

ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА И ВОЗМОЖНОСТИ ИХ РЕКОДИРОВАНИЯ

Е.П. Харченко

Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН, Санкт-Петербург, Россия

Резюме. Мир уже вступил в стадию возрастания вероятности возникновения новой пандемии, что побуждает к поиску новых вакцин против гриппа, поскольку эффективность существующих к нему вакцин лишь субоптимальная. С пандемией COVID-19 открылась возможность использования мРНК-вакцин, и перспектива поиска новых противогриппозных вакцин на основе мРНК гена гемагглютинаина (НА) представляется весьма привлекательной. Как правило, мРНК-вакцина является продуктом рекодирования, обеспечивающего стабильность мРНК. Однако результаты рекодирования мРНК бывают неоднозначными. Цель данного сообщения — проанализировать особенности генов и белков НА и рассмотреть возможности и ограничения их рекодирования. Источником первичных структур белков НА и их генов служили общедоступные в Интернете базы данных. Определяли аминокислотный состав и частоту дипептидов, нуклеотидный и динуклеотидный составы, %GC, трансляционный код и составы соседствующих ди- и трикодонов, распределение по первичной структуре НА явных и синонимических мутаций. Подтипы H1N1 и H3N2 имеют в их НА генах как частные, так и общие особенности (ограничения), различаясь не только по числу замен в самом белке, но и по числу и распределению синонимических кодонов в гене, не проявляющихся в первичной структуре самого белка НА, но выступающих, по-видимому, как скрытый фактор, обуславливающий низкую эффективность классических противогриппозных вакцин. Выявление нескольких ограничений в структуре генов НА предполагает, что любая ее модификация (в любом гене) должна не противоречить каждому из ограничений, установленных природой. Частота динуклеотидов CpG у всех исследованных штаммов низкая, но возможность оптимизации ее у штаммов H1N1 из-за запрета квартета в гене НА кодонов, кодирующих аргинин, особенно ограничена и может быть реализована через синонимические кодоны других аминокислот (аланина, пролина, треонина или серина). По сравнению с подтипом H1N1 у подтипа H3N2 можно предвидеть больше возможностей в конструировании стабильной мРНК гена НА.

Ключевые слова: вирус гриппа, ген гемагглютинаина, особенности структуры, рекодирование, мРНК, вакцины.

FEATURES OF INFLUENZA VIRUS HEMAGGLUTININ GENES AND THEIR RECODING POSSIBILITIES

Kharchenko E.P.

Sechenov Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences, St. Petersburg, Russian Federation

Abstract. The world has already entered the stage of increasing odds for a new pandemic, which prompts to seek out for new flu vaccines, because existing vaccines demonstrate only suboptimal effectiveness. With the COVID-19 pandemic,

Адрес для переписки:

Харченко Евгений Петрович
194223, Россия, Санкт-Петербург, пр. Мориса Тореза, 44,
ИЭФБ им. И.М. Сеченова РАН.
Тел.: 8 904 338-22-80 (моб.).
E-mail: neuro.children@mail.ru

Contacts:

Eugene P. Kharchenko
194223, Russian Federation, St. Petersburg, Toreza pr., 44,
Sechenov Institute of Evolutionary Physiology and Biochemistry.
Phone: +7 904 338-22-80 (mobile).
E-mail: neuro.children@mail.ru

Для цитирования:

Харченко Е.П. Особенности генов гемагглютинаина вирусов гриппа и возможности их рекодирования // Инфекция и иммунитет. 2024. Т. 14, № 4. С. 709–718. doi: 10.15789/2220-7619-FOA-17578

Citation:

Kharchenko E.P. Features of influenza virus hemagglutinin genes and their recoding possibilities // Russian Journal of Infection and Immunity = Infektsiya i immunitet, 2024, vol. 14, no. 4, pp. 709–718. doi: 10.15789/2220-7619-FOA-17578

© Харченко Е.П., 2024

DOI: <http://dx.doi.org/10.15789/2220-7619-FOA-17578>

the possibility of using mRNA vaccines has been opened up, and a prospect of finding hemagglutinin (HA) gene mRNA-based new influenza vaccines seems very attractive. As a rule, the mRNA vaccine is a product of recoding, which ensures the mRNA stability. However, the results of mRNA recoding can be ambiguous. The purpose of this report is to analyze the features of genes and proteins and to consider opportunities and limitations in their recoding. Primary structures of NA proteins and relevant genes were retrieved from Internet publicly available databases. The amino acid composition and frequency of dipeptides, nucleotide and dinucleotide compositions, %GC, translational code and compositions of neighboring di- and tricodones, distribution along primary structure for explicit and synonymous mutations were determined. H1N1 and H3N2 subtypes have both specific and general features (limitations) in their genes, differing not only in the number of protein substitutions, but also in the number and distribution of gene synonymous codons, which do not manifest in the protein primary structure, but appear, apparently, as a hidden factor, which causes the low effectiveness of classical influenza vaccines. The identification of several limitations in gene structure suggests that its any modification (in any gene) must not contradict each of the restrictions established by nature. The frequency of CpG dinucleotides in all studied strains is low, but a potential for optimizing it in H1N1 strains due to the prohibition of the quartet in the gene for arginine-encoding codons is especially limited and can be implemented through synonymous codons of other amino acids (alanine, proline, threonine or serine). Compared to the H1N1 subtype, the H3N2 subtype can be expected to have more possibilities in constructing stable NA gene mRNA.

Key words: *influenza virus, hemagglutinin gene, structural features, recoding, mRNA, vaccine.*

Введение

Пандемии гриппа возникают с интервалом 10–40 лет, и с 2020 года мир уже вступил в стадию возрастания вероятности возникновения новой пандемии, что побуждает к поиску против нее новых вакцин. Призыв к переосмыслению подходов к разработке вакцин против вирусов, вызывающих респираторные инфекции, представляется своевременным, поскольку эффективность существующих к ним вакцин лишь субоптимальная [11].

Стремительно развивающаяся синтетическая биология открыла возможности быстрого конструирования новых генов и их продуктов [12]. С пандемией COVID-19 прояснилась перспектива использования для иммунизации мРНК-вакцин [3, 7]. В распространении вирусов гриппа и в формировании популяционного иммунитета к ним роль гемагглютинина (HA) является определяющей. Можно предположить, что применительно к гриппу мРНК-вакцины на основе HA могут иметь превосходство над классическими вакцинами прежде всего в том, что мРНК-вакцина как носитель информации о вирусе позволяет избавиться от антигенного груза других белков вируса, не вовлеченных в формирование иммунитета к нему, ограничивая ее состав (в минимальном варианте) 4 молекулами мРНК HA, представляющими 4 разных подтипа вирусов гриппа, рекомендуемых ВОЗ для изготовления классических вакцин из цельных вирионов. В этом аспекте перспектива поиск новых противогриппозных вакцин на основе мРНК HA представляется весьма рациональной.

Как правило, мРНК-вакцина является продуктом рекодирования отдельного гена либо генома, например вируса, подвергнутого разным модификациям, обеспечивающим стабиль-

ность мРНК [9, 10, 13, 14]. Ныне признано, что мРНК не только задает первичную структуру белка, но и содержит в себе регуляторный код. О наличии последнего свидетельствуют возможности изменения синонимическими мутациями стабильности мРНК и ее вторичной структуры, эффективности трансляции, локализации и сплайсинга мРНК и котрансляционного свертывания белка [6, 8].

Успех в новом поиске противогриппозных вакцин на основе мРНК HA определяется прежде всего знанием особенностей кодирования ген HA и допустимыми возможностями его рекодирования. Цель данного сообщения — проанализировать особенности генов HA и рассмотреть возможности и ограничения их рекодирования.

Материалы и методы

Для сравнительного компьютерного анализа были использованы пандемические штаммы H1N1 A/Brevig Mission/1/18, H1N1 A/California/08/2009, H2N2 A/Japan/305/1957 и H3N2 A/Aichi/2/1968 и произвольно выбранные циркулирующие в эпидсезоне 2023–2024 гг. штаммы H1N1 A/New_Jersey/56/2023 и H1N1 A/Berlin/34/2023. Источником первичных структур HA и их генов служили общедоступные в Интернете базы данных (<http://www.ncbi.nlm.nih.gov>, <http://platform.gisaid.org>, <http://viralzone.expasy.org>).

52 гена белков человека, разных как по длине, так и по функциям (от гистона H4 до титина) были использованы для выявления редких дикодонов в первичных структурах генов. Их общая длина составила более 300 000 нуклеотидов. Источником первичных структур генов и белков служила общедоступная в Интернете база данных (<http://www.nextprot.org>). В них

определяли аминокислотный состав и частоту дипептидов, нуклеотидный и динуклеотидный составы, %GC, трансляционный код и составы соседствующих ди- и трикодонов, распределение по первичной структуре НА явных и синонимических мутаций

Для выявления ограничений в кодировании генов были построены также линейные последовательности оцифрованных трикодонов, считанных (имитируя процесс трансляции на рибосомах) со сдвигом рамки на один кодон [2]. Их числовым показателем служил индекс комплементарности (ИК) — сумма водородных связей, которые способны образовывать составляющие их кодоны с тРНК. ИК отдельных кодонов приведены в таблице генетического кода (рис. 1). Были подсчитаны частоты встречаемости разницы между ИК соседствующих трикодонов/кодонов для каждого гена (табл.).

В статье используется международный код аминокислот: А — аланин, С — цистеин, D — аспарагиновая кислота, Е — глутаминовая кислота, F — фенилаланин, G — глицин, H — гистидин, I — изолейцин, K — лизин, L — лейцин, M — метионин, N — аспарагин, P — пролин, Q — глутамин, R — аргинин, S — серин, T — треонин, V — валин, W — триптофан, Y — тирозин. Для обозначения нуклеиновых оснований используется следующая аббревиатура: А — аденин, G — гуанин, С — цитозин, Т — тимин.

Результаты

Чаще всего при рекодировании вирусов или их генов применяют оптимизацию и деоптимизацию частот кодонов и их пар и динуклеотидов CpG и UpA, модификации нуклеотидов [6, 10]. Поиски противогриппозных вакцин до последнего времени не учитывали роль синонимических мутаций в гене НА. Поэтому в этом разделе приводятся данные расширенного анализа, используемые при рекодировании генов.

Специфика кодирования генов НА у вирусов гриппа разных подтипов проявляется прежде

TCT S 7	TGT C 7	TTT F 6	TAT Y 6
TCC S 8	TGC C 8	TTC F 7	TAC Y 7
TCA S 7	TGA Z 0	TTA L 6	TAA Z 0
TCG S 8	TGG W 8	TTG L 7	TAG Z 0
ACT T 7	AGT S 7	ATT I 6	AAT N 6
ACC T 8	AGC S 8	ATC I 7	AAC N 7
ACA T 7	AGA R 7	ATA I 6	AAA K 6
ACG T 8	AGG R 8	ATG M 7	AAG K 7
CCT P 8	CGT R 8	CTT L 7	CAT H 7
CCC P 9	CGC R 9	CTC L 8	CAC H 8
CCA P 8	CGA R 8	CTA L 7	CAA Q 7
CCG P 9	CGG R 9	CTG L 8	CAG Q 8
GCT A 8	GGT G 8	GTT V 7	GAT D 7
GCC A 9	GGC G 9	GTC V 8	GAC D 8
GCA A 8	GGA G 8	GTA V 7	GAА E 7
GCG A 9	GGG G 9	GTG V 8	GAG E 8

Рисунок 1. Таблица генетического кода с приведенными значениями индексов комплементарности кодонов

Figure 1. Table of the genetic code with the given values of the codon complementarity indices

всего по их составу динуклеотидов. На рис. 2 представлены данные по их частотам у штаммов, вызвавших 4 последние пандемии гриппа.

Их выраженные различия практически по всем динуклеотидам предполагают применение разных подходов модификации их мРНК при рекодировании с целью использования в качестве противогриппозных вакцин

Следующим важным источником информации об особенностях кодирования генов НА служит его трансляционный код, позволяющий выявить уровень использования разных синонимических кодонов (СК), и он оказывается особенным у вирусов подтипа H1N1 тем, что из него исключен квартет кодонов CGC, CGT, CGA и CGG, кодирующих аргинин. Последний кодируются в НА лишь двумя оставшимися кодонами AGA и AGG. Примечательно, что этот запрет на квартет кодонов аргинина в гене НА сохраняется, по крайней мере, с 1918 г. и не рас-

Таблица. Различия трикодонов, считываемых со сдвигом на один шаг, по их ИК

Table. Differences of tricodons, read with a shift of one step, according to their complementarity indices

Источник гена гемагглютинаина вирусов гриппа A source of the hemagglutinin gene of influenza viruses	%GC гена %GC of the gene	Длина гена в кодонах The length of the gene in codons	Частота встречаемости разницы последовательных трикодонов по ИК Frequency of occurrence of the difference of successive tricodons in CI			
			Значения разницы Difference values			
			0	1	2	3
HA H1N1 A/Brevig Mission/1/1918	42	566	186	279	105	10
HA H2N2 A/JAPAN/305/1957	42	562	187	280	98	11
HA H3N2 A/AICHI/2/1968	45	566	223	243	105	9
HA H1N1 A/California/04/2009	41	566	207	267	94	12

H1N1 A/SouthCarolina/1/1918				H2N2 A/JAPAN/305/1957				H3N2 A/AICHI/2/1968				H1N1 A/California/04/2009							
A	G	T	C	A	G	T	C	A	G	T	C	A	G	T	C				
A	199	124	156	98	A	197	130	151	93	A	180	120	128	110	A	218	128	150	105
G	142	104	79	68	G	159	115	78	59	G	136	118	71	82	G	127	104	79	66
T	99	128	92	90	T	77	142	96	87	T	73	130	108	93	T	109	117	100	81
C	137	37	82	65	C	138	24	77	65	C	149	39	97	66	C	147	27	78	64

Рисунок 2. Состав динуклеотидов в генах НА пандемических штаммов вирусов гриппа
 Figure 2. The composition of dinucleotides in the genes of pandemic strains of influenza viruses

пространяется на другие белки вируса подтипа H1N1, а в гене НА вирусов подтипа H3N2 у разных штаммов не используется лишь 1 или 2 кодона из упомянутого квартета кодонов аргинина, чаще всего кодоны CGC и CGT (рис. 3).

Хотя по использованию СК гены НА A/Brevig Mission/1/18 и A/California/08/2009 существенно отличаются, общим для них является наиболее редкое использование кодонов, содержащих преимущественно G и C, и наиболее высокая частота (для большинства аминокислот) кодонов, содержащих A и T, что обуславливает низкое значение %GC в генах НА — 41–42%. Расширяет и уточняет описание особенностей кодирования генов встречаемость в них дикодонов. Из-за различий в использовании СК каждому штамму вируса одного подтипа (даже при идентичности первичной структуры белка НА) свойственна своя картина дикодонов. В качестве иллюстрации наличия ограничений в кодировании генов НА на рис. 4 представлена встречаемость дикодонов в гене НА пандемического штамма НА H1N1 A/Brevig Mission/1/1918. Заметим, что при наличии 61 смыслового кодона в генетическом коде минимальная длина гена белка, которая могла бы охватывать хотя бы по одному разу соседствование всех кодонов в дикодонах, должна составлять 3721 кодон. Длина же гена НА составляет всего 564 кодона, поэтому большинство возмож-

ных дикодонов на рис. 4 отсутствуют. Рис. 4 позволяет уточнить численность и состав дикодонов (а также и дипептидов) по первой или второй позициям. Из 564 возможных разных дикодонов для гена НА использованными оказались лишь 490 дикодонов, и из них 426 встречаются только один раз, что свидетельствует, с одной стороны, об ограничениях (уникальности) в соседствовании кодонов и, с другой стороны, о предпочтении у значительно меньшего числа кодонов находиться в соседстве с одним и тем же кодоном для одних и тех же дипептидов. Повторяемость дикодонов не более 3.

Для выяснения, используются ли в генах НА редкие дикодоны, приняв за последние те, что редко встречаются в генах человека, был первоначально выполнен поиск редко встречающихся дикодонов в генах человека. Выявлены 2 их группы (рис. 5). Одна из них включает очень редко встречающиеся дикодоны. Другая охватывает дикодоны, которые отсутствовали в исследованных генах человека. Ни один из дикодонов обеих групп не был выявлен в исследованных генах НА вирусов гриппа.

В природе НА вирусов гриппа относятся к числу наиболее изменчивых, поэтому проведен анализ первичной структуры их генов на существование ограничений в их кодировании. Оцифровка кодонов обозначила возможность

H1N1 A/Brevig Mission/1/1918										H1N1 A/California/04/2009													
TCT	S	7	TGT	C	10	TTT	F	9	TAT	Y	17	TCT	S	8	TGT	C	9	TTT	F	8	TAT	Y	13
TCC	S	6	TGC	C	6	TTC	F	10	TAC	Y	9	TCC	S	4	TGC	C	6	TTC	F	11	TAC	Y	14
TCA	S	16	TGA	Z	1	TTA	L	11	TAA	Z	0	TCA	S	18	TGA	Z	0	TTA	L	5	TAA	Z	1
TCG	S	3	TGG	W	11	TTG	L	9	TAG	Z	0	TCG	S	1	TGG	W	10	TTG	L	10	TAG	Z	0
ACT	T	7	AGT	S	6	ATT	I	10	AAT	N	26	ACT	T	9	AGT	S	7	ATT	I	18	AAT	N	27
ACC	T	8	AGC	S	11	ATC	I	8	AAC	N	16	ACC	T	2	AGC	S	9	ATC	I	7	AAC	N	14
ACA	T	20	AGA	R	12	ATA	I	14	AAA	K	21	ACA	T	23	AGA	R	14	ATA	I	12	AAA	K	27
ACG	T	2	AGG	R	8	ATG	M	8	AAG	K	12	ACG	T	3	AGG	R	4	ATG	M	7	AAG	K	15
CCT	P	2	CGT	R	0	CTT	L	2	CAT	H	10	CCT	P	2	CGT	R	0	CTT	L	1	CAT	H	8
CCC	P	5	CGC	R	0	CTC	L	6	CAC	H	3	CCC	P	4	CGC	R	0	CTC	L	5	CAC	H	7
CCA	P	9	CGA	R	0	CTA	L	9	CAA	Q	10	CCA	P	9	CGA	R	0	CTA	L	15	CAA	Q	7
CCG	P	3	CGG	R	0	CTG	L	14	CAG	Q	7	CCG	P	5	CGG	R	0	CTG	L	10	CAG	Q	8
GCT	A	6	GGT	G	10	GTT	V	5	GAT	D	15	GCT	A	5	GGT	G	9	GTT	V	7	GAT	D	13
GCC	A	7	GGC	G	2	GTC	V	9	GAC	D	10	GCC	A	9	GGC	G	4	GTC	V	5	GAC	D	13
GCA	A	18	GGA	G	22	GTA	V	9	GAA	E	26	GCA	A	17	GGA	G	13	GTA	V	17	GAA	E	24
GCG	A	6	GGG	G	10	GTG	V	9	GAG	E	9	GCG	A	2	GGG	G	14	GTG	V	7	GAG	E	11

Рисунок 3. Таблица трансляционного кода НА штаммов H1N1 A/Brevig Mission/1/1918 и A/California/04/2009

Figure 3. Translation code table for H1N1 strains A/Brevig Mission/1/1918 and A/California/04/2009

А (A)	Б (B)
TCC(S) – GGT(G)	TCC(S) – TCC(S)
TCC(S) – TTC(L)	TCC(S) – TAC(Y)
TCC(S) – ACA(T)	TCC(S) – CGT(R)
TCC(S) – AGA(R)	TCC(S) – CGC(R)
TCC(S) – ATT(I)	TCC(S) – ATA(I)
TCC(S) – GGT(G)	TCC(S) – CAA(Q)
AGT(S) – CAT(H)	TGC(C) – GCG(A)
AGC(S) – GTA(V)	STC(L) – GCG(A)
TGT(C) – GCG(A)	STA(L) – TCG(S)
TGC(C) – GGA(G)	STA(L) – CGG(R)
TTA(L) – CGC(R)	STA(L) – CCG(P)
STC(L) – GTA(V)	STA(L) – GCG(A)
STC(L) – GAA(E)	CGT(R) – AGC(S)
ACG(T) – CGT(R)	CGT(R) – ACG(T)
ACG(T) – ATA(I)	CGT(R) – CCG(P)
CGT(R) – TCG(S)	CGC(R) – GTA(V)
CGT(R) – CTA(L)	CCG(P) – TCG(S)
CGT(R) – CGC(R)	CCG(P) – CGA(R)
CGT(R) – GCG(A)	CAC(H) – TCG(S)
CGC(R) – CCG(P)	GCG(A) – CGA(R)
CGC(R) – CAT(H)	GCG(A) – ATA(I)
CGC(R) – GCG(A)	GGT(G) – ACG(T)
CGC(R) – GGA(G)	GGT(G) – CCG(P)
CGA(R) – TCG(S)	
CGA(R) – GCG(A)	
CGG(R) – TTA(L)	
CGG(R) – GCG(A)	
ATA(I) – TCG(S)	
ATA(I) – CGG(R)	
CCG(P) – CAC(H)	
CAA(Q) – CGC(R)	
GCG(A) – CGT(R)	
GCG(A) – ATG(M)	
GTA(V) – ACG(T)	

Рисунок 5. Список редко встречающихся (А) и не обнаруженных (Б) в исследованных генах человека дикодонов

Figure 5. List of rare (A) and undetected (B) dicodons in the studied human genes

23, 23, 22, 22, 23, 23, 21, 20, 21, 21, 22, 22, 23, 21, 21, 22, 23, 21, 20, 19, 22, 22, 23, 23, 22, 22, 20, 22, 23, 23, 22, 22, 22, 22, 20, 21, 22, 24, 23, 23, 22, 22, 21, 22, 23, 23, 23, 23, 24, 23, 23, 21, 21, 20, 20, 19, 18, 20, 20, 23, 23, 23, 21, 20, 23, 22, 22, 19, 20, 22, 24, 25, 24, 23, 23, 22, 24, 23, 24, 22, 21, 21, 22, 23, 24, 24, 24, 23, 23, 22, 20, 19, 20, 21, 22, 22, 22, 22, 21, 22, 21, 22, 21, 22, 23, 23, 22, 21, 22, 21, 21, 20, 22, 23, 24, 23, 22, 22, 22, 23, 22, 23, 21, 21, 19, 20, 20, 19, 21, 22, 23, 22, 23, 24, 25, 23, 22, 20, 21, 22, 21, 22, 21, 23, 23, 24, 24, 24, 22, 23, 23, 25, 24, 23, 21, 20, 20, 20, 20, 21, 23, 24, 23, 22, 21, 22, 23, 23, 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 23, 21, 22, 21, 21, 19, 21, 21, 22, 21, 22, 22, 23, 23, 22, 21, 23, 24, 25, 24, 23, 22, 21, 22, 20, 22, 21, 23, 21, 22, 21, 22, 23, 23, 20, 19, 19, 21, 22, 22, 22, 24, 24, 22, 21, 23, 24, 25, 22, 22, 20, 21, 21, 22, 24, 25, 24, 22, 20, 20, 21, 22, 21, 20, 20, 23, 24, 25, 23, 21, 20, 19, 21, 22, 23, 23, 21, 21, 19, 21, 22, 24, 22, 22, 21, 23, 23, 22, 21, 21, 22, 23, 23, 24, 23, 22, 21, 21, 22, 23, 24, 24, 23, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 22, 22, 21, 21, 21, 21, 21, 21, 21, 21, 21, 20, 20, 21, 23, 22, 20, 19, 20, 21, 23, 24, 23, 22, 20, 22, 23, 24, 23, 21, 21, 22, 24, 24, 23, 21, 22, 23, 22, 22, 22, 24, 25, 25, 24, 23, 22, 23, 22, 22, 21, 20, 22, 22, 23, 22, 23, 23, 21, 21, 20, 21, 22

Рисунок 6. Последовательность значений индексов комплементарности трикодонов НА штамма Н1N1 А/Brevig Mission/1/1918 при сдвиге рамки считывания на 1 кодон
Figure 6. The sequence of values of the tricondon complementarity indices of the H1N1 A/Brevig Mission/1/1918 HA gene when the reading frame is shifted by 1 codon

сокий показатель, как и в случае НА пандемических штаммов, по СК их генов — 21. Столь высокое преобладание синонимических замен и особенности их локализации возможно связаны с «выравниванием» разницы по ИК между соседствующими кодонами, позволяющим закрепить в структуре гена тем мутациям, что привели к замещениям аминокислот.

Обсуждение

Данные выполненного нами анализа свидетельствуют о наличии у генов НА вирусов гриппа разных подтипов как частных, так и общих особенностей (ограничений), которые необходимо учитывать при их рекодировании, имея в виду, что и штаммы одного подтипа в пределах эпидсезона могут иметь различия НА не только по числу замен в самом белке, которые могут затрагивать до 30–40 позиций, но и еще большие различия по числу и распределению СК в гене, не проявляющиеся в первичной структуре самого белка НА и выступающие, по-видимому, как скрытый фактор, обуславливающий низкую эффективность классических противогриппозных вакцин.

По динуклеотидам различия между генами НА выражены не только между подтипами, но и для штаммов одного и того же подтипа, выделенных в разное время. При рекодировании оптимизация и деоптимизация частот динуклеотидов по CpG и UpA проводится посредством введения в ген СК [6, 10]. Частота CpG у всех исследованных штаммов самая низкая, но возможность оптимизации ее у штаммов H1N1 из-за запрета квартета кодонов, кодирующих аргинин, ограничена и может быть реализована через СК других аминокислот (аланина, пролина, треонина или серина).

Следует сразу оговорить, что оптимизация частот динуклеотидов через СК должна быть подчинена соблюдению ограничения, накладываемого значениями разности ИК между соседствующими трикодонами, учитывая очень низкую частоту встречаемости разности, равной 3. Поясним это на конкретном примере — на фрагменте последовательности НА H1N1 A/Brevig Mission/1/18, представленного на рис. 8. Выберем в качестве мишени для оптимизации частоты динуклеотида CpG кодон АСА, кодирующий треонин и находящийся в третьей позиции выделенного серым цветом трикодона, и заменим его синонимическим триплетом АСG, имеющим значение ИК, равное 8 (см. рис. 1). В результате такой замены выделенный серым цветом трикодон превратится в трикодон GCA GAC АСG и с его ИК, равным 24, разница между соседними трикодонами, будет равна 3. Последняя очень редко встречается

61 L = TTA CTA*	91 S = AGC =
62 K R AAA AGA	92 S = TCA =
63 G = GGA GGG*	93 W = TGG =
64 I V ATA GTA	94 S = TCC =
65 A = GCC =	95 Y = TAT TAC*
66 P = CCA =	96 I = ATT =
67 L = TTA TTG*	97 V = GTA GTG*
68 Q H CAA CAT	98 E = GAA =
69 L = TTG =	99 T = ACA =
70 G = GGG GGT*	100 S P TCG CCT
71 K = AAA =	101 N S AAC AGT
72 C = TGT =	102 S = TCA =
73 N = AAT AAC*	103 E D GAG GAC
74 I = ATC ATT*	104 N = AAT =
75 A = GCC GCT*	105 G = GGA =
76 G = GGA GGC*	106 T = ACA ACG*
77 W = TGG =	107 C = TGT =
78 L I CTC ATC	108 Y = TAC =
79 L = TTG CTG*	109 P = CCA =
80 G = GGA =	110 G = GGA =
81 N = AAC AAT*	111 D = GAT =
82 P = CCG CCA*	112 F = TTC =
83 E = GAA GAG*	113 I = ATC =
84 C = TGC TGT*	114 D = GAC GAT*
85 D E GAT GAA	115 Y = TAT =
86 L S TTA TCA	116 E = GAA GAG*
87 L = CTG CTC*	117 E = GAA GAG*
88 L S CTC TCC	118 L = CTG CTA*
89 T = ACA =	119 R = AGG AGA*
90 A = GCG GCA*	120 E = GAG =

Рисунок 7. Различия первичных структур фрагмента НА пандемических штаммов A/Brevig Mission/1/1918 и A/California/04/2009

Figure 7. Differences in the primary structures of the HA fragment in the pandemic strains A/Brevig Mission/1/1918 and A/California/04/2009

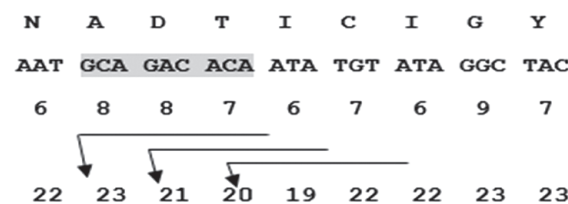


Рисунок 8. Иллюстрация расчета индексов комплементарности трикодонов, считываемых сдвигом на 1 кодон

Figure 8. An illustration of the calculation of a complementarity index in tricodons read by a shift of one codon

Примечание. 1-й ряд букв — последовательность аминокислот, 2-й ряд — обозначения кодонов, 3-й ряд — обозначения индексов комплементарности кодонов, 4-й ряд — обозначения индексов комплементарности трикодонов.

Note. The 1st row of letters is a sequence of amino acids, the 2nd row — a designation of codons, the 3rd row — a designation of codon complementarity indices, the 4th row — a designation of tricodon complementarity indices.

```

143 W W
144 A .V[1]A[213]T[164]
145 G G
146 V .I[2]V[375]A[1]
147 T .K[138]T[240]
148 Q Q
149 N N
150 G .E[1]G[377]
151 T .K[175]A[2]N[1]T[200]
152 S S
153 S .F[7]S[371]
154 A .A[309]S[69]
155 C C

```

Рисунок 9. Фрагмент доминирующей последовательности гемагглютинаина для штаммов H3N2 эпидсезона 2018–2019 гг.

Figure 9. Fragment of the dominant hemagglutinin sequence for H3N2 strains of the epidemic seasons 2018–2019

между соседними трикодонами в HA вирусов гриппа. Ее частота, варьируя у HA в диапазоне 1,6–2,1%, на 1% ниже чем у генов человека, то есть возможности модуляции частоты CpG при рекодировании, используемой, как правило, с целью стабилизации мРНК, у гена H1 HA, по сравнению с H3 HA, весьма ограничены.

По использованной нами модели считывания триплетов в мРНК каждый кодон, за исключением первых и последних трех кодонов, включается, как и при трансляции мРНК на рибосомах, в три последовательно считываемые трикодона, и очевидно, что ИК каждого кодона в трикодоне связан с ИК двух предшествующих и двух последующих кодонов. Это подводит к признанию существования в генах континуума связности кодонов по значениям их ИК и ограничения на их мутагенез. Свидетельством связности кодонов служит иллюстрируемая на рис. 7, на примере генов HA штаммов A/Brevig Mission/1/18 и A/California/04/2009, высокая частота и линейное распределение СК между теми кодонами, с которыми связана замена аминокислот. Как метко подмечено Комаром А.А., «синонимические кодоны размещены в мРНК стратегически (в определенных местах) таким образом, что они задают специфическую кинетику трансляции, обеспечивающую, в свою очередь, эффективное поэтапное котрансляционное сворачивание белка», то есть являются вторичным кодом для сворачивания белка [8].

У штаммов вируса гриппа одного и того же подтипа, как отмечалось выше, и эпидсезона различия HA связаны не только с заменами аминокислот, но и с еще большим различием в числе и распределении СК. Последние служат причиной образования у каждого штамма вируса гриппа своего конформера HA и своей мозаики обнажения и доступности в нем эпитопов. Социркуляция в эпидсезоне штаммов вируса гриппа с разными первичными структурами HA и составами СК в их генах и отличие от них

по этим же показателям вакцинного штамма являются (следует признать) другим фактором, обрекающим противогриппозные вакцины любой конструкции на невысокую эффективность. В США, например, эффективность вакцины против гриппа на субоптимальном уровне — 14–60%, а иммунитет к гриппу сохраняется лишь несколько месяцев [11].

Если таблица трансляционного кода (рис. 3) служит обобщающей информацией об использовании СК в мРНК, то двумерная комбинированная таблица (рис. 4) служит источником информации по дипептидам в белке HA и соответствующим им дикодоном в мРНК и полезна особенно тем, что в ней видно, какие СК сочетаются (или не сочетаются) друг с другом и как часто. Избирательность в соседствовании между кодонами служит, наряду с ограничениями в разнице по ИК между соседними кодонами, вторым барьером, который нужно обойти при рекодировании генов. Третьим же барьером являются редко встречающиеся и не обнаруженные в исследованных нами генах человека дикодоны, перечисленные на рис. 5. Эти дикодоны можно рассматривать как запретные для использования при рекодировании.

Учет перечисленных нами барьеров в рекодировании генома/гена посредством изменения состава СК переводит синтетическую биологию на другой уровень, способствуя избегания неудач. Ведь известно, что результаты рекодирования мРНК неоднозначны и могут привести к ухудшению свойств у новых вариантов синтетических геномов или генов. Более того, введенные в геном вируса синонимические замены могут реверсироваться к кодонам диких штаммов [6, 10], что можно рассматривать как существование ограничений в их рекодировании и подтверждение реальности выявленных нами барьеров, которые нельзя игнорировать, модифицируя гены или вирусы.

Нельзя обойти вниманием избирательность и неизменный консерватизм при исключении из гена H1 HA квартета кодонов, кодирующих аргинин. Если вновь обратиться к списку редко встречающихся и не обнаруженных в исследованных нами генах человека дикодонов (рис. 5), то нетрудно заметить, что в нем преобладают, с одной стороны, кодоны аминокислот (серин, лейцин и аргинин), кодируемых 6 кодонами, а с другой стороны, кодоны с динуклеотидом CpG, что побуждает объяснить исключение квартета кодонов CGC, CGT, CGA и CGG из гена H1 HA как проявление общего запрета по дикодоном. Однако эта аргументация отмечается ограниченностью исключения только генами H1 HA и не распространением ее на другие гены вирусов гриппа подтипа H1N1 и на другие подтипы вирусов гриппа.

Приемлемым объяснением исключения квартета кодонов CGC, CGT, CGA и CGG из гена H1 HA может быть связано с особенностями структуры самого белка HA, в частности ограничениями в его инвариантной последовательности. Так, следствием запрета в трансляционном коде гена H1 HA на кодоны CGC, CGT, CGA и CGG будут соответственно ограничения на замены по 8 аминокислотам (см. рис. 3): глицину, триптофану, цистеину, серину (в случае кодирования его кодонами с корнем AG), пролину, лейцину (в случае кодирования его кодонами с корнем СТ), гистидину и глутамину. Ограничения на цистеин, триптофан, глицин и лейцин связаны с присутствием их в инвариантных позициях первичной структуры белка H1, и запрет на использование в коде гена H1 всего квартета триплетов аргинина с корнем CG пресекает возможности мутирования кодонов этих аминокислот.

Выявление скрытой роли СК в невысокой эффективности противогриппозных вакцин еще более убеждает в том, что возможность успешного использования принципа «один иммуноген для всех штаммов подтипа вируса» маловероятна. Один из путей решения проблемы с увеличением эффективности противогриппозных вакцин применительно к мРНК вакцин мог бы заключаться в привлечении против штаммов одного подтипа 2–3 иммуногенов, то есть 2–3 мРНК, охватывающих более широкий спектр циркулирующих в эпидсезоне штаммов соответствующего подтипа, опираясь на представление о доминантных штаммах подтипа, упоминаемых в планах США и ВОЗ по поиску универсальных противогриппозных вакцин [4, 5]. Конструирование соответствующих им мРНК могло бы основываться на том же принципе (но более расширенном), что был предложен нами ранее для построения доминантной белковой структуры HA одного или нескольких эпидсезонов, в зависимости от целевой установки [1].

Приведем с помощью рис. 9 краткое изложение процедуры построения фрагмента аминокислотной последовательности доминантного H3 HA эпидсезона 2018–2019 гг., которая легко трансформируется в доминантную мРНК заменой аминокислот соответствующими кодонами. На рис. 9 первый вертикальный числовой ряд — нумерация позиций аминокислот в HA, второй вертикальный ряд аминокислот — собственно сама построенная доминантная последовательность

HA эпидсезона, параллельно ей, правее — «скелет» молекулы с расположением инвариантных (представлены обозначениями аминокислот) и вариабельных (представлены точками) сайтов. За последними следует горизонтальный ряд всех замещающих аминокислот, обнаруженных в этих позициях HA у разных штаммов эпидсезона, с указанием (в скобках) для каждой аминокислоты частоты ее встречаемости. Строилась доминантная последовательность HA вставлением в цепь с инвариантными блоками доминантной аминокислоты из приводимого горизонтального ряда заменяемых аминокислот [1].

Как следует из рис. 9, доминантная последовательность HA не охватывает все циркулирующие штаммы, и нередки ситуации (например, позиции 147 и 151), когда в соответствующей позиции доминантная аминокислота по численности не намного превосходит другую аминокислоту в этой же вариабельной позиции. Последнюю условно можно назвать как субдоминантную. Поэтапно можно построить несколько субдоминантных HA, замещая каждую вариабельную позицию инвариантной последовательности HA новыми уменьшающимися по частоте встречаемости аминокислотами. В совокупности с доминантной последовательностью HA они обеспечивают более широкий охват HA штаммов соответствующего подтипа. В отличие от мРНК вакцин, использование описанного подхода применительно к классическим вакцинам из вирионов сопряжено с возрастанием антигенной нагрузки вакцины за счет белков вирусов, не вовлеченных в формирование к ним иммунитета.

Завершая обсуждение результатов биоинформационного анализа первичной структуры генов HA вирусов гриппа хотелось бы подчеркнуть полезность его в выявлении тех их особенностей, которые потенциально могли бы ограничивать успешность их рекодирования с целью получения на их основе мРНК вакцины. Выявление нескольких ограничений в структуре генов предполагает, что любая ее модификация (в любом гене) должна не противоречить каждому из ограничений, установленных природой. Сравнивая HA подтипов H1N1 и H3N2, можно предвидеть больше возможностей в конструировании стабильной мРНК у подтипа H3N2 из-за исключения из трансляционного кода у гена подтипа H1N1 квартета кодонов, кодирующих аргинин.

Список литературы/References

1. Харченко Е.П. Поиски универсальной противогриппозной вакцины: возможности и ограничения // Эпидемиология и вакцинопрофилактика. 2019. Т. 18, № 5. С. 70–84. [Kharchenko E.P. The search for a universal influenza vaccine: possibilities and limitations. *Epidemiologiya i vaksinoprofilaktika = Epidemiology and Vaccine Prevention*, 2019, vol. 18, no. 5, pp. 70–84. (In Russ.)] doi: 10.31631/2073-3046-2019-18-5-70-84

2. Харченко Е.П. Проблемы и коллизии вакцинологии // Эпидемиология и вакцинопрофилактика. 2023. Т. 22, № 6. С. 183–200. [Kharchenko E.P. Problems and Collisions of Vaccinology. *Epidemiologiya i vaksinoprofilaktika = Epidemiology and Vaccine Prevention*, 2023, vol. 22, no. 6, pp. 183–200. (In Russ.)] doi: 10.31631/2073-3046-2023-22-6-183-200
3. Altmann D.M., Boyton R.J. COVID-19 vaccination: the road ahead. *Science.*, 2022, vol. 375, no. 6585, pp. 1127–1132. doi: 10.1126/science.abn1755
4. Erbelding E.J., Post D.J., Stemmy E.J., Roberts P.C., Augustine A.D., Ferguson S., Paules C.I., Graham B.S., Fauci A.S. A universal influenza vaccine: the strategic plan for the National Institute of Allergy and Infectious Diseases. *J. Infect. Dis.*, 2018, vol. 218, no. 3, pp. 347–354. doi: 10.1093/infdis/jiy103
5. Global influenza strategy 2019–2030. World Health Organization. URL: <https://www.who.int/publications/i/item/978924151532015>
6. Gonçalves-Carneiro D., Bieniasz P.D. Mechanisms of attenuation by genetic recoding of viruses. *mBio*, 2021, vol. 12, no. 1: e02238-20. doi: 10.1128/mBio.02238-20
7. Knyazev S., Chhugani K., Sarwal V., Ayyala R., Singh H., Karthikeyan S., Deshpande D., Baykal P.I., Comarova Z., Lu A., Porozov Y., Vasylyeva T.I., Wertheim J.O., Tierney B.T., Chiu C.Y., Sun R., Wu A., Abedalthagafi M.S., Pak V.M., Nagaraj S.H., Smith A.L., Skums P., Pasaniuc B., Komissarov A., Mason C.E., Bortz E., Lemey P., Kondrashov F., Beerenwinkel N., Lam T.T., Wu N.C., Zelikovsky A., Knight R., Crandall K.A., Mangul S. Unlocking capacities of genomics for the COVID-19 response and future pandemics. *Nat. Methods*, 2022, vol. 19, no. 4, pp. 374–380. doi: 10.1038/s41592-022-01444-z
8. Komar A.A. [Synonymous codon usage — a guide for co-translational protein folding in the cell]. *Mol. Biol. (Mosk.)*, 2019, vol. 53, no. 6, pp. 883–898. doi: 10.1134/S0026898419060090
9. Lin B.C., Kaissarian N.M., Kimchi-Sarfaty C. Implementing computational methods in tandem with synonymous gene recoding for therapeutic development. *Trends Pharmacol. Sci.*, 2023, vol. 44, no. 2, pp. 73–84. doi: 10.1016/j.tips.2022.09.008
10. Martínez M.A., Jordan-Paiz A., Franco S., Nevot M. Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol.*, 2016, vol. 24, no. 2, pp. 134–147. doi: 10.1016/j.tim.2015.11.002
11. Morens D.M., Taubenberger J.K., Fauci A.S. Rethinking next-generation vaccines for coronaviruses, influenzaviruses, and other respiratory viruses. *Cell. Host Microbe*, 2023, vol. 31, no. 1, pp. 146–157. doi: 10.1016/j.chom.2022.11.016
12. Ostrov N., Nyerges A., Chiappino-Pepe A., Rudolph A., Baas-Thomas M., Church G.M. Synthetic genomes with altered genetic codes. *Curr. Opin. Syst. Biol.*, 2020, vol. 24, pp. 32–40. doi: 10.1016/j.coisb.2020.09.007
13. Tian Y., Deng Z., Yang P. mRNA vaccines: A novel weapon to control infectious diseases. *Front. Microbiol.*, 2022, no. 13: 1008684. doi: 10.3389/fmicb.2022.1008684
14. Yang L., Tang L., Zhang M., Liu C. Recent advances in the molecular design and delivery technology of mRNA for vaccination against infectious diseases. *Front. Immunol.*, 2022, no. 13: 896958. doi: 10.3389/fimmu.2022.896958

Автор:

Харченко Е.П., д.б.н., ведущий научный сотрудник Института эволюционной физиологии и биохимии им. И.М. Сеченова РАН, Санкт-Петербург, Россия.

Author:

Kharchenko E.P., DSc (Biology), Senior Researcher, Sechenov Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences, St. Petersburg, Russian Federation.

Поступила в редакцию 27.01.2024
Принята к печати 08.08.2024

Received 27.01.2024
Accepted 08.08.2024