

ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА
FEATURES OF HEMAGGLUTININ GENES OF INFLUENZA VIRUSES 10.15789/2220-7619-FOA-17578
**ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА
И ВОЗМОЖНОСТИ ИХ РЕКОДИРОВАНИЯ**

Харченко Е. П.¹

¹Института эволюционной физиологии и биохимии им. И. М. Сеченова РАН.

**FEATURES OF AND POTENTIAL FOR RECODING INFLUENZA VIRUSE
HEMAGGLUTININ GENES**

Kharchenko Eugene P. ^a

^a Sechenov Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences, St. Petersburg, Russia.

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 в 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 выступающие, 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 HA.

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

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, 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.

Keywords: influenza virus, hemagglutinin gene, structural features, recoding, mRNA, vaccine.

1 Введение

Пандемии гриппа возникают с интервалом 10–40 лет, и с 2020 г. мир уже вступил в стадию возрастания вероятности возникновения новой пандемии, что побуждает к поиску против нее новых вакцин. Призыв к переосмысливанию 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 проявилась перспектива использования для иммунизации мРНК-вакцин [3,7]. В распространении вирусов гриппа и в формировании популяционного иммунитета к ним роль гемагглютинина (НА) является определяющей. Можно предположить, что применительно к гриппу мРНК-вакцины на основе НА 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 НА, представляющими 4 разных подтипов вирусов гриппа, рекомендуемых ВОЗ для изготовления классических вакцин из цельных вирионов. В этом аспекте перспектива поиска новых противогриппозных вакцин на основе мРНК НА представляется весьма рациональной.

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

29 ее вторичной структуры, эффективности трансляции, локализации и
30 сплайсинга мРНК и котрансляционного свертывания белка [6, 8].

31 Успех в новом поиске противогриппозных вакцин на основе мРНК НА
32 определяется прежде всего знанием особенностей кодирования ген НА и
33 допустимыми возможностями его рекодирования. Цель данного сообщения -
34 проанализировать особенности генов НА и рассмотреть возможности и
35 ограничения в их рекодирования.

36 2 Материалы и методы

37 Для сравнительного компьютерного анализа были использованы
38 пандемические штаммы H1N1 A/Brevig Mission/1/18, H1N1
39 A/California/08/2009, H2N2 A/Japan/305/1957 и H3N2 A/Aichi/2/1968 и
40 произвольно выбранные циркулирующие в эпидсезоне 2023-2024 гг.. штаммы
41 H1N1 A/New_Jersey/56/2023 и H1N1 A/Berlin/34/2023. Источником
42 первичных структур НА и их генов служили обще доступные в Интернете
43 базы данных ([http:// www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov), <http://platform.gisaid.org>,
44 <http://viralzone.expasy.org>).

45 52 гена белков человека, разных как по длине, так и по
46 функциям (от гистона H4 до титина) были использованы для выявления
47 редких дикодонов в первичных структурах генов. Их общая длина составила
48 более 300000 нуклеотидов. Источником первичных структур генов и
49 белков служила общедоступная в Интернете база данных
50 (<http://www.nextprot.org>). В них определяли аминокислотный состав и
51 частоту дипептидов, нуклеотидный и динуклеотидный составы, %GC,
52 трансляционный код и составы соседствующих ди- и трикодонов,
53 распределение по первичной структуре НА явных и синонимических мутаций

54 Для выявления ограничений в кодировании генов были построены также
55 линейные последовательности оцифрованных трикодонов, считанных
56 (имитируя процесс трансляции на рибосомах) со сдвигом рамки на один
57 кодон [2]. Их числовым показателем служил индекс комплементарности

58 (ИК) - сумма водородных связей, которые способны образовывать
59 составляющие их кодоны с тРНК. ИК отдельных кодонов приведены в таблице
60 генетического кода (рис. 1). Были подсчитаны частоты встречаемости разницы
61 между ИК соседствующих трикодонов/кодонов для каждого гена (таблица
62 1).

63 В статье используется международный код аминокислот: А – аланин, С
64 – цистеин, D – аспа- рагиновая кислота, E – глутаминовая кислота, F –
65 фенилаланин, G — глицин, H – гистидин, I – изолейцин, K – лизин, L – лейцин,
66 M – метионин, N – аспарагин, P – пролин, Q – глутамин, R – арги- нин, S –
67 серин, T – треонин, V – валин, W – трип- тофан, Y – тирозин. Для обозначения
68 нуклеиновых оснований используется следующая аббревиатура: A – аденин,
69 G – гуанин, C– цитозин, T – тимин.

70 3 Результаты

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

77 Специфика кодирования генов HA у вирусов гриппа разных подтипов
78 проявляется прежде всего по их составу динуклеотидов. На рис.2
79 представлены данные по их частотам у штаммов, вызвавших 4 последние
80 пандемии гриппа.

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

84 Следующим важным источником информации об особенностях
85 кодирования генов HA служит его трансляционный код, позволяющий
86 выявить уровень использования разных синонимических кодонов (СК), и он

87 оказывается особенным у вирусов подтипа H1N1 тем, что из него исключен
88 квартет кодонов CGC, CGT, CGA и CGG, кодирующих аргинин. Последний
89 кодируются в HA лишь двумя оставшимися кодонами AGA и AGG.
90 Примечательно, что этот запрет на квартет кодонов аргинина в гене HA
91 сохраняется, по крайней мере, с 1918 г. и не распространяется на другие белки
92 вируса подтипа H1N1, а в гене HA вирусов подтипа H3N2 у разных штаммов
93 не используется лишь 1 или 2 кодона из упомянутого квартета кодонов
94 аргинина, чаще всего кодоны CGC и CGT.

95 Хотя по использованию СК гены HA A/Brevig Mission/1/18 и
96 A/California/08/2009 существенно отличаются, общим для них является
97 наиболее редкое использование кодонов, содержащих преимущественно G и
98 C, и наиболее высокая частота (для большинства аминокислот) кодонов,
99 содержащих A и T, что обуславливает низкое значение %GC в генах HA – 41-
100 42% Расширяет и уточняет описание особенностей кодирования генов
101 встречаемость в них дикодонов. Из-за различий в использовании СК каждому
102 штамму вируса одного подтипа (даже при идентичности первичной структуры
103 белка HA) свойственна своя картина дикодонов. В качестве иллюстрации
104 наличия ограничений в кодировании генов HA на рисунке 3 представлена
105 встречаемость дикодонов в гене HA пандемического штамма HA H1N1
106 A/Brevig Mission/1/1918. Заметим, что при наличии 61 смысловых кодонов в
107 генетическом коде минимальная длина гена белка, которая могла бы
108 охватывать хотя бы по одному разу соседствование всех кодонов в дикодонах,
109 должна составлять 3721 кодонов. Длина же гена HA составляет всего 564
110 кодона, поэтому большинство возможных дикодонов на рисунке 3
111 отсутствуют. Рисунок 3 позволяет уточнить численность и состав дикодонов
112 (а также и дипептидов) по первой или второй позициям. Из 564 возможных
113 разных дикодонов для гена HA использованными оказались лишь 490
114 дикодонов, и из них 426 встречаются только один раз, что свидетельствует, с
115 одной стороны, об ограничениях (уникальности) в соседствовании кодонов и,

116 с другой стороны, о предпочтения у значительно меньшего числа кодонов
117 находиться в соседстве с одним и тем же кодоном для одних и тех же
118 дипептидов. Повторяемость дикодонов не более 3.

119 Для выяснения, используются ли в генах НА редкие дикодоны, приняв
120 за последние те, что редко встречаются в генах человека, был первоначально
121 выполнен поиск редко встречающихся дикодонов в генах человека. Выявлены
122 2 их группы (рис. 5). Одна их них включает очень редко встречающиеся
123 дикодоны. Другая охватывает дикодоны, которые отсутствовали в
124 исследованных генах человека. Ни один из дикодонов обеих групп не был
125 выявлен в исследованных генах НА вирусов гриппа.

126 В природе НА вирусов гриппа относятся к числу наиболее изменчивых,
127 поэтому проведен анализ первичной структуры их генов на существование
128 ограничений в их кодировании. Оцифровка кодонов обозначила возможность
129 разделения таблицы генетического кода (рис.1) на четыре квадрата и
130 соответственно выделить три группы триплетов по значениям их ИК. Первая
131 группа, включающая триплеты кодирующие пролин, аланин, аргинин и
132 глицин, характеризуется наибольшими (8 и 9) значениями их ИК, и
133 составленные из них трикодоны имеют максимальное значение ИК – 27.
134 Тирозин, фенилаланин, лейцин (кодоны ТТА и TTG), изолейцин, метионин,
135 аспарагин и лизин составляют вторую группу кодонов с ИК, равными 6 или
136 7. Трикодоны из них могут иметь минимальное значение ИК – 18. Кодоны
137 остальных аминокислот имеют промежуточные значения ИК – 7 или 8.
138 Поскольку считывание кодонов проводится со сдвигом рамки на один кодон,
139 то разница между ИК соседствующих трикодонов может принимать значения
140 от 0 до 3. На рисунке 6 представлена в качестве иллюстрации линейная
141 последовательность ИК трикодонов гена НА штамма H1N1 A/Brevig
142 Mission/1/1918, считанных со сдвигом рамки на один кодон, имитируя
143 процесс трансляции на рибосомах.

144 Как уже отмечалось в разделе «Материалы и методы», при каждом
145 сдвиге считывания трикодонов заменяется лишь один триплет и
146 максимальная разница между ИК соседних трикодонов будет равна 3. У
147 гена НА штамма H1N1 A/Brevig Mission/1/1918 максимальное и
148 минимальное значения ИК составляют соответственно 25 и 19. Лишь у 10
149 соседствующих пар трикодонов (таблицы 1) ИК имеют
150 разницу 3, а у остальных она не более 2, т. е. кодирование гена НА штамма
151 H1N1 A/Brevig Mission/1/1918 имеет явные ограничения. Об общности
152 (универсальности) этих ограничений свидетельствуют приведенные в
153 таблице 1 показатели разниц ИК трикодонов и генов НА других
154 пандемических штаммов. У генов НА пандемических штаммов вирусов
155 гриппа не более 2.1% соседствующих пар трикодонов (таблица 1) ИК
156 имеют разницу 3, а разницу по ИК 0 или 1 имеют 82-84% соседствующих
157 трикодонов, т.е. даже разница между ИК, равной 2, в генах НА не
158 предпочтительна. Следует специально оговорить, что при свойственной ГК
159 композиции выявленные общие ограничения в кодировании генов, благодаря
160 вырожденности генетического кода, не ведут к запретам соседствования
161 аминокислот в первичной структуре белков.

162 В качестве иллюстрации, что каждый штамм вирусов гриппа уникален
163 хотя бы по распределению в его НА гене СК, для сравнения были выбраны две
164 пары штаммов. Одна из них включает пандемические штаммы A/Brevig
165 Mission/1/1918 и A/California/04/2009, разделенные по времени возникновения
166 более чем в 90 лет, другая пара – два штамма подтипа H1N1, выделенные в
167 ноябре 2023 г. Сравнение первичных структур белков и генов пандемических
168 штаммов 1918 г. и 2009 г. выявил, что белки НА A/California/04/2009
169 отличается от НА A/South Carolina/1/1918 по 79 позициям, которым
170 сопутствуют различия по 191 СК. На рис. 7 представлен фрагмент
171 последовательностей белков и генов НА обоих штаммов со специальной
172 пометкой (*) позиций, различающихся по СК. Очень примечательна весьма

173 частая сблоченность последних и чередование их с мутациями, вызвавших
174 замены аминокислот. Два штамма H1N1 эпидсезона 2023/2024 гг.
175 естественно имели более скромные различия их первичных структур белков
176 НА (всего по 14 позициям) и более высокий показатель, как и в случае НА
177 пандемических штаммов, по СК их генов – 21. Столь высокое преобладание
178 синонимических замен и особенности их локализации, возможно, связаны с
179 «выравниванием» разницы по ИК между соседствующими кодонами,
180 позволяющим закрепиться в структуре гена тем мутациям, что привели к
181 замещениям аминокислот.

182 4 Обсуждение

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

193 По динуклеотидам различия между генами НА выражены не только
194 между подтипами, но и для штаммов одного и того же подтипа, выделенных в
195 разное время. При рекодировании оптимизация и деоптимизация частот
196 динуклеотидов по CpG и UpA проводится посредством введения в ген СК
197 [6,10]. Частота CpG у всех исследованных штаммов самая низкая, но
198 возможность оптимизации ее у штаммов H1N1 из-за запрета квартета
199 кодонов, кодирующих аргинин, ограничена и может быть реализована через
200 СК других аминокислот (аланина, пролина, треонина или серина).

201 Следует сразу оговорить, что оптимизация частот динуклеотидов через
202 СК должна быть подчинена соблюдению ограничения, накладываемого
203 значениями разности ИК между соседствующими трикодонами, учитывая
204 очень низкую частоту встречаемости разности, равной 3. Поясним это на
205 конкретном примере - на фрагменте последовательности НА H1N1 A/Brevig
206 Mission/1/18, представленного на рис 8 . Выберем в качестве мишени для
207 оптимизации частоты динуклеотида CpG кодон АСА, кодирующий треонин и
208 находящийся в третьей позиции выделенного серым цветом трикодона, и
209 заменим его синонимическим триплетом АСГ, имеющим значение ИК, равное
210 8 (см. рис.1). В результате такой замены выделенный серым цветом трикодон
211 превратится в трикодон GCA GAC АСГ и с его ИК, равным 24, разница между
212 соседними трикодонами, будет равна 3. Последняя очень редко встречается
213 между соседними трикодонами в НА вирусов гриппа. Ее частота, варьируя у
214 НА в диапазоне 1.6-2.1%, на 1% ниже чем у генов человека, т.е. возможности
215 модуляции частоты CpG при рекодировании, используемой, как правило, с
216 целью стабилизации мРНК, у гена H1 НА, по сравнению с H3 НА, весьма
217 ограничены.

218 По использованной нами модели считывания триплетов в мРНК
219 каждый кодон, за исключением первых и последних трех кодонов,
220 включается, как и при трансляции мРНК на рибосомах, в три последовательно
221 считываемые трикодона, и очевидно, что ИК каждого кодона в трикодоне
222 связан с ИК двух предшествующих и двух последующих кодонов. Это
223 подводит к признанию существования в генах континуума связности
224 кодонов по значениям их ИК и ограничения на их мутагенез. Свидетельством
225 связности кодонов служит иллюстрируемая на рис.7, на примере генов НА
226 штаммов A/Brevig Mission/1/18 и A/California/04/2009, высокая частота и
227 линейное распределение СК между теми кодонами, с которыми связана
228 замена аминокислот. Как метко подмечено Комаром А.А., «синонимические
229 кодоны размещены в мРНК стратегически (в определенных местах) таким

230 образом, что они задают специфическую кинетику трансляции,
231 обеспечивающую, в свою очередь, эффективное поэтапное котрансляционное
232 сворачивание белка», т.е. являются вторичным кодом для сворачивания белка
233 [8].

234 У штаммов вируса гриппа одного и того же подтипа, как отмечалось
235 выше, и эпидсезона различия НА связаны не только с заменами аминокислот,
236 но и с еще большим различием в числе и распределении СК. Последние служат
237 причиной образования у каждого штамма вируса гриппа своего конформера
238 НА и своей мозаики обнажения и доступности в нем эпитопов. Социркуляция
239 в эпидсезоне штаммов вируса гриппа с разными первичными структурами НА
240 и составами СК в их генах и отличие от них по этим же показателям
241 вакцинного штамма являются (следует признать) другим фактором,
242 обрекающим противогриппозные вакцины любой конструкции на невысокую
243 эффективность. В США, например, эффективность вакцины против гриппа на
244 субоптимальном уровне – 14–60%, а иммунитет к гриппу сохраняется лишь
245 несколько месяцев [11].

246 Если таблица трансляционного кода (рис.2) служит обобщающей
247 информацией об использовании СК в мРНК , то двумерная комбинированная
248 таблица (рис. 4) служит источником информации по дипептидам в белке НА
249 и соответствующим им дикодонам в мРНК и полезна особенно тем , что в ней
250 видно, какие СК сочетаются (или не сочетаются) друг с другом и как часто.
251 Избирательность в соседствовании между кодонами служит, наряду с
252 ограничениями в разнице по ИК между соседними кодонами, вторым
253 барьером, который нужно обходить при рекодирования генов. Третьим же
254 барьером являются редко встречающиеся и не обнаруженные в
255 исследованных нами генах человека дикодоны, перечисленные на рис.5 . Эти
256 дикодоны можно рассматривать как запретные для использования при
257 рекодировании.

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

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

279 Приемлемым объяснением исключения квартета кодонов CGC, CGT,
280 CGA и CGG из гена H1 NA может быть связано с особенностями структуры
281 самого белка NA, в частности ограничениями в его инвариантной
282 последовательности. Так следствием запрета в трансляционном коде гена H1
283 NA на кодоны CGC, CGT, CGA и CGG будут соответственно ограничения
284 на замены по 8 аминокислотам (см. рис. 2): глицину, триптофану, цистеину,
285 серину (в случае кодирования его кодонами с корнем AG), пролину, лейцину
286 (в случае кодирования его кодонами с корнем CT), гистидину и глутамину.

287 Ограничения на цистеин, триптофан, глицин и лейцин связаны с присутствием
288 их в инвариантных позициях первичной структуры белка Н1, и запрет на
289 использование в коде гена Н1 всего квартета триплетов аргинина с корнем СG
290 пресекает возможности мутирования кодонов этих аминокислот

291 Выявление скрытой роли СК в невысокой эффективности
292 противогриппозных вакцин еще более убеждает в том, что возможность
293 успешного использования принципа «один иммуноген для всех штаммов
294 подтипа вируса» маловероятна. Один из путей решения проблемы с
295 увеличением эффективности противогриппозных вакцин применительно к
296 мРНК вакцин мог бы заключаться в привлечении против штаммов одного
297 подтипа 2-3 иммуногенов, т.е. 2-3 мРНК, охватывающих более широкий
298 спектр циркулирующих в эпидсезоне штаммов соответствующего подтипа,
299 опираясь на представление о доминантных штаммах подтипа, упоминаемых
300 в планах США и ВОЗ по поиску универсальных противогриппозных вакцин
301 [4,5]. Конструирование соответствующих им мРНК могло бы основываться
302 на том же принципе (но более расширенном), что был предложен нами ранее
303 для построения доминантной белковой структуры НА одного или нескольких
304 эпидсезонов, в зависимости от целевой установки [1].

305 Приведем с помощью рисунка 9 краткое изложение процедуры
306 построения фрагмента аминокислотной последовательности доминантного
307 Н3 НА эпидсезона 2018-2019 гг., которая легко трансформируется в
308 доминантную мРНК заменой аминокислот соответствующими кодонами. На
309 рисунке 9 первый вертикальный числовой ряд – нумерация позиций
310 аминокислот в НА, второй вертикальный ряд аминокислот – собственно сама
311 построенная доминантная последовательность НА эпидсезона, параллельно
312 ей, правее – «скелет» молекулы с расположением инвариантных
313 (представлены обозначениями аминокислот) и переменных (представлены
314 точками) сайтов. За последними следует горизонтальный ряд всех
315 замещающих аминокислот, обнаруженных в этих позициях НА у разных

316 штаммов эпидсезона, с указанием (в скобках) для каждой аминокислоты
317 частоты ее встречаемости. Строилась доминантная последовательность НА
318 вставлением в цепь с инвариантными блоками доминантной аминокислоты из
319 приводимого горизонтального ряда заменяемых аминокислот [1].

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

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

ТАБЛИЦЫ

Таблица 1. Различия трикодонов, считываемых со сдвигом на один шаг, по их ИК.

Table 1. 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 разниц последовательных трикодонов по ИК Frequency of occurrence of the difference of successive tricodon 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

РИСУНКИ

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.

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	GAA	E	7
GCG	A	9	GGG	G	9	GTG	V	8	GAG	E	8

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.

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 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.

H1N1 A/Brevig Mission/1/1918					H1N1 A/California/04/2009				
TCT	S	7	TGT	C 10	TCT	S	8	TGT	C 9
TCC	S	6	TGC	C 6	TCC	S	4	TGC	C 6
TCA	S	16	TGA	Z 1	TCA	S	18	TGA	Z 0
TCG	S	3	TGG	W 11	TCG	S	1	TGG	W 10
TTT	F	9	TTA	L 11	TTT	F	8	TTA	L 5
TAT	Y	17	TAA	Z 0	TAT	Y	13	TAA	Z 1
TAC	Y	9	TAG	Z 0	TAC	Y	14	TAG	Z 0
ACT	T	7	ATT	I 10	ACT	T	9	ATT	I 18
ACC	T	8	AAT	N 26	ACC	T	2	AAT	N 27
ACA	T	20	AAC	N 16	ACA	T	23	AAC	N 14
ACG	T	2	AAA	K 21	ACA	T	3	AAA	K 27
AGT	S	6	AAG	K 12	ACG	T	3	AAG	K 15
AGC	S	11	ATG	M 8	AGT	S	7	ATG	M 7
AGR	12	ATG	M 8	AGC	S	9	ATG	M 7	
ATA	I	14	AAG	K 12	AGR	12	AAG	K 15	
AAA	K	21	AAG	K 12	ATA	I	12	AAG	K 15
AAT	N	26	AAG	K 12	AAA	K	27	AAG	K 15
AAC	N	16	AAG	K 12	AAT	N	27	AAG	K 15
AAA	K	21	AAG	K 12	AAC	N	14	AAG	K 15
AAG	K	12	AAG	K 12	AAA	K	27	AAG	K 15
CCT	P	2	AAG	K 12	AAT	N	27	AAG	K 15
CCG	P	3	AAG	K 12	AAC	N	14	AAG	K 15
CGT	R	0	AAG	K 12	AAA	K	27	AAG	K 15
CGC	R	0	AAG	K 12	AAT	N	27	AAG	K 15
CGA	R	0	AAG	K 12	AAC	N	14	AAG	K 15
CTT	L	2	AAG	K 12	AAA	K	27	AAG	K 15
CTC	L	6	AAG	K 12	AAT	N	27	AAG	K 15
CTA	L	9	AAG	K 12	AAC	N	14	AAG	K 15
CTG	L	14	AAG	K 12	AAA	K	27	AAG	K 15
CAT	H	10	AAG	K 12	AAT	N	27	AAG	K 15
CAC	H	3	AAG	K 12	AAC	N	14	AAG	K 15
CAQ	10	AAG	K 12	AAA	K	27	AAG	K 15	
CAA	Q	10	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAC	N	14	AAG	K 15
CAG	Q	7	AAG	K 12	AAA	K	27	AAG	K 15
CAG	Q	7	AAG	K 12	AAT	N	27	AAG</	

ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА
 FEATURES OF HEMAGGLUTININ GENES OF INFLUENZA VIRUSES 10.15789/2220-7619-FOA-17578

K 37.AAG	1		2		1	1		1		2	1		1
1													
P 38.CCT				1				1					
P 39.CCC													
P 40.CCA	1		1		1					1	1	1	1
1										2	1		1
P 41.CCG												1	
2													
H 42.CAT												1	1
1												2	1
1												1	1
H 43.CAC	1									1	1		
1													
Q 44.CAA	1				1	1	2			1			
1										1	1		
Q 45.CAG			1		1					1			
1													
A 46.GCT					1			2		1			
1													
A 47.GCC			1							2	1		
1													
A 48.GCA			1	2	1			1	1	1		1	1
1													
1										2		1	2
1													
A 49.GCG			1					1		1			
1													
G 50.GGT	1			1		1	1	1		1			1
1													
G 51.GGC												1	1
1													
G 52.GGA	1	1	1	1	3	1		1		1	1	2	1
1													
1													
G 53.GGG	1	1						1		1		2	
1													
V 54.GTT								1		1			1
1													
V 55.GTC	1	1			1	1		1		1			1
1													
V 56.GTA					1	1				1		2	1
1													
V 57.GTG		1				1		1	1	1			1
1													
D 58.GAT			1		2	1		1		1		3	1
1													
D 59.GAC		1						1	1	3			1
1													
E 60.GAA		2	1		1	1		1		2	2		1
1													
1													
E 61.GAG	1	3			1	1				1	2		1

Примечания: Первый вертикальный ряд и первый горизонтальный ряд сверху - обозначения аминокислот; второй вертикальный ряд и второй горизонтальный ряд сверху – нумерация кодонов; третий вертикальный ряд – кодоны.

Notes: The first vertical row and the first horizontal row from above are amino acid designations; the second vertical row and the second horizontal row from above are codon numbering; the third vertical row are codons.

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.

А	Б
TCC(S)-GGT(G)	TCG(S)-TCA(S)
TCG(S)-TCC(S)	TCG(S)-TCG(S)
TCG(S)-CTC(L)	TCG(S)-TAC(Y)
TCG(S)-ACA(T)	TCG(S)-CGT(R)
TCG(S)-AGA(R)	TCG(S)-CGC(R)
TCG(S)-ATT(I)	TCG(S)-ATA(I)
TCG(S)-GGT(G)	TCG(S)-CAA(Q)
AGT(S)-CAT(H)	TGC(C)-GCG(A)
AGC(S)-GTA(V)	CTC(L)-GCG(A)
TGT(C)-GCG(A)	CTA(L)-TCG(S)
TGC(C)-GGA(G)	CTA(L)-CGG(R)
TTA(L)-CGC(R)	CTA(L)-CCG(P)
CTC(L)-GTA(V)	CTA(L)-GCG(A)
CTC(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 6. Последовательность значений индексов комплементарности HA штамма H1N1 A/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.

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

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.

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 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.

N	A	D	T	I	C	I	G	Y
AAT	GCA	GAC	ACA	ATA	TGT	ATA	GGC	TAC
6	8	8	7	6	7	6	9	7
22	23	21	20	19	22	22	23	23

Примечания: 1-ый ряд букв – последовательность аминокислот, 2-ой ряд – обозначения кодонов, 3-ий ряд - обозначения индексов комплементарности кодонов, 4-ый ряд – обозначения индексов комплементарности трикодонов.

Notes: 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.

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

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

```
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
```

ТИТУЛЬНЫЙ ЛИСТ_МЕТАДААННЫЕ

Блок 1. Информация об авторе ответственном за переписку

Харченко Евгений Петрович – д. б. н., ведущий научный сотрудник

Института эволюционной физиологии и биохимии им. И. М. Сеченова РАН;

адрес: 194223, Санкт-Петербург, пр. Тореза, 44;

телефон: 8(904)338-22-80;

e-mail: neuro.children@mail.ru.

Kharchenko Eugene P. – Dr. Sci. (Biol.), leader researcher Sechenov Institute of
Evolutionary Physiology and Biochemistry, 44;

address: Toreza pr., St. Petersburg, Russian Federation, 194223;

telephone: 8(904)338-22-80;

e-mail: neuro.children@mail.ru.

Блок 2. Метаданные статьи

ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА И
ВОЗМОЖНОСТИ ИХ РЕКОДИРОВАНИЯ

FEATURES OF HEMAGGLUTININ GENES OF INFLUENZA VIRUSES AND
THE POSSIBILITY OF THEIR RECODING

Сокращенное название статьи для верхнего колонтитула:

ОСОБЕННОСТИ ГЕНОВ ГЕМАГГЛЮТИНИНА ВИРУСОВ ГРИППА

FEATURES OF HEMAGGLUTININ GENES OF INFLUENZA VIRUSES

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

Keywords: influenza virus, hemagglutinin gene, structural features, recoding, mRNA, vaccine.

Оригинальные статьи.

Количество страниц текста – 13, количество таблиц – 1, количество рисунков – 9.

27.01.2024

СПИСОК ЛИТЕРАТУРЫ

1	Харченко Е. П. Поиски универсальной противогриппозной вакцины: возможности и ограничения. Эпидемиология и Вакцинопрофилактика. 2019; Т. 18 ,No 5 , С .70–84.	[.Kharchenko EP. The Search for a Universal Influenza Vaccine: Possibilities and Limitations. Epidemiology and Vaccinal Prevention. 2019; vol. 18 ,. no. 5 : 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 , No. 6, С.183-200 doi. 10.31631/2073-3046-2023-22-6-183-200	Kharchenko E.P. Problems and Collisions of Vaccinology. Epidemiology and Vaccinal 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: 1127-1132.	doi: 10.1126/science.abn1755.

4	Erbelding EJ., Post DJ. Stemmy EJ. et al. Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases. Journal of Infectious Diseases 2018. Vol. 218 :. 347–354.	doi: 10.1093/infdis/jiy103.
5	Global influenza strategy 2019 -2030. World Health Organization. https://www.who.int/publications/i/item/9789241515320 15 March 2019	https://www.who.int/publications/i/item/9789241515320
6	Gonçalves-Carneiro D, Bieniasz P.D.. Mechanisms of attenuation by genetic recoding of viruses. mBio, 2021, vol. 12 : e02238-20.	doi: 10.1128/mBio.02238-20.
7	Knyazev S. , Chhugani K. , Sarwal V. , Ayyala R. et al. Unlocking capacities of genomics for the COVID-19 response and future pandemics. Nat Methods, 2022, vol. 19, no. 4 :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. Vol 2019, vol. 53, no. 6 : 777-790;	doi:10.1134/S0026893319060098
9		Lin B. C., Kaissarian N.M., Kimchi-Sarfaty C. Implementing computational methods in tandem with synonymous gene recoding for therapeutic development. Trends in Pharmacological Sciences, 2023, Vol. 44, no. 2	doi: 10.1016/j.tips.2022.09.008
10		10. Martínez M. A., Jordan-Paiz A., Franco S., Nevot M. Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness. Trends in Microbiology, 2015	doi: 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. 31:146-157.	doi: 10.1016/j.chom.2022.11.016
12	Ostrov N ., Nyerges A., Chiappino-Pepe A . Rudolph A. et al. Synthetic genomes with altered genetic codes. Current Opinion in Systems Biology , 2020, vol. 24:32–40.	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, vol. 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, vol. 13: 896958.	doi: 10.3389/fimmu.2022.896 958
--	--	---	---------------------------------------