

Эволюция кодирования аминокислот иРНК.

Похоже, что способ трансляции белка по информации нуклеотидов иРНК не всегда шел в режиме триплетного кодирования. Когда-то считывание шло шагами по 1-му основанию. Это и ограничило количество применяемых аминокислот до 20-ти из 64-х возможных.

С другой стороны, мы видим, что формирование системы кодонов иРНК вначале шло на основе симметрии оснований относительно среднего члена триплета. Главный элемент триплета – среднее основание, а потом правое и левое. Это подтверждается фактом применения симметричных кодов в 15-ти кодонах неповторяющихся аминокислот из 16 возможных. Очевидно, что оставшиеся кодоны формировались по такому же принципу, хоть и без использования симметрии. Когда-то эти ветви эволюции кодирования иРНК встретились. И получилось то, что есть.

Существующий сегодня способ формирования таблицы кодонов по последовательности от 1-го к 3-му основаниям привел к тому, что кодоны некоторых аминокислот, например Arg, оказываются в нескольких местах таблицы, запутывая общую картину. Система кодирования с учетом эволюционных изменений меняет понимание группировки кодонов и не создает повторов в таблице.

Данная статья написана по материалам статьи [1].

Когда-то меня заинтересовал вопрос триплетного кодирования аминокислот.

Как это происходит?

Я начал свои поиски. Начал я с «классики»..., с таблицы кодирующих триплетов.

Вот эта таблица очень близка к «классической»:

Таблица 1.

Стандартный генетический код									
1-е основание	2-е основание								3-е основание
	U		C		A		G		
U	UUU	(Phe/F)	UCU	(Ser/S)	UAU	(Tyr/Y)	UGU	(Cys/C)	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L)	UCA		UAA	Стон	UGA	Стон	A
	UUG		UCG		UAG	Стон	UGG	(Trp/W)	G
C	CUU	(Leu/L)	CCU	(Pro/P)	CAU	(His/H)	CGU	(Arg/R)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	CGA	A		
	CUG		CCG		CAG	(Gln/Q)	CGG		G
A	AUU	(Ile/I)	ACU	(Thr/T)	AAU	(Asn/N)	AGU	(Ser/S)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K)	AGA	(Arg/R)	A
	AUG ^[A]	(Met/M)	ACG		AAG	AGG	G		
G	GUU	(Val/V)	GCU	(Ala/A)	GAU	(Asp/D)	GGU	(Gly/G)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E)	GGA		A
	GUG		GCG		GAG		GGG		G

Видите? Первое основание, второе,... третье... Из них получаются триплеты, кодирующие ту или иную аминокислоту при синтезе белка рибосомой.

Но, непонятно, почему одна и та же аминокислота вдруг оказывается в разных частях таблицы с разными триплетными кодами... Например, аргинин, серин...

Как же так получилось, что триплеты разных групп таблицы кодируют одну и ту же аминокислоту? Может быть случайность так распорядилась?

Может быть...

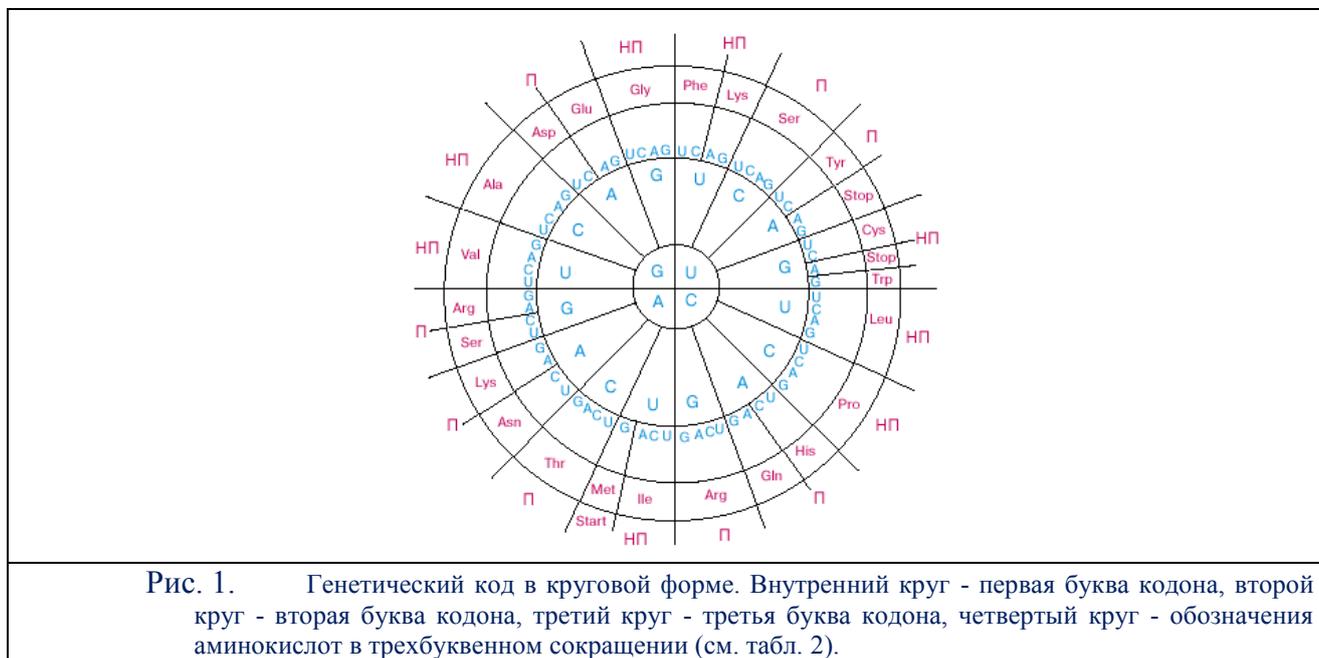
Есть разные отображения формирования триплетов для синтеза белка.

Например, так:

Таблица 2.

A	Ala	Аланин	GCA GCC GCG GCU
C	Cys	Цистеин	UGC UGU
D	Asp	Аспарагиновая кислота	GAC GAU
E	Glu	Глутаминовая кислота	GAA GAG
F	Phe	Фенилаланин	UUC UUU
G	Gly	Глицин	GGA GGC GGG GGU
H	His	Гистидин	CAC CAU
I	Ile	Изолейцин	AUA AUC AUU
K	Lys	Лизин	AAA AAG
L	Leu	Лейцин	UUA UUG CUA CUC CUG CUU
M	Met	Метионин	AUG
N	Asn	Аспарагин	AAC AAU
P	Pro	Пролин	CCA CCC CCG CCU
Q	Gln	Глутамин	CAA CAG
R	Arg	Аргинин	AGA AGG CGA CGC CGG CGU
S	Ser	Серин	AGC AGU UCA UCC UCG UCU
T	Thr	Треонин	ACA ACC ACG ACU
V	Val	Валин	GUA GUC GUG GUU
W	Trp	Триптофан	UGG
Y	Tyr	Тирозин	UAC UAU

Возможно, для кого-то будет интересным и круговое изображение этой же таблицы. Оно на [рис.1](#)



А теперь цитата [6]:

Как известно, оснований, которыми различаются нуклеотиды, всего четыре. В РНК это аденин (А), гуанин (G), цитозин (С) и урацил (U) (Т-тимин в ДНК), а обычных аминокислот, входящих в белки, - 20 (рис. 1). Следовательно, задача сводится к тому, чтобы четырьмя основаниями записать двадцать аминокислот. И отсюда следует, что код должен быть не менее чем триплетным, поскольку по одному основанию и даже по два ($4 \times 4 = 16$) недостаточно, а по три даже много ($4 \times 4 \times 4 = 64$). Сколько же кодонов из 64 имеют смысл, а какие бессмысленны? Соответствует ли каждой аминокислоте один или несколько кодонов? Ответы на эти вопросы были получены к 1965 году, когда генетический код был полностью расшифрован [см. 6]. Удобнее всего представить код в круговой форме (рис. 1). Буквы в центре круга - первые буквы кодонов, вокруг расположены буквы, соответствующие второму положению в кодоне, и, наконец, третий круг - третье положение в кодоне. Четвертое кольцо образуют аминокислотные остатки, представленные в виде трехбуквенных сокращений. Во внешнем круге отмечены физико-химические свойства аминокислот, а именно являются ли они полярными (п) или неполярными (нп). Сразу видно, что каждой аминокислоте соответствует от одного (Met, Trp) до шести (Leu, Arg, Ser) кодонов, то есть код обладает свойством избыточности, или вырожденности (табл. 1).

Кодон для метионина одновременно служит инициатором - сигналом начала синтеза полипептида. Кодонов, не кодирующих аминокислот, оказалось всего три: UAA, UAG, UGA. Поначалу их называли бессмысленными кодонами или нонсенсами (это название сохранилось в научном обиходе до сих пор), однако вскоре выяснилось, что они вовсе не бессмысленны, а представляют собой сигналы терминации синтеза белка. Действительно, в дальнейшем, когда начали расшифровывать нуклеотидные последовательности генов, убедились, что первый же встреченный на иРНК кодон AUG (Met) задает фазу последующего считывания троек, то есть служит той самой фиксированной точкой, с которой начинается считывание. Любой последующий AUG просто кодирует Met. В конце гена обязательно стоит UAA, или UAG, или UGA, а то и два нонсенса подряд.

У меня, как у дилетанта сразу возникли вопросы, которые наверно у ученых-биологов не возникают. Как клетка смогла определить, какой элемент кодона является первым, а какой третьим? На какой половинке ДНК формируется иРНК с информацией для трансляции белка?

Ну и так далее. Вопросов оказалось много.

Оказалось, что... иРНК образуется на... любой половинке ДНК, а потом ведется поиск «старт-кодонов» и «стоп-кодонов». По их наличию определяется возможность начала процессинга иРНК или подготовки «зрелой» РНК, годной для трансляции белка...

Но, совершенно точно, считать клетка не умеет.

И потому, ей всё равно, где первый элемент триплета и в какую сторону идет движение рибосомы. Тогда у меня возник новый вопрос: Если кодонов 64, то почему они кодируют только 20 аминокислот?

Аминокислот в реальности больше чем двадцать, но даже для 21-ой и 22-ой места в таблице триплетов уже не находится. Почему? С другой стороны некоторые аминокислоты имеют до 6 разных триплетов кодирования. Не многовато ли?

Говорят, что так получилось от того, что одна аминокислота применяется чаще, а другая реже. Хотя, мы же знаем, при неперекрываемом кодировании для кодирования любой аминокислоты достаточно одного кодового триплета. Повторяй один кодон сколько угодно раз, и получай столько молекул нужной аминокислоты в белке.

Легко, просто, понятно. И энергозатраты минимальны.

Если же кодирование аминокислот в последовательности оснований РНК сразу возникло как триплетное, то даже полная случайность и хаотичность начального возникновения кодонов не должны были дать конечное количества именно в 20 аминокислот. Оно должно быть ближе к пределу 64 аминокислот, из трехсот-то возможных!

Читаем здесь [9]:

1. Транспортные РНК (тРНК) состоят примерно из 70 нуклеотидов. Каждая тРНК имеет акцепторный конец, к которому присоединяется аминокислотный остаток, и адаптерный конец, несущий тройку нуклеотидов, комплементарную какому-либо кодону иРНК (см. рис. 2), потому этот триплет назвали антикодоном. Первый и второй нуклеотиды кодона строго следуют правилам комплементарности (А - U; G - C) при взаимодействии с соответствующими нуклеотидами антикодона, а вот взаимодействие с третьим нуклеотидом кодона позволяет себе некоторую нестрогость, неоднозначность спаривания. Благодаря этой неоднозначности каждое семейство кодонов для одной аминокислоты, различающихся по третьему нуклеотиду, может "обслуживаться" одним антикодоном. С учетом этих правил для считывания всей кодовой таблицы

достаточно всего 31 тРНК. Тем не менее все не так просто, и уже у бактерий есть 45 разных тРНК. Их кодируют 78 генов. У дрожжей этих генов уже 400, у мушки дрозофилы - около 750, а у лягушки - уже примерно 8000, то есть получается, что одну молекулу тРНК могут кодировать несколько одинаковых или очень близких по структуре генов, и чем "дальше" в эволюции, тем больше таких генов для кодирования одинаковых тРНК.

Вот видите? Технические возможности этого способа кодирования дают сразу 31 тРНК или даже 45 тРНК для осуществления триплетного синтеза белка. Это ощутимо больше применяемых 20 аминокислот.

Тогда, можно предположить, что ... сначала все было не совсем так, чем сейчас.

Видимо, в начале... «рамка считывания» рибосомы каждый раз сдвигалась **только на один знак, а считывалось всё время по три знака, как триплет.**

Вот пример:

Есть какая-то последовательность нуклеотидов, например: АСГТАГТСААТС...

И... смотрим таблицу 3.

Следите за цветом букв в результате считывания. Например, смотрим на голубой в первых трех считываниях. Видите, как он меняет позицию при переходах?

Только каждый четвертый - новый триплет. На 12 полученных триплетов совсем новых - 4. Это 1-ый, 4-ый, 7-ой, 10-ый,...

Видимо такой способ движения рамки по цепи РНК когда-то и определил количество примененных аминокислот в современном рибосомном способе производства белка - 20 аминокислот.

Как это происходило?

Как так получается, что триплетов по 3 из 4 возможно 64, а аминокислот 20?

Вот например, если все знаки триплета полностью изменяются только за три сдвига «рамки считывания», то полностью независимых вариантов в первых 64 триплетах при таком сдвиге рамки будет только:

$$64:3 \approx 21;$$

1)

Это, как раз, количество применяемых аминокислот и команда «Стоп». Остальные 42 варианта триплетов становятся переходными между этими независимыми [3].

И похоже, мы неверно понимаем явление вырожденности кодонов. Это не расширение возможностей системы в кодировании информации, а «ошибки её прошлого». Это отголосок той, исходной системы кодирования. Как и вырожденность кода триплета – вынужденная мера, связанная с первоначальным способом считывания кода. Так уж получилось в ходе эволюции. Вырожденность кода аминокислоты в триплете возникла от исходной перекрываемости кода [3], хоть наука и говорит обратное.

Конечно, когда-то рибосома перешла на способ формирования прямого триплетного сдвига «рамки считывания». И вроде теперь-то уже можно использовать все 64 варианта триплетов, отдельных для каждой аминокислоты?

Но - нет. Количество аминокислот уже не увеличилось.

Потому, что полного слома той первичной, уже почти забытой системы сдвига рамки считывания не произошло, а значит осталась и та система формирования количества используемых аминокислот. Сегодня мы так и имеем 20 аминокислот и 64 кодона. На каждую аминокислоту приходится от 1 до 6 различных вариантов кодонов.

Таблица 3.

№	Положение рамки считывания	Считывается
1	<u>АС</u> ГТАГТСААТС...	АСГ
2	АС <u>ГТ</u> АГТСААТС...	СГТ
3	АСГ <u>ТА</u> ГТСААТС...	ГТА
4	АСГТА <u>ГТ</u> СААТС...	ТАГ
5	АСГТАГ <u>Т</u> СААТС...	АГТ
6	АСГТАГТ <u>С</u> ААТС...	ГТС
7	АСГТАГТ <u>СА</u> АТС...	ТСА
8	АСГТАГТ <u>САА</u> ТС...	САА
9	АСГТАГТ <u>СААТ</u> С...	ААТ
10	АСГТАГТ <u>СААТ</u> С...	АТС
11	АСГТАГТ <u>СААТ</u> С...	ТС..

Только постепенный переход от одностороннего движения «рамки считывания» к триплетному может дать имеющийся фактический результат. Все осложнения и дополнения этого процесса получения информации для синтеза белка по шаблону мРНК были уже чуть позже.

Но...

Это лишь вынужденный механизм формирования триплетов до перехода на прямое триплетное считывание. Он не объясняет способа отбора применяемых триплетов для кодирования аминокислот в процессе трансляции белка. Возможно, что эти 20 триплетных кодов и «стоп-кодон» формировались по какому-то доступному для клеток способу, не связанному со счетом оснований триплета?

Например так:

Первые триплеты формировались исходя из их симметричности!

Применим принцип симметричности в поиске нужных сочетаний и проверим, насколько мы правильно поняли путь природного кодирования аминокислот в ДНК.

Для этого соберем все варианты симметричных кодов в таблицу 4.

Отличный результат..., 15 из 16 возможных аминокислот получили симметричные коды [2].

Но, осталось еще 5 аминокислот и СТОП.

Видимо Природа шла тем же путем, ... и споткнулась на том же месте.

Все симметричные варианты использованы, запаса для расширения системы нет, а кодов не хватает. Далее использовались несимметричные варианты, но не все и не сразу.

Мы запомним этот очень показательный пример использования симметрий в кодировании триплетов. И применим основу его формирования к полной таблице триплетов.

На круговой диаграмме (рис.1.) обратим внимание, например, на аминокислоты Ser, Leu или Arg. Их группы кодирования расположены в разных сегментах диаграммы и, на первый взгляд, никак между собой не связаны.

Тогда воспользуемся таблицей 2 и выпишем нужные данные для анализа.

Мы увидим:

Аминокислота Arg кодируется как - AGA, GGA, CGA, CGC, CGG, CGU.

Кодоны для аминокислоты Leu - UUA, UUG, CUA, CUC, CUG, CUU.

В обоих случаях все коды крутятся вокруг средних нуклеотидов триплета.

Для Arg это нуклеотид G, для Leu это нуклеотид U. Они становятся и центрами симметрии триплетов. Вокруг этих центров группируются все коды триплетов этих аминокислот. Это основной вариант формирования кодов триплетов. Так формируется основной объем кодирования.

Есть другой вариант:

Для аминокислоты SER – AGG, AGU, UCG, UCU, UCC, UCA.

Здесь мы видим два центра триплета - G и C.

И это уникальный вариант кодирования...

Таблица 4.

№	Amino acid	RNA triplet
1	Phe	UUU
2	Pro	CCC
3	Lys	AAA
4	Gly	GGG
5	Pe	AUA
6	Arg	AGA
7	Thr	ACA
8	Tyr	UAU
9	Cys	UGU
10	Ser	UCU
11	Glu	GAG
12	Val	GUG
13	Ala	GCG
14	His	CAC
15	Leu	CUC
16	Gin	CAA
17	Asn	AAC
18	Tip	UGG
19	Stop	UAA
20	Asp	GAC
21	Start - Met	AUG
	Stop	UAG
	Stop	UGA

Но общее очевидно: Главным элементом для формирования кодонов стал... средний элемент триплета. Это и центр возможной симметрии триплета. Так рибосома определяет «начало» триплета. Потом идут «фланги», правый и левый.

Попробуем в основу кодирования поставить центр триплета и составим полную таблицу триплетов для всех аминокислот на новых принципах.

Данные занесем... в таблицу 5.

Таблица 5.

№	Центр триплета		Аминокислота	1	2	3	4
1	A	Lys	Лизин	AAA		AAG	
2	A	Asn	Аспарагин		AAC		AAU
3	A	Gln	Глутамин	CAA		CAG	
4	A	His	Гистидин		CAC		CAU
5	A	Glu	Глутаминовая кислота	GAA		GAG	
6	A	Asp	Аспарагиновая кислота		GAC		GAU
7	A	Stop	терминация	UAA		UAG	
8	A	Tyr	Тирозин		UAC		UAU
9	C	Thr	Треонин	ACA	ACC	ACG	ACU
10	C	Pro	Пролин	CCA	CCC	CCG	CCU
11	C	Ala	Аланин	GCA	GCC	GCG	CCU
12	GC	Ser	Серин	UCA	UCC	UCG	UCU
					AGC		AGU
13	G	Arg	Аргинин	AGA		AGG	
				CGA	CGC	CGG	CGU
14	G	Gly	Глицин	GGA	GGC	GGG	GGU
15	G	Stop	терминация	UGA			
16	G	Cys	Цистеин		UGC		UGU
17	G	Trp	Триптофан			UGG	
18	U	Ile	Изолейцин	AUA	AUC		AUU
19	U	Met	Метионин (Start)			AUG	
20	U	Leu	Лейцин	CUA	CUC	CUG	CUU
				UUA		UUG	
21	U	Val	Валин	GUA	GUC	GUG	GUU
22	U	Phe	Фенилаланин		UUC		UUU

Выделим красным симметрии. Как триплеты, так и аминокислоты, чтобы понять, насколько равномерно распределены симметрии по всей таблице...

Заметим, что у Arg использовано две симметрии CGC и AGA.

Но для всех аминокислот в других группах симметрий не хватило. В основном, для аминокислот с малым количеством кодонов. Похоже, что они включались в работу в последнюю очередь...

Ну это так, собственные домысливания...

Отметим, что использование двух центров симметрии для одной аминокислоты в кодировании, это явление уникальное. Оно применено только для серина. И больше нигде.

Все остальные аминокислоты имеют только один центр кодирования. При этом на одну аминокислоту приходится только один симметричный код. В пятнадцати случаях.

В остальных триплетах такой симметрии нет. И тогда в ход идут другие варианты. Уже не симметричные.

Смотрим, например, лейцин: *Левое основание может быть С или U. Для левого С правое основание может быть А, С, G, U, а для левого U правое основание может быть А или G. Но в центре всегда U!*

В абсолютном большинстве триплетов одной аминокислоты постоянным остается только центральное основание, а правое и левое ... изменяются.

Количество групп кодонов теперь снижено до 4 возможных. По букве центра кодона. Есть триплет, есть его центр. Есть правое и левое основание и их взаимные изменения.

Становится очевидна причина такого расположения кодонов и их внутренняя системность. Теперь каждая аминокислота четко фиксируется в своей строке таблицы.

декабрь 2021г.

г.Волгодонск

Литература:

1. Никитин А.В., О кодировании аминокислот в иРНК // «Академия Тринитаризма», М., Эл № 77-6567, публ.24693, 09.08.2018 <http://www.trinitas.ru/rus/doc/0016/001f/00163760.htm>
2. Никитин А.В., Эволюционный путь саморазвития искусственного интеллекта // «Академия Тринитаризма», М., Эл № 77-6567, публ.14738, 19.03.2008 <http://trinitas.ru/rus/doc/0016/001c/00161450.htm>
3. Никитин А.В., Проблемы понимания системы кодирования ДНК // «Академия Тринитаризма», М., Эл № 77-6567, публ.16181, 27.11.2010 <http://www.trinitas.ru/rus/doc/0016/001c/00161731.htm>
4. С. Г. ИНГЕ-ВЕЧТОМОВ Трансляция как способ существования живых систем, или в чем смысл "бессмысленных" кодонов <http://nature.web.ru/db/msg.html?mid=1157633&s>
5. С. Г. ИНГЕ-ВЕЧТОМОВ. Трансляция как способ существования живых систем, или в чем смысл "бессмысленных" кодонов. Продолжение. <http://nature.web.ru/db/msg.html?mid=1157633&uri=1.html>