аминокислоте Ser-Ser; это одна из шести неоднозначностей, которая была установлена ранее [9]. Из теоремы следует, что не существуют однородные э.п. больше 3-го порядка, или э.п.-4 – э.п.-6. Для доказательства этого факта отметим, что все перечисленные выше э.п. из (1)–(4) имеют одно общее свойство. Оно состоит в том, что кодировки аминокислот в каждом из этих перекрытий присутствуют в сдвинутых фазах: сдвиги на -1 либо +1 нуклеотид, причем независимо от цепи ДНК, или нет ни одного э.п., где пара перекрывающихся кодонов была бы в одной и той же фазе.

Из (2) следует, что подобные э.п. отсутствуют при перекрытиях из различных цепей ДНК, поэтому они невозможны в структурах однородных э.п.4–э.п.6.

2. Введение основных множеств

Как следует из предыдущего раздела, однородные элементарные перекрытия имеют место только для перекрытий с участием не более трех аминокислот. Для выбора рабочих множеств рассмотрим все такие перекрытия.

Прежде всего следует исключить из рассмотрения все однородные перекрытия, в которых участвуют две цепи ДНК. Рассмотрение этих перекрытий требует введения двойной цепи ДНК — это дополнительное условие в задаче. Исключая подобные однородные перекрытия, мы исходим из принципа построения алгоритма с минимальным количеством условий. Поэтому в нашем рассмотрении остаются только однородные перекрытия, принадлежащие одной и той же цепи ДНК: для пар аминокислот (1) – их всего 5, и аналогичные перекрытия для трех аминокислот (3) — их всего 4. Тем самым нами были выбраны главные рабочие множества э.п., а именно те, в которых присутствуют данные однородные перекрытия. Окончательная версия этих множеств представлена на стр. 312-319 в [2].

Рассмотрим более подробно вопрос о множествах с этими перекрытиями. Ранее [2] нами было введено понятие элементарного перекрытия применительно только к перекрытиям пар генов. Обобщим это понятие для трех генов, принадлежащих одной и той же цепи ДНК. Под термином элементарное перекрытие понимаем перекрытие для кодонов одиночных аминокислот по максимальному числу позиций. На фрагменте А рис. 3 для аминокислоты А_{ma}, кодируемой триплетом $p_1p_2p_3$, указаны альтернативные аминокислоты А_{ma1} и А_{ma2}, кодировки которых $p_0p_1p_2$ и $p_2p_3p_4$ соответственно образованы сдвигами на -1 и +1 нуклеотид в той же цепи ДНК (→). Предполагается при этом, что все значения p_0 – p_4 принадлежат каноническому набору из четырех нуклеотидов. На основе фрагмента А рис. 3 можно построить три вида сочетаний аминокислот, представленных на фрагменте Б рис. 3 и обозначенных соответственно u_1 , u_2 , u_3 : одно u_1 для перекрытия по

одной позиции, два u_2 для перекрытия по двум позициям и одно u_3 для перекрытия 3 аминокислот.

Следует отметить, что в генетических экспериментах более 20 лет назад нами были обнаружены все допустимые генетическим кодом элементы u_1 , u_2 , u_3 , которые будем рассматривать соответственно как элементы множеств U_1 , U_2 , U_3 . Число элементов U_2 равно 160, а элементов U_3 – 307; полный их перечень представлен в [2]. Что касается множества U_1 , то в ходе решения были использованы лишь небольшая часть элементов этого множества и его обобщенные характеристики, которые приводятся в [2].

Ниже в рисунках 4 и 5 представлены только фрагменты этих множеств из Приложений 1 и 2, а некоторые их характеристики представлены в Табл. 1.

1	Met	Met	Met	Met			
Met	Tyr	His	Asn	Asp			
(4+2)	1	2	3	4			
	Met	Met					
	Trp	Cys					
	5	6					
20	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Arg	Gln	Lys	Glu	Pro	Thr	Ala	Ser
(7+5)	149	150	151	152	153	154	155
	Arg	Arg	Arg	Arg	Arg		
	Asp	Glu	Val	Ala	Gly		
	156	157	158	159	160		

Рис.4. Приводится перечень некоторых элементов множества U_2 . Речь идет о двух наборах таких элементов, соответствующих первой (Met) и последней (Arg) аминокислотам из соответствующего набора, представленных в первом столбце Табл. 1. Каждое из представлений содержит по две строки: первая соответствует сдвигу между кодонами, равному -1, а вторая – равному +1. Под названием аминокислоты указывается число элементов, соответствующих этим сдвигам, а нижние номера соответствуют номерам в полном перечне этих элементов.

Trp	Trp	Trp	Trp	Cys	Cys	Cys	Cys
Met	Met	Met	Met	Met	Met	Met	Met
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
1	2	3	4	5	6	7	8
Asp	Asp	Asp	Asp	Glu	Glu	Glu	Glu
Arg	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
279	280	281	282	283	284	285	286
Val	Val	Val	Val	Ala	Ala	Ala	Ala
Arg	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
287	288	289	290	291	292	293	294
Gly	Gly	Gly	Gly	Asp	Asp	Asp	
Arg	Arg	Arg	Arg	Arg	Arg	Arg	
Pro	Thr	Ala	Ser	Gln	Lys	Glu	
295	296	297	298	299	300	301	
Glu	Glu	Glu	Gly	Gly	Gly		
Arg	Arg	Arg	Arg	Arg	Arg		
Gln	Lys	Glu	Gln	Lys	Glu		
302	303	304	305	306	307		

Рис. 5. Приводятся элементы с номерами 1-8 и 279-307 из множества U_3 , соответствующие первой (Met) и последней (Arg) аминокислотам из перечня аминокислот, представленных в табл. 1. Элементы приводятся в том представлении, которое используется в данной задаче. Каждый из элементов состоит из трех строк: верхней, средней и нижней. Названные аминокислоты Met и Arg приводятся в средней строке.

Таблица 1

Некоторые данные множеств U_2 и U_1 для аминокислот Ama_i , $i \in (1,20)$. Первые 4 столбца указывают на число перекрытий, а в скобках – перечень аминокислот для перекрытий: по 1 и 2 основаниям (m_{12}), по 2 и 3 основаниям (m_{23}) по 1 основанию с третьим основанием (m_{1-3}), по 3 основаниям с первым основанием (m_{3-1}). В столбцах 3 и 4 речь идет только о перекрытиях с Lys, Phe, Pro, Gly, поэтому это число не может быть более 4. В последнем столбце приводятся вычисленные значения K. Обозначения: X: a, d; Y: b, c; M: a, b, c; N: a, b, c, d.

i	Ama_i	U_2			U_1		K
		m_{12}	m_{23}	m_{1-3}	m_{3-1}		
1	Met	4(Tyr,His,Asn,Asp)	2(Trp,Cys)	3(Lys,Pro,Gly)	1(Gly)	abd	
2	Trp	3(Met,Val,Leu)	1(Gly)	3(Phe,Pro,Gly)	1(Gly)	bdd	
3	Phe	4(Phe,Ile,Val,Leu)	3(Phe,Ser,Leu)	3(Phe,Pro,Gly)	2(Phe,Pro)	bbY	
4	Tyr	3(Ile,Val,Leu)	3(Met,Ile,Thr)	3(Phe,Pro,Gly)	2(Phe,Pro)	baY	
5	His	4(Pro,Thr,Ala,Ser)	3(Met,Ile,Thr)	3(Phe,Pro,Gly)	2(Phe,Pro)	caY	
6	Asn	3(Gln,Lys,Glu)	3(Met,Ile,Thr)	3(Lys,Pro,Gly)	2(Phe,Pro)	aaY	
7	Asp	2(Gly,Arg)	3(Met,Ile,Thr)	3(Lys,Pro,Gly)	2(Phe,Pro)	daY	
8	Cys	3(Met,Val,Leu)	2(Val,Ala)	3(Phe,Pro,Gly)	2(Phe,Pro)	bdY	
9	Gln	4(Pro,Thr,Ala,Ser)	4(Asn,Lys,Ser,Arg)	3(Phe,Pro,Gly)	2(Lys,Gly)	caX	
10	Lys	3(Gln,Lys,Glu)	4(Asn,Lys,Ser,Arg)	3(Lys,Pro,Gly)	2(Lys,Gly)	aaX	
11	Glu	2(Gly,Arg)	4(Asn,Lys,Ser,Arg)	3(Lys,Pro,Gly)	2(Lys,Gly)	daX	
12	Ile	4(Tyr,His,Asn,Asp)	4(Phe,Tyr,Ser,Leu)	3(Lys,Pro,Gly)	3(Phe,Lys,Pro)	abM	
13	Val	4(Cys,Gly,Ser,Arg)	6(Trp,Phe,Tyr,Cys,Ser,Leu)	3(Lys,Pro,Gly)	4	dbN	
14	Pro	4(Pro,Thr,Ala,Ser)	5(His,Gln,Pro,Leu,Arg)	3(Phe,Pro,Gly)	4	ccN	
15	Thr	4(Tyr,His,Asn,Asp)	5(His,Gln,Pro,Leu,Arg)	3(Lys,Pro,Gly)	4	acN	
16	Ala	4(Cys,Gly,Ser,Arg)	5(His,Gln,Pro,Leu,Arg)	3(Lys,Pro,Gly)	4	dcN	
17	Gly	3(Trp,Gly,Arg)	5(Asp,Glu,Val,Ala,Gly)	3(Lys,Pro,Gly)	4	ddN	
18	Ser	7(Phe,Gln,Lys,Glu,Ile,Val,Leu)	7(His,Gln,Val,Pro,Ala,Leu,Arg)	4	4	bcN adY	
19	Leu	8(Phe,Ile,Val,Pro,Thr,Ala,Ser,Leu)	6(Trp,Phe,Tyr,Cys,Ser,Leu)	3(Phe,Pro,Gly)	4	cbN bbX	
20	Arg	7(Gln,Lys,Glu,Pro,Thr,Ala,Ser)	5(Asp,Glu,Val,Ala,Gly)	4	4	cdN adX	

	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly	Ser	Leu	Arg
Met													Cys			Cys	Trp			
Trp							Gly				Gly		Gly			Gly				
Phe		Leu	Phe	Leu	Ser			Leu	Ser					Ser				Phe	A9	Ser
Tyr		Met	Ile	Ile	Thr			Met	Thr				Thr	Thr				Ile	A6	Thr
His		Met	Ile	Ile	Thr			Met	Thr					Thr				Ile	A6	Thr
Asn		Met	Ile	Ile	Thr			Met	Thr					Thr				Ile	A6	Thr
Asp		Met	Ile	Ile	Thr			Met	Thr					Thr				Ile	A6	Thr
Cys		Val	Val	Val	Ala			Val	Ala					Ala				Val	A8	Ala
Gln	Asn					Lys	Arg			Lys	Arg	Asn	Ser		Asn	Ser	Arg	Lys		Lys
Lys	Asn					Lys	Arg			Lys	Arg	Asn	Ser		Asn	Ser	Arg	Lys		Lys
Glu	Asn					Lys	Arg			Lys	Arg	Asn	Ser		Asn	Ser	Arg	Lys		Lys
Ile	Tyr	Leu	Phe	Leu	Ser			Leu	Ser			Tyr		Ser	Tyr			Phe	A9	Ser
Val	Tyr	Leu	Phe	Leu	Ser			Leu	Ser			Tyr	Cys	Ser	Tyr	Cys	Trp	Phe	A9	Ser
Pro	His	Leu	Leu	Leu	Pro	Gln	Arg	Leu	Pro	Gln	Arg	His	Arg	Pro	His	Arg	Arg	A1	A11	A5
Thr	His	Leu	Leu	Leu	Pro	Gln	Arg	Leu	Pro	Gln	Arg	His	Arg	Pro	His	Arg	Arg	A1	A11	A5
Ala	His	Leu	Leu	Leu	Pro	Gln	Arg	Leu	Pro	Gln	Arg	His	Arg	Pro	His	Arg	Arg	A1	A11	A5
Gly	Asp	Val	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly	A3	A8	A4
Ser	His	A10	A10	A10	A7	Gln	Arg	A10	A7	Gln	Arg	His	Arg	A7	His	Arg	Arg	A2	A12	A13
Leu	Tyr	Leu	Phe	Leu	Ser			Leu	Ser			Tyr	Cys	Ser	Tyr	Cys	Trp	Phe	A9	Ser
Arg	Asp	Val	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly	A3	A8	A4

Рис. 6. Сжатое представление для 307 элементов главного множества – U_3 : для каждой A_{α} этого множества (оно указывается в соответствующей клетке) дается аминокислота A_{α_1} по оси абсцисс, а по оси ординат – A_{α_2} (см. рис. 1А).

	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly
Met													Cys			Cys	Trp
Trp							Gly			Gly			Gly			Gly	Gly
Phe			Phe														
Tyr		Met	Ile	Ile	Thr			Met	Thr					Thr			
His		Met	Ile	Ile	Thr			Met	Thr					Thr			
Asn		Met	Ile	Ile	Thr			Met	Thr					Thr			
Asp		Met	Ile	Ile	Thr			Met	Thr					Thr			
Cys		Val	Val	Val	Ala			Val	Ala					Ala			
Gln	Asn					Lys				Lys		Asn			Asn		
Lys	Asn					Lys				Lys		Asn			Asn		
Glu	Asn					Lys				Lys		Asn			Asn		
Ile	Tyr		Phe									Tyr			Tyr		
Val	Tyr		Phe									Tyr	Cys		Tyr	Cys	Trp
Pro	His				Pro	Gln			Pro	Gln		His		Pro	His		
Thr	His				Pro	Gln			Pro	Gln		His		Pro	His		
Ala	His				Pro	Gln			Pro	Gln		His		Pro	His		
Gly	Asp	Val	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly

Рис. 7. Редуцированная область рис. 6: в ней присутствуют области, по которым проводится расчет кода. Все заштрихованные области отсекаются по указанным выше причинам, а главная область в расчете представлена без штриховки.

3. Постановка задачи. Введение сжатого множества

ПОСТАНОВКА ЗАДАЧИ. Пусть имеем набор из 4 букв: N: a, b, c, d, а также триплеты – любые тройки из этих букв, всего их 64. При этом каждая из 20 канонических аминокислот может кодироваться произвольным сочетанием таких триплетов. Задача состоит в поиске всех генетических кодов, отвечающих всем элементам обозначенных выше трех множеств U_1 , U_2 , U_3 , соответствующих генетическим экспериментам.

Для дальнейшего используем стандартные трехбуквенные сокращения для каждой из 20 аминокислот.

В [11] говорилось о шагах расчета данной задачи, но не говорилось, каким образом выделялись требуемые элементы на шаге. Для такого выделения было введено в рассмотрение специальное представление основного рабочего множества. Подробно данный результат публикуется впервые.

Введем в рассмотрение одно сжатое представление для 307 элементов главного множества- U_3 . На рис.6 для каждой A_{α} этого множества (оно указывается в соответствующей клетке) дается аминокислота A_{α_1} по оси абсцисс, а по оси ординат – A_{α_2} (см. рис. 3А). Оказалось, что полученное представление не является однородным, а содержит множественные неоднозначности: это случаи, когда одним и тем же значениям A_{α_1} и A_{α_2} соответствует более одного значения A_{α} . Оказалось, что число неоднозначностей в пределах от 2 до 4.

Эти случаи заштрихованы на рис. 6, они обозначены через A1 -A13, т.е. их всего 13, хотя на рисунке приведены 34 штриховки. Дело в том, что на этом рисунке A6, A9 и A10 представлены по 4 раза, A1, A5, A7, A8 и A11 – по три раза, A3 и A4 – по 2 раза, а A2, A12 и A13 – только по одному разу. Имеем A1 Gln, Leu; A2: Gln, Val, Leu;... A13: Gln, Pro,

Следует отметить, что указанные неоднозначности соответствуют значениям Ser, Leu, Arg как по оси абсцисс, так и по оси ординат. Однако наиболее значительная область из рис. 6 – та, которая соответствует случаям где по обоим осям нет ни одной из аминокислот из набора Ser, Leu, Arg. Для наших расчетов последнюю область редуцируем, исключив из нее все клетки, содержащие Ser, Leu, Arg. На рис. 7 штриховка соответствует названным трем аминокислотам, а ненулевые элементы незаштрихованной области имеют следующее свойство: каждое значение A_{α} является единственным для соответствующей пары A_{α_1} и A_{α_2} .

Таблица 2

Шаги 1-9 итерационного процесса вычисления генетического кода. В последней строке указано число вычисленных кодонов по окончании шага. Обозначения: X: a, d; Y: b, c; M: a, b, c; N: a, b, c, d.

	1	2	3	4	5	6	7	8	9
Met							abd		
Trp				bdd					
Phe	bbb	bbY							
Tyr						baY			
His						caY			
Asn						aaY			
Asp						daY			
Cys				bdY					
Gln			caX						
Lys	aaa	aaX							
Glu			daX						
Ile							abM		
Val			dbN						
Pro	ccc	ccN							
Thr					acN				
Ala			dcN						
Gly	ddd	ddN							
Ser								bca bcc ddY	bcd bcb
Leu								cbX cbb bbX	cbc
Arg								cdN adX	
Σ	4	12	24	27	31	35	43	58	61

Указанное выше свойство позволило обратиться к первому этапу расчета, когда производится расчет кодировок для всех элементов значение A_{ma} на основе кодировок для соответствующей пары A_{ma_1} и A_{ma_2} . Результаты пошагового решения задачи представлены в Табл. 2, но важнейшим этапом исследования стал вопрос о поиске начального приближения.

4. Начальное приближение

РЕШЕНИЕ ЗАДАЧИ. Используем стандартные трехбуквенные сокращения для каждой из 20 аминокислот, указанные в первом столбце Табл. 1. Имеем набор A^0

$$A^0: Ama_i, \quad i \in (1, 20). \quad (5)$$

Введем определение. Обратимся к введенным ранее однородным перекрытиям. Как и ранее, назовем сочетание аминокислот, построенное на основе элементарного генетического перекрытия, однородным, если в нем участвует одна и та же аминокислота. Для однородных элементов из множества имеет место

Свойство. Пусть кодировки Ama для однородного u_3 имеют одно из трех следующих представлений:

$$n_1N_1N_2, \quad N_3n_2N_4, \quad N_5N_6n_3, \quad (6)$$

где малыми буквами обозначены единичные составляющие набора N , а большими – какие-либо подмножества этого набора, вплоть до N . Тогда однородный u_3 может существовать только при использовании хотя бы одного базового триплета или триплета с тремя одинаковыми буквами.

Для доказательства осуществим последовательно подстановки каждого из представлений (6) в u_3 :

$$\begin{array}{ccc}
 Ama & Ama & Ama \\
 Ama & Ama & Ama \\
 Ama & Ama & Ama \\
 \text{н.п.} & n_1n_1n_1n'_1n'_2 & n'_3n_2n_2n_2n'_4 & n'_5n'_6n_3n_3n_3
 \end{array} \quad (7)$$

где n'_i – одиночная составляющая набора N_i , где $i \in (1, 6)$, а строка н.п. – нуклеотидные последовательности, которые образуются после указанной подстановки. В первом случае в (3) базовый кодон $n_1n_1n_1$ был использован при кодировке аминокислоты Ama из нижней позиции, во втором – $n_2n_2n_2$ – из средней позиции, а в третьем – $n_3n_3n_3$ – из верхней позиции.

Обратимся к однородным u_3 из множества U_3 , которых оказалось 4:

$$\begin{array}{cccc}
 \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly} \\
 \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly} \\
 \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly}
 \end{array} \quad (8)$$

В рамках предположения, указанного в Свойстве, предлагается следующий пошаговый процесс поиска генетического кода; см. Таблицу 2.

Шаг 1. Аминокислотам из (8) присвоим соответствующие базовые кодоны. Такое присвоение не является однозначным. Однако в нашем подходе набор букв N: a, b, c, d не сопоставлен с каноническим набором 4 нуклеотидов; об этом речь пойдет в конце работы. Поэтому далее будем оперировать лишь с одним из представлений для аминокислот из (4), которым присвоим соответственно следующие базовые триплеты:

$$\text{Lys: aaa, Phe: bbb, Pro: ccc, Gly: ddd} \quad (9)$$

Для дальнейших расчетов обратимся к некоторым обобщенным данным по множествам U_2 и U_1 , которые приводятся в Таблице 1.

Шаг 2. Из Табл. 1 следует, что как в столбце m_{12} (указывается число и дается перечень перекрываемых аминокислот по 1 и 2 основаниям), так и в столбце m_{23} (аналогичные данные по 2 и 3 основаниям) не содержится взаимных перекрытий между аминокислотами из (8). Такие перекрытия имеют место только по одной позиции и принадлежат множеству U_1 . Имеем:

$$\begin{array}{cccc} \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly} \\ \text{Lys} & \text{Phe} & \text{Lys} & \text{Lys} \\ (\text{Gly}) & (\text{Pro}) & (\text{Pro}) & (\text{Phe}) \\ & & (\text{Gly}) & (\text{Pro}) \\ & & (\text{Phe}) & (\text{Gly}) \end{array} \quad (10)$$

$$\begin{array}{cccc} \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly} \\ \text{Lys} & \text{Phe} & \text{Pro} & \text{Gly} \\ (\text{Pro}) & (\text{Pro}) & (\text{Phe}) & (\text{Lys}) \\ (\text{Gly}) & (\text{Gly}) & (\text{Gly}) & (\text{Pro}) \end{array}$$

где первые 4 элемента u_1 соответствуют перекрытиям для 3 позиций (в скобках указаны альтернативные варианты, см. столбец m_{3-1} из табл. 1), последующие 4 – перекрытиям для первых позиций (см. столбец m_{1-3}). Формальная подстановка базовых кодонов из (9) в (10) приводит кодировки всех 4 аминокислот к тому, что они становятся неоднозначными по 1-й и 3-й позициям. Для уточнения приведем вывод всего для двух аминокислот из (9) – Lys и связанной с ней согласно первому перекрытию из (10) Gly. Согласно сказанному имеем: Lys должна кодироваться набором триплетов X_1aX , а Gly – X_1dN , где X_1 : a, d, c; X: a, d. Тогда имеет место перекрытие Lys с Gly по двум позициям, что невозможно согласно Табл. 1. Это также не допускает две другие возможности: Lys не может кодироваться набором триплетов X_1aa , если Gly - ddN , а также Lys не может кодироваться набором триплетов aaX , если Gly - X_1dd . Остаются

еще две возможности: и Lys и Gly кодируются неоднозначными кодонами по одним и тем же позициям. Случай кодировок Lys: X₁aa, Gly: X₁dd невозможен, так как при этом не может быть удовлетворено условие из табл. 1: число m₂₃ для Gly равно 5 согласно Табл. 1 (а для подобной кодировки Gly может быть максимум -4). Поэтому остается единственный из возможных вариантов для рассматриваемых кодировок, когда имеют место неоднозначности в третьей позиции. Аналогично можно установить кодировки для оставшейся пары Phe и Pro. В итоге получим:

$$\text{Lys: aaX, Phe: bbY, Pro: ccN, Gly: ddN,} \quad (11)$$

где Y: b, c.

Из Табл. 1 следует, что значение m₁₂ не превышает числа 4 для аминокислот из (5) с номерами от 1 до 17. Число 4 означает, что первые и вторые позиции могут быть однозначными, чего никак нельзя сказать о m₁₂ для Ser, Leu, Arg, для которых подобные значения равны 7, 8, 7 соответственно. В связи с этим на последующих шагах 3-7 будут рассматриваться только аминокислоты из (5) с номерами до 17. Отметим, что расчет кодировок для всех аминокислот из (5) проводится по методу, опубликованному в [3].

5. Пошаговый расчет в области однозначности

Поиск решения на шаге 3 иллюстрирует рис. 8 (шаг 3), где представлена указанная выше редуцированная область однозначности. Для каждой из четырех аминокислот из (9) проводим по две полосы (серая штриховка на рисунке) для A_{ma1} (горизонтальная полоса, а цифра 3 вне рисунка указывает номер шага) и A_{ma2} (вертикальная полоса). В итоге на пересечениях этих полос находим всего 4 аминокислоты: Gln, Glu, Val, Ala, которые выделены на рисунке полужирным шрифтом. С учетом принятой стандартной записи и кодировок из (9) имеем

	Lys	Lys	Phe	Pro	
	Gln	Glu	Val	Ala	
	Pro	Gly	Gly	Gly	(12)
н.п.	ссаа	ддаа	ддбб	ддсс	

где н.п. – нуклеотидная последовательность. Из (12) имеем однозначные кодировки для 4 аминокислот: Gln, Glu, Val, Ala, а с учетом (9) находим:

$$\text{Gln: caX, Glu: daX, Val: dbN, Ala: dcN.} \quad (13)$$

Тем самым показано, что использование введенного редуцированного множества приводит на шаге к минимальным затратам, по сравнению с прямым перебором для 307 элементов главного множества – U₃.

Шаг 4. Поиск решения пояснен на рис. 8 (шаг 4), где цифра 4 вне рисунка указывает на новые полосы, соответствующие решению из (13), а именно полосы для Gln, Glu, Val, Ala. В итоге имеем три перекрытия:

$$\begin{array}{ccc}
 \text{Gly} & \text{Val} & \text{Ala} \\
 \text{Trp} & \text{Cys} & \text{Cys} \\
 \text{Val} & \text{Val} & \text{Val}
 \end{array} \quad (14)$$

на основе которых получим

$$\text{Trp: bdd,} \quad \text{Cys: bdY} \quad (15)$$

Шаг 5. Во множестве U_3 нет элементов, у которых присутствуют по две аминокислоты из наборов (11), (13), (15): все полосы для цифры 5, которая указана вне рисунка, не имеют ни одного требуемого пересечения. В связи с этой особенностью мы обратились к множеству U_2 и рассмотрим одну из допустимых возможностей. Эта возможность не была случайной: она имела максимальное число перекрытий по сравнению с другими случаями. В итоге на основе двух элементов u_2 имеем:

$$\begin{array}{cc}
 \text{Thr} & \text{Thr} \\
 \text{Gln} & \text{Pro}
 \end{array} \quad (16)$$

Получим две кодировки для Thr: xca, xcc, где x пока неизвестно. Для поиска всех кодировок для Thr обратимся к ее значениям m_{1-3} и m_{3-1} . Из равенства $m_{1-3}=3$ (Lys, Pro, Gly) следует, что x может принимать не более двух значений: a, d, т.к. в перечисленном наборе отсутствует Phe. Однако значение d невозможно, т.к. dca и dcc являются кодировками Ala. Кроме того, из равенства $m_{3-1}=4$ (это означает, что первые позиции кодировок 4 аминокислот: Lys, Phe, Pro, Gly перекрываются с кодировками третьей позиции Thr) следует, что третья позиция в кодировках Thr равна N. Таким образом, имеем:

$$\text{Thr: acN.} \quad (17)$$

	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly
Met													Cys			Cys	Trp
Trp							Gly				Gly		Gly			Gly	Gly
3	Phe		Phe														
Tyr		Met	Ile	Ile	Thr			Met	Thr					Thr			
His		Met	Ile	Ile	Thr			Met	Thr					Thr			
Asn		Met	Ile	Ile	Thr			Met	Thr					Thr			
Asp		Met	Ile	Ile	Thr			Met	Thr					Thr			
Cys		Val	Val	Val	Ala			Val	Ala					Ala			
Gln	Asn					Lys				Lys		Asn			Asn		
3	Lys	Asn				Lys				Lys		Asn			Asn		
Glu	Asn					Lys				Lys		Asn			Asn		
Ile	Tyr		Phe									Tyr			Tyr		
Val	Tyr		Phe									Tyr	Cys		Tyr	Cys	Trp
3	Pro	His			Pro	Gln			Pro	Gln		His		Pro	His		
Thr	His				Pro	Gln			Pro	Gln		His		Pro	His		
Ala	His				Pro	Gln			Pro	Gln		His		Pro	His		
3	Gly	Asp	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly

	3			4				3			4			3			
	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly
3	Met												Cys			Cys	Trp
	Trp						Gly				Gly		Gly			Gly	Gly
3	Phe		Phe														
	Tyr	Met	Ile	Ile	Thr			Met	Thr					Thr			
	His	Met	Ile	Ile	Thr			Met	Thr					Thr			
	Asn	Met	Ile	Ile	Thr			Met	Thr					Thr			
	Asp	Met	Ile	Ile	Thr			Met	Thr					Thr			
	Cys	Val	Val	Val	Ala			Val	Ala					Ala			
4	Gln	Asn				Lys				Lys		Asn			Asn		
3	Lys	Asn				Lys				Lys		Asn			Asn		
4	Glu	Asn				Lys				Lys		Asn			Asn		
	Ile	Tyr									Tyr				Tyr		
4	Val	Tyr	Phe								Tyr		Cys		Tyr	Cys	Trp
3	Pro	His			Pro	Gln			Pro	Gln		His		Pro	His		
	Thr	His			Pro	Gln			Pro	Gln		His		Pro	His		
4	Ala	His			Pro	Gln			Pro	Gln		His		Pro	His		
3	Gly	Asp	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly

Puc. 8. IIIar 4

Met	5	Trp	3	Phe	3	Tyr	His	Asn	Asp	Cys	5	Gln	4	Lys	3	Glu	4	Ile	4	Val	4	Pro	3	Thr	4	Ala	4	Gly	3		
Met																															
Trp									Gly							Gly												Cys		Trp	
Phe				Phe																											Gly
Tyr		Met		Ile		Ile	Thr			Met		Thr											Thr								
His		Met		Ile		Ile	Thr			Met		Thr											Thr								
Asn		Met		Ile		Ile	Thr			Met		Thr											Thr								
Asp		Met		Ile		Ile	Thr			Met		Thr											Thr								
Cys		Val		Val		Val	Ala			Val		Ala											Ala								
Gln	Asn							Lys						Lys				Asn							Asn						
Lys	Asn							Lys						Lys				Asn							Asn						
Glu	Asn							Lys						Lys				Asn						Asn							
Ile	Tyr					Phe												Tyr						Tyr							
Val	Tyr			Phe														Tyr		Cys				Tyr	Cys					Trp	
Pro	His						Pro	Gln				Pro	Gln				His	Arg	Pro	Pro				His							
Thr	His						Pro	Gln				Pro	Gln				His	Arg	Pro	Pro				His							
Ala	His						Pro	Gln				Pro	Gln				His	Arg	Pro	Pro				His							
Gly	Asp	Val		Val		Val	Ala	Glu	Gly	Val		Ala	Glu	Glu			Asp	Gly	Ala	Gly			Ala	Asp	Gly					Gly	

Рис. 8. Шаг 5

	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly
5	Met												Cys			Cys	Trp
3	Trp						Gly				Gly		Gly			Gly	Gly
	Phe		Phe											Ser			
	Tyr	Met	Ile	Ile	Thr			Met	Thr					Thr			
	His	Met	Ile	Ile	Thr			Met	Thr					Thr			
	Asn	Met	Ile	Ile	Thr			Met	Thr					Thr			
	Asp	Met	Ile	Ile	Thr			Met	Thr					Thr			
5	Cys	Val	Val	Val	Ala			Val	Ala					Ala			
4	Gln	Asn				Lys				Lys		Asn			Asn		
3	Lys	Asn				Lys				Lys		Asn			Asn		
4	Glu	Asn				Lys				Lys		Asn			Asn		
	Ile	Tyr		Phe								Tyr			Tyr		
4	Val	Tyr		Phe								Tyr	Cys		Tyr	Cys	Trp
3	Pro	His			Pro	Gln			Pro	Gln		His		Pro	His		
6	Thr	His			Pro	Gln			Pro	Gln		His		Pro	His		
4	Ala	His			Pro	Gln			Pro	Gln		His		Pro	His		
3	Gly	Asp	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	ASP	Gly	Ala	ASP	Gly	Gly

Рис. 8. Шаг 6

	Met	Trp	Phe	Tyr	His	Asn	Asp	Cys	Gln	Lys	Glu	Ile	Val	Pro	Thr	Ala	Gly
5	Met												Cys			Cys	Trp
3	Trp						Gly				Gly		Gly			Gly	Gly
3	Phe		Phe														
7	Tyr	Met	Ile	Ile	Thr			Met	Thr					Thr			
7	His	Met	Ile	Ile	Thr			Met	Thr					Thr			
7	Asn	Met	Ile	Ile	Thr			Met	Thr					Thr			
7	Asp	Met	Ile	Ile	Thr			Met	Thr					Thr			
5	Cys	Val	Val	Val	Ala			Val	Ala					Ala			
4	Gln	Asn				Lys				Lys		Asn	Ser		Asn		
3	Lys	Asn				Lys				Lys		Asn	Ser		Asn		
4	Glu	Asn				Lys				Lys		Asn	Ser		Asn		
4	Ile	Tyr		Phe								Tyr			Tyr		
4	Val	Tyr		Phe								Tyr	Cys		Tyr	Cys	Trp
3	Pro	His			Pro	Gln			Pro	Gln		His		Pro	His		
6	Thr	His			Pro	Gln			Pro	Gln		His		Pro	His		
4	Ala	His			Pro	Gln			Pro	Gln		His		Pro	His		
3	Gly	Asp	Val	Val	Ala	Glu	Gly	Val	Ala	Glu	Gly	Asp	Gly	Ala	Asp	Gly	Gly

Шаг 6. На рис 8 (шаг 6) после введения полос 6, соответствующих Thr по вертикали и горизонтали были выделены 4 аминокислоты полужирным шрифтом: 3 раза Asn, один раз Tyr, 3 раза His и один раз Asp. Эти аминокислоты соответствуют перекрытиям:

$$\begin{array}{cccc}
 \text{Thr} & \text{Thr} & \text{Thr} & \text{Thr} \\
 \text{Tyr} & \text{Asn} & \text{His} & \text{Asp} \\
 \text{Val} & \text{Gln} & \text{Pro} & \text{Gly} \\
 & (\text{Lys}) & (\text{Thr}) & \\
 & (\text{Glu}) & (\text{Ala}) &
 \end{array} \quad (18)$$

на основе которых вычислим 4 кодировки для Tyr: bac, Asn: aac His: cac, Asp: dac. С учетом данных из столбца m_{3-1} окончательно находим:

$$\text{Tyr: baY, Asn: aaY, His: caY, Asp: daY} \quad (19)$$

Шаг 7. Для этих аминокислот проводим по 4 полосы по горизонтали и вертикали соответственно. При этом оказались выделенными по 8 позиций лишь 2 аминокислоты – Met и Ile, см. полужирный шрифт на рис 8 (шаг7).

Выделим два перекрытия,

$$\begin{array}{cc}
 \text{Trp} & \text{Phe} \\
 \text{Met} & \text{Ile} \\
 \text{Tyr} & \text{Tyr}
 \end{array} \quad (20)$$

которые оказались достаточными для расчета. С учетом данных из столбца m_{3-1} окончательно находим:

$$\text{Met: abd, Ile: abM}, \quad (21)$$

где M: a, b, c.

Шагом 7 завершается поиск кодировок для введенной области однозначности: их оказалось 43 – для первых 17-ти аминокислот; они даны в таблице 2. Дальнейший расчет сводился к поиску решений для Ser, Leu, Arg, присутствующих в сжатом множестве рис. 6.

6. Поиск решения в области неоднозначности

Шаг 8. Поиск решений в области, где значения A_{m_1} и A_{m_2} принадлежат не только этим 17 аминокислотам. На основе рис. 7 получим:

$$\begin{array}{cccc}
 \text{Gln} & \text{Pro} & \text{Ala} & \text{Val} \\
 \text{Ser} & \text{Ser} & \text{Ser} & \text{Ser} \\
 \text{Phe} & \text{Ile} & \text{Gln} & \text{Gln} \\
 \\
 \text{Tyr} & \text{Phe} & \text{Trp} & \text{Tyr} & \text{Trp} \\
 \text{Leu} & \text{Leu} & \text{Leu} & \text{Leu} & \text{Leu}
 \end{array}$$

Thr	Ala	Pro	Val	Val		(22)
	Glu	Val	Ala	Gly	Gly	Gly
	Arg	Arg	Arg	Arg	Arg	Arg
	Lys	Ala	Ala	Pro	Thr	Lys

Из этих перекрытий находим следующие кодировки:

$$\text{Ser: bca, bcc, adY; Leu: cbX, cbb, bbX; Arg: cdN, adX.} \quad (23)$$

Шаг 9. Из области неоднозначности на рис.2 выделим случаи, содержащие по две аминокислоты из набора Ser, Leu, Arg и дающие решения, отличные от (23). Имеем:

Arg	Leu	Ser	
Ser	Ser	Leu	(24)
Val	Ile	Pro	

Из первого и второго перекрытия находим: Ser: bcd, bcb, а из третьего – Leu: cbc. Окончательные кодировки для Ser, Leu, Arg представлены в Табл. 2. Полное число смысловых триплетов для всех 20 аминокислот из в этой таблице равно 61. Дополнительной проверкой установлено, что все элементы области неоднозначности не содержат никаких других решений. Три триплета: baX, bda – не определяются при использовании любых элементов множеств U_1, U_2, U_3 ; они дополняют полное число триплетов до 64.

При переходе от набора букв a, b, c, d к каноническим нуклеотидам A, C, T, G можно получить 24 подобных генетических кода. Всего один из них – стандартный, при a=A, b =T, c =C, d=G, а триплеты baX, bda становятся TAA, TAG, TGA; они играют роль терминаторных кодонов – кодонов, останавливающих белковый синтез.

Библиографический список

1. Козлов Н.Н., Энеев Т.М. Введение в математическую теорию генетического кода. ДАН, 2017, Т.473. № 3. С. 12-15.

N.N. Kozlov: Fundamentals of a Mathematical Theory of Genetic Code. Doklady Mathematics, 2017, Vol. 95, №2, pp. 1-3

2. Козлов Н.Н. Генетический код: взгляд математика. Монография. Palmarium Academic Publishing, 2014, 336 с. ISBN: 978-3-639-63268-2N.N. Kozlov. Genetic Code: A Mathematician's Point of View. Palamarium Academic, Hamburg, 2014, 336 p. ISBN: 978-3-639-63268-2

3. Козлов Н.Н. Метод расчета генетического кода. ДАН 2015.Т. 462. № 1. С. 15-18. Kozlov, N.N. Method of Computing a Genetic Code. *Doklady Mathematics* 2015, Volume 91, №3, pp.263-266

4. Козлов Н.Н. О перекрытиях более двух генов. Теорема для однородных перекрытий. ДАН 2015.Т. 462. № 4. С. 391-393. Kozlov, N.N. On Overlaps of More than Two Genes: A Theorem for Homogeneous Overlaps. *Doklady Mathematics* 2015, Volume 91, № 3, pp. 1-3

5. Козлов Н.Н. Теорема для генетического кода. ДАН 2002. Т. 382, № 5, С. 593-597,

Kozlov, N.N. A Theorem on the Genetic Code. *Doklady Mathematics*, 65, No. 1, 83-87 (2002).

6. Козлов Н.Н. Множества, порождаемые генетическим кодом. ДАН 2008. Т.423, № 3. С.295-299.

Kozlov, N.N. Sets generated by genetic code. *Doklady Mathematics* 2008, Vol.78, No.3. pp. 851-855.

7. Nakayama T., Asai S., Takahashi Y., et al., Overlapping of genes in the human. *Genome NJBS* 2007, vol.3, no. 1, p. 14-19.

8. Козлов Н.Н. Неоднозначности во множествах, порождаемых генетическим кодом. ДАН 2010. Т.432. № 2. С.157- 161.

Kozlov, N.N. Ambiguity in sets generated by the genetic code. *Doklady Mathematics* 2010, Volume 81, No.3 pp 364-367

9. Козлов Н.Н. Одна функция неоднозначностей из множеств, порождаемых генетическим кодом // Математическое моделирование, т. 24, № 6. С.57-66, 2012.

Kozlov, N.N. One function of ambiguities from the sets generated by the genetic code *Mathematical Models and Computer Simulations*, January 2013, Volume 5, Issue 1, pp 17-24.

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

Kozlov N.N. Three Functions of Ambiguities Generated by the Genetic Code. *Mathematical Models and Computer Simulations*, 2015, Volume 7, Issue 5, pp 401-408.

11. Козлов Н.Н. Расчет генетического кода // ДАН, 2010. Т.433. №.2. С. 158-162.

Kozlov, N.N. Computation of the genetic code. *Doklady Mathematics* 2010, Volume 82, No.3 pp 535-539

12. Козлов Н.Н. О расчете генетического кода // Математическое моделирование, т. 23, № 6, с. 3-17, 2011.

Kozlov, N.N. Computation of the genetic code *Mathematical Models and Computer Simulations*, February 2012, Volume 4, Issue 1, pp 36-46

Приложение 1

Ниже приводятся все элементы множества U_2 .

1	Met	Met	Met	Met
Met	Tyr	His	Asn	Asp
(4+2)	1	2	3	4
	Met	Met		
	Trp	Cys		
	5	6		
2	Trp	Trp	Trp	
Trp	Met	Val	Leu	
(3+1)	7	8	9	
	Trp			
	Gly			
	10			
3	Phe	Phe	Phe	Phe
Phe	Phe	Ile	Val	Leu
(4+3)	11	12	13	14
	Phe	Phe	Phe	
	Phe	Ser	Leu	
	15	16	17	
4	Tyr	Tyr	Tyr	
Tyr	Ile	Val	Leu	
(3+3)	18	19	20	
	Tyr	Tyr	Tyr	
	Met	Ile	Thr	
	21	22	23	
5	His	His	His	His
His	Pro	Thr	Ala	Ser
(4+3)	24	25	26	27
	His	His	His	
	Met	Ile	Thr	
	28	29	30	

6	Asn	Asn	Asn	
Asn	Gln	Lys	Glu	
(3+3)	31	32	33	
	Asn	Asn	Asn	
	Met	Ile	Thr	
	34	35	36	
7	Asp	Asp		
Asp	Gly	Arg		
(2+3)	37	38		
	Asp	Asp	Asp	
	Met	Ile	Thr	
	39	40	41	
8	Cys	Cys	Cys	
Cys	Met	Val	Leu	
(3+2)	42	43	44	
	Cys	Cys		
	Val	Ala		
	45	46		
9	Gln	Gln	Gln	Gln
Gln	Pro	Thr	Ala	Ser
(4+4)	47	48	49	50
	Gln	Gln	Gln	Gln
	Asn	Lys	Ser	Arg
	51	52	53	54
10	Lys	Lys	Lys	
Lys	Gln	Lys	Glu	
(3+4)	55	56	57	
	Lys	Lys	Lys	Lys
	Asn	Lys	Ser	Arg
	58	59	60	61
11	Glu	Glu		
Glu	Gly	Arg		
(2+4)	62	63		

	Glu	Glu	Glu	Glu		
	Asn	Lys	Ser	Arg		
	64	65	66	67		
12	Ile	Ile	Ile	Ile		
Ile	Tyr	His	Asn	Asp		
(4+4)	68	69	70	71		
	Ile	Ile	Ile	Ile		
	Phe	Tyr	Ser	Leu		
	72	73	74	75		
13	Val	Val	Val	Val		
Val	Cys	Gly	Ser	Arg		
(4+6)	76	77	78	79		
	Val	Val	Val	Val	Val	Val
	Trp	Phe	Tyr	Cys	Ser	Leu
	80	81	82	83	84	85
14	Pro	Pro	Pro	Pro		
Pro	Pro	Thr	Ala	Ser		
(4+5)	86	87	88	89		
	Pro	Pro	Pro	Pro	Pro	
	His	Gln	Pro	Leu	Arg	
	90	91	92	93	94	
15	Thr	Thr	Thr	Thr		
Thr	Tyr	His	Asn	Asp		
(4+5)	95	96	97	98		
	Thr	Thr	Thr	Thr	Thr	
	His	Gln	Pro	Leu	Arg	
	99	100	101	102	103	
16	Ala	Ala	Ala	Ala		
Ala	Cys	Gly	Ser	Arg		
(4+5)	104	105	106	107		
	Ala	Ala	Ala	Ala	Ala	
	His	Gln	Pro	Leu	Arg	
	108	109	110	111	112	
17	Gly	Gly	Gly			
Gly	Trp	Gly	Arg			

(3+5)	113	114	115					
	Gly	Gly	Gly	Gly	Gly			
	Asp	Glu	Val	Ala	Gly			
	116	117	118	119	120			
18	Ser	Ser	Ser	Ser	Ser	Ser	Ser	
Ser	Phe	Gln	Lys	Glu	Ile	Val	Leu	
(7+7)	121	122	123	124	125	126	127	
	Ser	Ser	Ser	Ser	Ser	Ser	Ser	
	His	Gln	Val	Pro	Ala	Leu	Arg	
	128	129	130	131	132	133	134	
19	Leu	Leu	Leu	Leu	Leu	Leu	Leu	
Leu	Phe	Ile	Val	Pro	Thr	Ala	Ser	I
(8+6)	135	136	137	138	139	140	141	
	Leu	Leu	Leu	Leu	Leu	Leu		
	Trp	Phe	Tyr	Cys	Ser	Leu		
	143	144	145	146	147	148		
20	Arg	Arg	Arg	Arg	Arg	Arg	Arg	
Arg	Gln	Lys	Glu	Pro	Thr	Ala	Ser	
(7+5)	149	150	151	152	153	154	155	
	Arg	Arg	Arg	Arg	Arg			
	Asp	Glu	Val	Ala	Gly			
	156	157	158	159	160			

Приложение 2

Ниже приводятся все 307 элементов множества U_3 . Элементы приводятся в том представлении, которое используется в данной задаче. Каждый из элементов состоит из трех строк: верхней, средней и нижней.

Trp	Trp	Trp	Trp	Cys	Cys	Cys	Cys
Met	Met	Met	Met	Met	Met	Met	Met
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
1	2	3	4	5	6	7	8

Gly	Gly	Gly
Trp	Trp	Trp
Met	Val	Leu
9	10	11

Phe	Phe	Phe	Phe	Ser	Ser
Phe	Phe	Phe	Phe	Phe	Phe
Phe	Ile	Val	Leu	Phe	Ile
12	13	14	15	16	17

Ser	Ser	Leu	Leu	Leu	Leu
Phe	Phe	Phe	Phe	Phe	Phe
Val	Leu	Phe	Ile	Val	Leu
18	19	20	21	22	23

Met	Met	Met	Ile	Ile	Ile	Thr	Thr	Thr
Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr
Ile	Val	Leu	Ile	Val	Leu	Ile	Val	Leu
24	25	26	27	28	29	30	31	32

Met	Met	Met	Met	Ile	Ile
His	His	His	His	His	His
Pro	Thr	Ala	Ser	Pro	Thr
33	34	35	36	37	38

Ile	Ile	Thr	Thr	Thr	Thr
His	His	His	His	His	His
Ala	Ser	Pro	Thr	Ala	Ser
39	40	41	42	43	44

Met	Met	Met	Ile	Ile	Ile
-----	-----	-----	-----	-----	-----

Asn	Asn	Asn	Asn	Asn	Asn
Gln	Lys	Glu	Gln	Lys	Glu
45	46	47	48	49	50

Thr	Thr	Thr
Asn	Asn	Asn
Gln	Lys	Glu
51	52	53

Met	Met	Ile	Ile	Thr	Thr
Asp	Asp	Asp	Asp	Asp	Asp
Gly	Arg	Gly	Arg	Gly	Arg
54	55	56	57	58	59

Val	Val	Val	Ala	Ala	Ala
Cys	Cys	Cys	Cys	Cys	Cys
Met	Val	Leu	Met	Val	Leu
60	61	62	63	64	65

Asn	Asn	Asn	Asn	Lys	Lys	Lys	Lys
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
66	67	68	69	70	71	72	73

Ser	Ser	Ser	Ser	Arg	Arg	Arg	Arg
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
74	75	76	77	78	79	80	81

Asn	Asn	Asn	Lys	Lys	Lys
Lys	Lys	Lys	Lys	Lys	Lys
Gln	Lys	Glu	Gln	Lys	Glu
82	83	84	85	86	87

Ser	Ser	Ser	Arg	Arg	Arg
Lys	Lys	Lys	Lys	Lys	Lys
Gln	Lys	Glu	Gln	Lys	Glu
88	89	90	91	92	93

Asn	Asn	Lys	Lys	Ser	Ser	Arg	Arg
Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu
Gly	Arg	Gly	Arg	Gly	Arg	Gly	Arg
94	95	96	97	98	99	100	101

Phe	Phe	Phe	Phe	Tyr	Tyr	Tyr	Tyr
Ile	Ile	Ile	Ile	Ile	Ile	Ile	Ile
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
102	103	104	105	106	107	108	109

Ser	Ser	Ser	Ser	Leu	Leu	Leu	Leu
Ile	Ile	Ile	Ile	Ile	Ile	Ile	Ile
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
110	111	112	113	114	115	116	117

Trp	Trp	Trp	Trp	Phe	Phe	Phe	Phe
Val	Val	Val	Val	Val	Val	Val	Val
Cys	Gly	Ser	Arg	Cys	Gly	Ser	Arg
118	119	120	121	122	123	124	125

Tyr	Tyr	Tyr	Tyr	Cys	Cys	Cys	Cys
Val	Val	Val	Val	Val	Val	Val	Val
Cys	Gly	Ser	Arg	Cys	Gly	Ser	Arg
126	127	128	129	130	131	132	133

Ser	Ser	Ser	Ser	Leu	Leu	Leu	Leu
Val	Val	Val	Val	Val	Val	Val	Val
Cys	Gly	Ser	Arg	Cys	Gly	Ser	Arg
134	135	136	137	138	139	140	141

His	His	His	His	Gln	Gln	Gln	Gln
Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
142	143	144	145	146	147	148	149

Pro	Pro	Pro	Pro	Leu	Leu	Leu	Leu
Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
150	151	152	153	154	155	156	157

Arg	Arg	Arg	Arg
Pro	Pro	Pro	Pro
Pro	Thr	Ala	Ser
158	159	160	161

His	His	His	His	Gln	Gln	Gln	Gln
Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
162	163	164	165	166	167	168	169

Pro	Pro	Pro	Pro	Leu	Leu	Leu	Leu
Thr	Thr	Thr	Thr	Thr	Thr	Thr	Thr
Tyr	His	Asn	Asp	Tyr	His	Asn	Asp
170	171	172	173	174	175	176	177

Arg	Arg	Arg	Arg
Thr	Thr	Thr	Thr
Tyr	His	Asn	Asp
178	179	180	181

His	His	His	His	Gln	Gln	Gln	Gln
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
Cys	Gly	Ser	Arg	Cys	Gly	Ser	Arg
182	183	184	185	186	187	188	189

Pro	Pro	Pro	Pro	Leu	Leu	Leu	Leu
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
Cys	Gly	Ser	Arg	Cys	Gly	Ser	Arg
190	191	192	193	194	195	196	197

Arg	Arg	Arg	Arg
Ala	Ala	Ala	Ala
Cys	Gly	Ser	Arg
198	199	200	201

Asp	Asp	Asp	Glu	Glu	Glu	Val	Val	Val
Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly
Trp	Gly	Arg	Trp	Gly	Arg	Trp	Gly	Arg
202	203	204	205	206	207	208	209	210

Ala	Ala	Ala	Gly	Gly	Gly
Gly	Gly	Gly	Gly	Gly	Gly
Trp	Gly	Arg	Trp	Gly	Arg
211	212	213	214	215	216

His	His	His	His	Gln	Gln	Gln	Gln
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser
Phe	Ile	Val	Leu	Phe	Ile	Val	Leu
217	218	219	220	221	222	223	224

Pro	Pro	Pro	Pro	Leu	Leu	Leu	Leu
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser
Phe	Ile	Val	Leu	Phe	Ile	Val	Leu

225	226	227	228	229	230	231	232
Arg	Arg	Arg	Arg	Val	Val	Val	Ala
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser
Phe	Ile	Val	Leu	Gln	Lys	Glu	Gln
233	234	235	236	237	238	239	240
Ala	Ala						
Ser	Ser						
Lys	Glu						
241	242						
Trp	Trp	Trp	Trp	Phe	Phe	Phe	Phe
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
243	244	245	246	247	248	249	250
Tyr	Tyr	Tyr	Tyr	Cys	Cys	Cys	Cys
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
251	252	253	254	255	256	257	258
Ser	Ser	Ser	Ser	Leu	Leu	Leu	Leu
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
259	260	261	262	263	264	265	266
Trp	Trp	Trp	Trp	Tyr	Tyr	Tyr	Tyr
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu
Phe	Ile	Val	Leu	Phe	Ile	Val	Leu
267	268	269	270	271	272	273	274
Cys	Cys	Cys	Cys				
Leu	Leu	Leu	Leu				
Phe	Ile	Val	Leu				
275	276	277	278				
Asp	Asp	Asp	Asp	Glu	Glu	Glu	Glu
Arg	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
279	280	281	282	283	284	285	286
Val	Val	Val	Val	Ala	Ala	Ala	Ala

Arg	Arg	Arg	Arg	Arg	Arg	Arg	Arg
Pro	Thr	Ala	Ser	Pro	Thr	Ala	Ser
287	288	289	290	291	292	293	294

Gly	Gly	Gly	Gly	Asp	Asp	Asp
Arg	Arg	Arg	Arg	Arg	Arg	Arg
Pro	Thr	Ala	Ser	Gln	Lys	Glu
295	296	297	298	299	300	301

Glu	Glu	Glu	Gly	Gly	Gly
Arg	Arg	Arg	Arg	Arg	Arg
Gln	Lys	Glu	Gln	Lys	Glu
302	303	304	305	306	307

Файл оцЦюрих

IRED Conference Review Result



Conference Name: "FIFTH INTERNATIONAL CONFERENCE ON ADVANCES IN BIO-INFORMATICS,
BIO-TECHNOLOGY AND ENVIRONMENTAL ENGINEERING - ABBE 2017

Paper Title: "COMPUTATION OF THE GENETIC CODE"

Paper ID: "ABBE-17-946"

Conference Date: 02-03 September, 2017

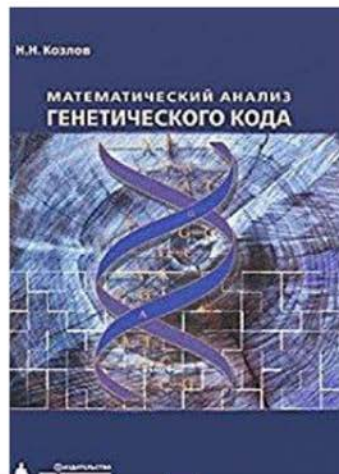
Official Conference Website: www.abbe.theired.org

Evaluation:					
	Poor	Fair	Good	Very Good	Outstanding
Originality	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Innovation	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
technical merit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
applicability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Presentation and English	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Match to Conference Topic	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recommendation to Editors					
	Strongly Reject	Reject	Marginally Accept	Accept	Strong Accept
Recommendation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Файл Amazon

18.06.18

Mathematical analysis of the genetic code / Matematicheskiy analiz geneticheskogo koda



Артикул: [5996301199](#) (посмотреть на [Amazon](#))
Лучшее предложение: **\$81.49**

Все предложения: [Новых 1 за \\$81.49](#) [Б/У 1 за \\$66.45](#)



Математический анализ генетического кода
Н. Н. Козлов

275 ₺ Москва

275: 60= 4.6 \$ 81.5:4.6 =17.

Оглавление

Введение	3
1. Теорема для однородных перекрытий	3
2. Введение основных множеств	8
3. Постановка задачи. Введение сжатого множества	14
4. Начальное приближение	16
5. Пошаговый расчет в области однозначности	18
6. Поиск решения в области неоднозначности	25
Библиографический список.....	26
Приложение 1	28
Приложение 2	32
Файл оцЦюрих.....	38
Файл Amazon	39