

УДК 578.834.1:578.7]:615.371

DOI: 10.14427/jipai.2021.3.45

Поиск безопасных иммунодоминантных пептидов SARS COV-2 in silico

А.М. Цыганков, В.В. Янченко

Витебский государственный медицинский университет

Searching for safe SARS COV-2 immunodominant peptides in silico

A.M. Tsygankov, V.V. Yanchenko

Vitebsk State Medical University, Belarus

Аннотация

Молекулы HLA функционируют для контроля постоянства внутренней среды организма. HLA I класса экспрессируются на поверхности всех ядродержащих клеток и представляют пептиды всех белков, которые синтезируются в каждой конкретной клетке. При заражении клетки вирусом пептиды его белков связываются с HLA I класса и информируют систему иммунитета о наличии вируса в этой клетке. В качестве актуального примера выбран вирус SARS COV-2. В данной публикации рассматривается алгоритм выбора иммунодоминантных пептидов вируса SARS COV-2, которые можно использовать для определения иммунореактивности организма, иммунопрофилактики и иммунотерапии. Проанализированы белки исходного штамма hCoV-19/Wuhan/WIV04/2019 (далее – WIV04), изолята MW674675.1. из Республики Беларусь (далее – MW674675.1.), штамма B.1.617.2 из Индии (далее – дельта или MZ314977.1.), и вызывающий беспокойство вариант B.1.1.529 (далее – омикрон или OL677199.1). С помощью различных программ на платформах открытого доступа был осуществлен поиск иммунодоминантных пептидов из протеома SARS COV-2: NCBI – GenBank, Protein; IEDB – MHC-I Processing Predictions, Population Coverage; GISAID – CoVsurver: Mutation Analysis of hCoV-19, UNIPROT – BLAST; ToxinPred. В результате поисковой работы были выбраны 27 иммунодоминантных пептидов – мишеней для цитотоксических Т-лимфоцитов посредством представления через HLA-A *02:01. Мы надеемся, что данная информация будет способствовать распространению и реализации идей, направленных на наиболее безопасную профилактику вирусных инфекций. При прогнозировании *in silico* основ для пептидных вакцин можно скорейшим и эффективным образом подойти к практической составляющей, что приведет к персонализированной вакцинации людей как чрезвычайно эффективной мере.

Ключевые слова

HLA, SARS COV-2, пептиды, иммунитет, вакцина.

Summary

HLA molecules function to control the constancy of the internal environment of the body. HLA class I molecules are expressed on the surface of all nucleated cells and represent the peptides of all proteins that are synthesized in each individual cell. When a cell is infected with a virus, the peptides of its proteins bind to HLA class I and inform the immune system of the presence of the virus in that cell. The SARS COV-2 virus is chosen as a relevant example. This publication discusses an algorithm for selecting immunodominant peptides of SARS COV-2 virus that can be used to determine body immunoreactivity, immunoprophylaxis and immunotherapy. Proteins from the original strain hCoV-19/Wuhan/WIV04/2019 (hereafter WIV04), the isolate MW674675.1. from RB (hereafter MW674675.1.), strain B.1.617.2 from India (hereafter delta or MZ314977.1.), and the worrisome variant B.1.1.529 (hereafter omicron or OL677199.1) were analyzed. Immunodominant peptides from the SARS COV-2 proteome were searched using various programs on open access platforms: NCBI – GenBank, Protein; IEDB – MHC-I Processing Predictions, Population Coverage; GISAID – CoVsurver: Mutation Analysis of hCoV-19, UNIPROT – BLAST; ToxinPred. The search work resulted in the selection of 27 immunodominant peptides – targets for cytotoxic T-lymphocytes by representation through HLA-A *02:01. We hope that this information will contribute to the dissemination and implementation of ideas aimed at the safest prevention of infections. In predicting *in silico* bases for peptide vaccines, the practical component can be approached in an early and efficient manner, leading to personalized vaccination of humans as an extremely effective measure.

Keywords

HLA, SARS COV-2, peptides, immunity, vaccine.

антигенов происходит через вакуолярный путь или через путь от эндосомы к цитозолю. В вакуолярном пути антигены расщепляются в эндосомах катепсином S и загружаются на HLA молекулу. На втором пути антигены транспортируются в цитозоль для протеасомной деградации. После этого пептиды транспортируются обратно в эндосомы или в эндоплазматический ретикулум (ЭР). Там они обрезаются и загружаются на HLA I. Пептид-связывающая бороздка HLA имеет размер около 2,5 нм в длину и 1 нм в ширину. Она содержит разные аминокислотные остатки, что позволяет связывать и удерживать широкий спектр пептидов антигенов. Важно отметить, что эта бороздка закрыта с обоих концов, что ограничивает размер представляемого пептида, обычно это 8–12 аминокислотных остатков, в зависимости от аллеля HLA I класса. Аминокислотные остатки пептида связываются с карманами бороздки и «заякориваются». Первичные якоря обычно расположены во втором положении и на С-конце пептида, в то время как положение вторичных якорей менее ограничено и зависит от аллеля HLA I класса. Например, пептиды с лейцином во втором положении и валином или лейцином в девятом положении имеют высокое сродство к HLA-A*02, гидрофобный С-концевой участок присутствует во всех пептидах HLA I класса. Молекулы HLA I класса также могут представлять и более длинные пептиды (до 25 аминокислотных остатков) из-за выступа слабых аффинных участков пептидной цепи и сохранения положений якорного аминокислотного остатка [12]. Чтобы получить подобную информацию о пептидах, связанных с пептид-связывающей бороздкой, в мире используются современные методы работы с этими структурами: иммуноаффинная хроматография, мягкая кислотная элюция [13-14].

Восприимчивость человека к SARS COV-2 была доказана для носителей аллелей HLA-A*25:01, HLA-A*26; HLA-B*46:01, HLA-B*27:07; HLA-C*01:02; HLA-C*05; DRB1*15:01 и DRB1*07:01; DQB1*06:02 [15-20]. С учетом того, что штаммы вируса различаются по своим протеомным характеристикам, для сравнения был выбран штамм B.1.617.2 или дельта как наиболее распространенный на момент написания статьи [21]. Варианты дельта подразделяются в системе обозначения линий Pango на варианты AY (их обозначение не подразумевает каких-либо функциональных биологических отличий) [22]. Для анализа был выбран штамм MZ314977.1 из базы открытого доступа GenBank. Также в

работе проанализирован вариант – омикрон, происходящий из Южной Африки, вызывающий беспокойство – OL677199.1, выделенный исследователями Канады.

Материалы и методы

Применяли пошаговый алгоритм поиска безопасных иммунодоминантных пептидов SARS COV-2 для конкретного аллеля HLA I класса:

- Шаг первый: выбор изолятов вируса SARS COV-2 – базы данных GenBank, Protein (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=SARS-CoV-2,%20taxid:2697049).
 - Шаг второй: анализ мутаций избранных вирусов – CoVsurver: Mutation Analysis of hCoV-19 (<https://www.gisaid.org/epiflu-applications/covsurver-mutations-app>).
 - Шаг третий: анализ аффинности всех пептидов из протеома каждого вируса – MHC I Processing Predictions (версии 2013-02-22) [23-26].
 - Шаг четвертый: проверка in silico на токсичность – ToxinPred [27, 28].
 - Шаг пятый: проверка in silico на гомологию человеческим белкам – BLAST (protein-protein) [29] для исключения возможных аутоиммунных реакций.
 - Шаг шестой: анализ аффинности всех мимикричных пептидов (белков человека) длиной 14 аминокислотных остатков (далее – а.о.) – MHC I Processing Predictions (версии 2013-02-22) для исключения возможных аутоиммунных реакций.
 - Шаг седьмой: прогноз популяционного реагирования на отобранные иммунодоминантные пептиды – Population Coverage (IEDB) [24].
- Результаты и обсуждение

Первым шагом был выбор 4 вариантов вируса SARS COV-2 в базе данных GenBank на платформе NCBI: референс WIV04, единственный изолят (Республика Беларусь) MW674675.1., дельта вариант MZ314977.1. (Индия) и OL677199.1 (Канада). Пример «severe acute respiratory syndrome coronavirus 2 isolate WIV04, complete genome». Для анализа белков вирусов загружали их аминокислотные последовательности в формате FASTA: 1 ATTAAAGGTT... 298891 ... AAAAAAAAAA.

Вторым шагом был анализ мутаций с помощью программы CoVsurver: Mutation Analysis of hCoV-19 и аминокислотных последовательностей из файлов MW674675.1., MZ314977.1., OL677199.1 в сравнении с WIV04 (табл. 1).

За общее происхождение вирусов может говорить общая мутация P323L в NSP12. Обращает на себя внимание количественный и качественный состав мутаций в S белке (рис 2.). Меньше всего у MW674675.1., больше всего мутаций у OL677199.1. Что, естественно, отражается на том, какие пептиды могут быть представлены вместе с HLA-A молекулами, а какие нет.

Третьим шагом был анализ аффинности *in silico* всех возможных вариантов вирусных пептидов с помощью программы MHC I Processing Predictions при использовании кода а.о. в формате FASTA, пример:

```
>QHR63266.1 nonstructural protein NS7b
[Severe acute respiratory syndrome coronavirus 2]
MIELSLIDFYLCFLAFLFLVLMIMLIIFWFSLEL
QDHNETCHA.
```

Применялись следующие условия поиска иммунодоминантных пептидов: формат последовательности – рекомендованный IEDB; MHC человека – HLA-A *02:01; длина а.о. пептида – 9; иммунопротеасома; максимальное расширение предшественника – 1; альфа-фактор – 0,2; формат вывода – таблица XHTML и IC50 < 50 нмоль. Так был найден 61 пептид, отвечающий нашим требованиям (табл. 2).

На четвертом шаге все пептиды проверялись *in silico* на токсичность с помощью программы ToxinPred. Среди выбранных пептидов не было обнаружено ни одного токсичного пептида по результатам компьютерного моделирования.

На пятом шаге пептиды (за исключением 7 уникальных из T-Scan) были проверены на гомологию с белками человека с помощью программы BLAST (protein-protein): база данных – RefSeq Genome Database, Homo Sapiens, optimization – megablast, степень схожести а.о. и охват ≥80%.

Шестой шаг подразумевал использование MHC I Processing Predictions для анализа аффинности всех пептидов (белков человека) длиной 14 а.о. Чтобы получить пептид длиной 14 а.о. из последовательности гомологичного человеческого белка, нужно выбрать последовательность а.о. как перед, так и после схожего фрагмента. Например, для пептида ALSKGFHFV позиция А соответствует №1, а позиция V соответствует №9; жирным шрифтом выделена сходная последовательность **LSEGFHF** /848-854/ из белка KICSTOR complex protein SZT2 isoform. Значит, необходимо, чтобы позиция №1 была меньше на 1, т.е. 847 (R). Достраиваем позиции №9-14 и получаем ACSGEG. В итоге получаем последовательность в виде /847-860/.

К особенности окончательного исключения из списка безопасных иммунодоминантных пептидов исходного вирусного пептида мы отнесли утверждение: если хотя бы один мимикричный пептидов белка человека из проанализированного фрагмента демонстрировал IC 50 (концентрацию полумаксимального ингибирования) < 50 нМ, то он считался опасным в плане возможного развития аутоиммунных реакций на клетки человека, несущие такой белок.

В результате анализа установлено, что только 7 из отобранных пептидов вируса не гомологичны белкам человека, а для 54 вирусных пептидов найдены 107 белков человека со схожей аминокислотной последовательностью (выделено жирным шрифтом в табл. 3). Так, 27 пептидов (выделены курсивом в табл. 3) оказались годны для дальнейших исследований, причем 13 пептидов принадлежали OL677199.1. Из 27 пептидов к ORF1ab принадлежало 8 пептидов, 7 пептидов – S белку, 3 пептида – ORF10, по 2 пептида – оболочечному, мембранному белку и ORF3а, по 1 пептиду – нуклеопротеину, ORF7а, ORF6.

Для 4 вариантов SARS COV-2 были найдены 2 пептида мимикричных пептидам белков CUB and sushi domain-containing protein и KICSTOR complex protein SZT2 isoform X1. Из таблицы видно, что пептид **RLSEGFHFACSGEG** по причине прочного связывания (IC=16,8 нМ) и пептид **TSVGPITAKGFHFV** с IC 63,5 нМ по причине высокой аффинности к HLA-A *02:01, установленной в результате проверки, нецелесообразно использовать в составе вакцин. Гомологичные человеческим вирусные пептиды могут быть триггерами развития аутоиммунных реакций, поэтому только негомологичные пептиды подлежат дальнейшим исследованиям. Так, например, в составе комплекса KICSTOR функционирует в ветви сигнального пути TORC1. Привлекает, независимым от аминокислот образом, комплекс GATOR1 к лизосомальным мембранам и обеспечивает его взаимодействие с GATOR2 и RAG ГТФаз. Функционирует перед RAG ГТФазой и необходим для негативной регуляции сигнализации mTORC1 в отсутствие аминокислот. В отсутствие комплекса KICSTOR mTORC1 конститутивно локализуется в лизосоме и активируется. Комплекс KICSTOR также, вероятно, участвует в регуляции mTORC1 глюкозой [33]. CUB and sushi domain-containing protein участвует в развитии дендритов нейронов. Слабо экспрессируется в большинстве тканей, за исключением мозга (в том числе плода): мозжечок, substantia nigra, таламус, спинной мозг, гиппокамп, также экспрессируется в яичках

Таблица 1. Мутации SARS COV-2

Белок	MW674675.1.	Общие мутации	MZ314977.1	Общие мутации	OL677199.1
Spike	T259I D1118Y	D614G	S982A, A570D, T716I, D1118H	D614G N501Y P681H Y144, H69, V70 del	A67V, V70del, T95I, G142D, V143del, Y145del, N211del, L212I, ins214E, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, Y505H, T547K, , H655Y, N679K, N764K, D796Y, N856K, Q954H, N969K, L981F
NSP1	S142del, R175C, F143del, K141del: <i>только для MW674675.1.</i>				
NSP2	G465S, K312R, T170I: <i>только для MW674675.1.</i>				
NSP3	-	-	A890D, T749A, T183I, I1412T	-	K38R, S1265del, L1266I, A1892
NSP4	-				T492I
NSP5	-				P132H
NSP6	-	-	F108 del	S106del, G107del	L105del I189V
NSP8	T141M				
NSP12	P323L				
NSP13	-	-	K460R	-	-
N	S194L, G137R	-	S235F D3L	R203K G204R	P13L, E31del, R32del, S33del
NS3	V77F, S60F				
NS7a	V93F				
NS7b	L6M				
NS8	-	-	Y73C, Q27stop, Q24R, R52I	-	-
E	-				T9I

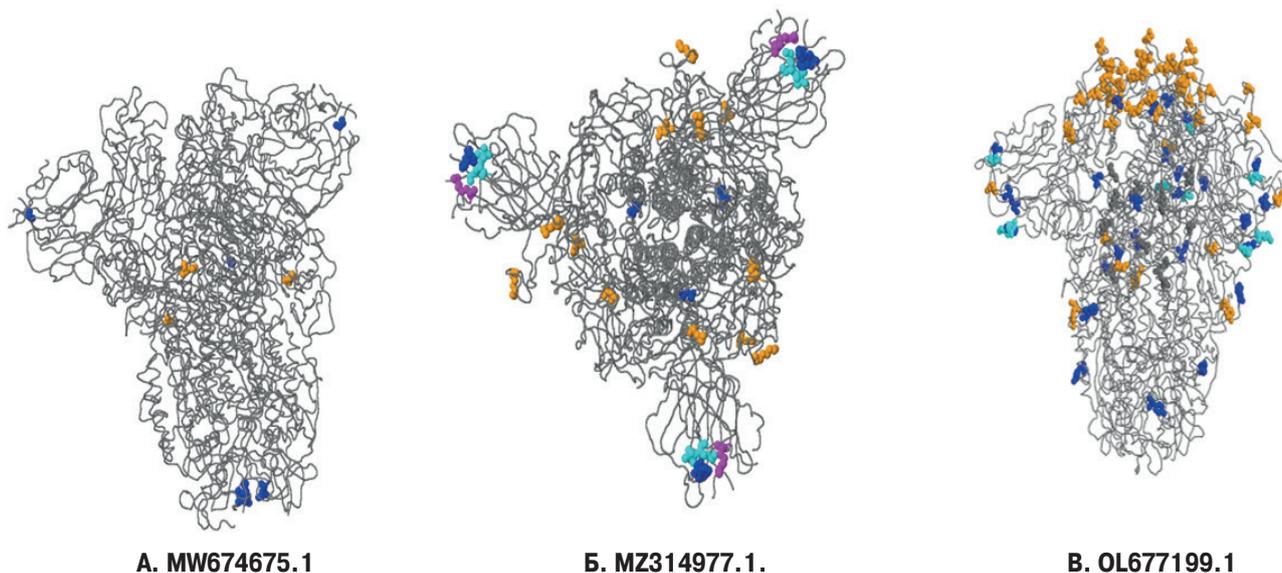


Рис. 2. S белки SARS COV-2 с выделением мест мутаций в сравнении с WIV04

Таблица 2. Пептиды с высоким аффинитетом

Источник	Уникальные последовательности (позиция в белке)	Идентичные пептиды
WIV04 ¹	-	-
MW 674675.1. ²	-	-
MZ 314977.1. ³	ORF1ab polyprotein: YTMADLVYA (4514-4522) YLTNDVSFL (3115-3123)	Spike protein (998-1006) ³ RLQSL
T-Scan [30]	KLPDDFTGCVKLAQCQVL, RLDKVEAEV, FLPGVYSV, FLYENAFLLP, WLDMDVDTSL, YLITPVHV	QTYV (997-1005) ⁴
OL 677199.1 ⁴	26 уникальных пептидов ниже	ORF3a protein LLYDANYFL (139-147) ¹²³⁴ , YLYALVYFL (97-105) ²⁴ ; (107-115) ¹³ , ALSKGF- HFV (72-80) ²⁴ ; Membrane glycoprotein GLMWL- SYFI (89-97) ¹²³⁴ ; FLWLLWPV (53-60) ²⁴ ; Spike protein YLQPRTFLL (278-286) ² ; (269-277) ¹³ ; (266-274) ⁴ ; ORF7b protein FLAFLFLV (8-16) ² ; (13-21) ¹³ ; Enve- lope protein FLAFVVFL (20-28) ¹²³⁴ ; ORF1ab poly- protein: FLAHIQWMV (3122-3130) ¹³ ; (3118-3126) ² ; ILFTRFFYV (2332-2340) ¹³ , (2329-2337) ² ; FLLNKE- MYL (3183-3191) ¹³ , (3179-3187) ² ; YLDAYNMMI (6419-6427) ¹³ , (6416-6424) ² ; TMADLVYAL (4515- 4523) ¹³ , (4512-4520) ² ; YLNTLTLAV (6852-6860) ¹³ , (6849-6857) ² ; TLMNVTLV (3710-3719) ¹³ , (3707-3715) ² ; LLLDDFVEI (6749-6757) ¹³ , (6746-6754) ² ; FVNEFY- AYL (5133-5141) ¹³ , (5140-5148) ² ; FLPRVFSV (2884- 2892) ¹³ , (2880-2888) ² ; VLWAHGFEL (6109-6117) ¹³ , (6106-6114) ² SMWALIISV (3732-3740) ¹³ , (3729-3737) ² ; FLLPSLATV (3639-3647) ¹³ , (3636-3644) ² ; NLIDSY- FVV (4456-4464) ¹³ , (4453-4461) ² ; ALLADKFPV (6245- 6253) ¹³ , (6242-6250) ² ; FLNRFTTTL (3481-3489) ¹³ , (3478-3486) ² ; MMISAGFSL (6425-6433) ¹³ , (6422-6430) ² ; RIMTWLDMV (3662-3670) ¹³ , (3659-3667) ² ; FLAR- GIVFM (3743-3751) ¹³ , (3740-3748) ² ; YLATALTL (1675- 1683) ¹³ , (1672-1680) ² ; MLDMYSVML (5291-5299) ¹³ , (5288-5296) ² ; KLIEYTDEFA (2901-2909) ¹³ , (2898-2906) ² ; VLAWLYAAV (3467-3475) ¹³ , (3464-3472) ² ; Nucleopro- tein: LLLDRLNQL (222-230) ¹ ; (218-226) ⁴ .
ORF10: NVFAFPFTI (5-13); Nucleoprotein FGMSRIGMEV (312-320); ORF8: YIDIGNYTV (73-81), FLEYHDVRV (108-116); ORF7a: KLFIRQEEV (85-93), FLIVAAIVFI (91-100); ORF6: FHLVDFQVTI (2-11); M: KLEEWNLV (15-23); FVLAAYRI (65-73); E: SLVKPSFYV (49-57), FLLVTLAIL (36-44); SVLLFLAFV (17-25); VLLFLAFVV (18-26); ORF3a: TVYSHLLLV (89-97); ALVYFLQSI (110-118); WLIVGVALL (45-53), ALLAVFQSA (51-59); Spike protein: FIA-GLIAIV (1217-1225), FQFCNDPFL (131-139), SIIAYTMSL (688-696), LFNKVTLA (818-826); VLNDIFSRL (973-981), VVFLHVITYV (1057-1065), FLHVITYVPA (1059-1067), FLPFFSNV (55-62), FTISVTTEI (715-723).		

[34]. Это могло бы объяснять поражение ЦНС, бесплодие, нарушение деления клеток.

Седьмым шагом было использование программы Population Coverage (IEDB) для прогнозирования взаимодействий HLA-A*02:01 с 27 подходящими пептидами из таблицы 3.

Выбор сделан для Европы и мира, МНС I класс.

Для аллеля HLA-A*02:01 предсказан высокий охват населения мира (39,08%) и Европы (47%), но невысокое количество безопасно распозна-

ваемых пептидных эпитопов вируса (менее 13%) (табл. 4), этим и обусловлено высокое распространение SARS COV-2 и высокое количество смертей от COVID-19.

Чем больше будет HLA аллелей задействовано для представления высокоаффинных эпитопов, тем лучше результат in vitro. На основании предложенного нам алгоритма необходимо подбирать безопасные иммунодоминантные пептиды для других часто встречающихся HLA аллелей для включения их в состав пептидных вакцин. Также

Таблица 3. Подбор пептидов (антигенная мимикрия)

Пептиды вирусов	Последовательность а.о. белка человека /позиция в белке/ белки человека	IC (нмоль)	Итог
LLDRLNQL	LLDRVNDLIEFRI /815-828/ dynein axonemal heavy chain 5	52.8	-
FLWLLWPV	FLWLKWPGRLLLR /525-538/ tripartite motif-containing protein 26	40	-
ALSKGFHFV	TSVGPITAKGFHFV /1853-1866/ CUB and sushi domain-containing protein RLSEGFHFACSGEG /847-860/ KICSTOR complex protein SZT2 isoform X1	63.5 16.8	-
KLIEYTDEFA	KLIGVYTDFAFRIT /495-508/ polynucleotide 5'-hydroxyl-kinase NOL9 KLLEYTDSNNIASLFLT /1248-1264/ ATP-binding cassette sub-family C member 8 isoform X1	66.1 50.6	-
FLLPSLATV	FLLPSLKDRQPKC /1822-1835/ ubiquitin carboxyl-terminal hydrolase 34 RQLLPSTATVRAAD /2178-2191/ protein bassoon QARPSLATVQTPSH /363-376/ nuclear receptor coactivator 6 isoform X1	1841.5 23.1 3940.3	-
LLYDANYFL	AISPWKTYQLVYFL 526-539/ aminopeptidase B isoform a LWGAILHSALIYFL /1053-1066/ protein sel-1 homolog 3 isoform 1	45.9 15.8	-
GLMWLSYFI	SLMWLSFRKVQEPQ /172-185/vitamin K epoxide reductase complex subunit 1 isof.	12.4	-
MMISAGFSL	MFISAGFNLSSSLT /198-211/ olfactory receptor 5M11 TGISAGFNLKESY /301-314/ prestin isoforms	1484 9.6	-
RIMTWLDMV	ELCTWLDKVEVELL /5929-5942/ dystonin isoform X4	23.5	-
SMWALIISV	GLWALIIMVFFKTY /373-386/ GPI-anchor transamidase precursor STWALIHSVSDAFS /117-130/ H(+)/Cl(-) exchange transporter 5 isoforms PPWALIAIAVVAGLL /61-75/ synaptotagmin-2	17.9 17.7 261.8	-
NLIDSYFVV	NLLDSYFVDLVIQK /133-146/ centromere protein O isoforms NLIDYFVPFLPLEY /261-274/ torsin-1A precursor	29.7 8.7	-
YTMADLVYA	YTMAALVIYLRFHN /128-141/ synaptophysin-like protein 2	7.6	-
YLATALTL	VLATVLLTLRAWRR /414-427/ protein delta homolog 2 isoform X1 ILATALLILCPFGL /199-212/ olfactory receptor 10C1 HPVTALLTLGPVKT /8588-8601/ mucin-16 GLAGALLTLLCMGL /8-21/ beta-1,3-N-acetylglucosaminyltransferase manic fringe YLAALLTVSEEVRS /2438-2451/ neurofibromin isoform 1	21.2 18.9 539.5 24.1 12.1	-
FLAFLFLV	FLLFLLFLVDIMAN /38-51/ myc target protein 1 isoform 1 FLAHLFLVAIDQT /580-593/ adhesion G protein-coupled receptor E2 isoforms	12.6 2.6	-
VLAWLYAAV	ELAWLYQEIRHHLW /290-303/ NACHT domain- and WD repeat-containing protein ALAWLYPEIGYCQG /190-203/ small G-protein signaling modulator isoforms	36.6 11.7	-
FLARGIVFM	FLSRGIVGTGSASY /140-153/ protein spinster homolog 3 isoform 1 KRARGNVFMYHAPE /2602-2615/ thyroglobulin precursor SLARDIVFEPYWMI /1008-1021/ transient receptor potential cation channel subfamily M member 6 isoform a ALARKGVQLGLVAF /30-43/ mitochondrial chaperone BCS1 isoforms FLERGESSQGEKEE /702-715/ probable guanine nucleotide exchange factor MCF2L2 isoform X1/4	51.2 13597 42.1 9531.6 21319	-
LLDDFVEI	IALLDDFVEITKKYA /242-255/ wolframin GLLDDFVEKNVLKL /24-37/ evolutionarily related interleukin-1beta converting enzyme LLLDDFFKSAVENC /389-402/ phosphogluconate dehydrogenase	507.3 45.5 15.7	-
ALLADKFPV	ARLAEKFPVISVPH /321-334/ Sorting nexin 18	11.2	-

FLNRFTTTL	FLNRFTQVFDPLT /1884-1897/ midasin FLNRFSTVCEEKLA /40-53/ WASH complex subunit 3 isoform 1	9.8 197.3	-
YLNTLT LAV	VLSTLT LAVFPTLY /131-144/ ALG10 protein, alpha2-glucosyltransferase SLNTLT LN VKSEKV /42-55/ lotillin-1 isoform 1	11.2 18.8	-
FLIVAAIVFI	FLIVAEIVLGLLVW /43-56/ CKLF-like MARVEL transmembrane domain-containing protein 8 isoform 1 FLIVPAIVGSALLH /37-50/ monocyte to macrophage differentiation factor	23.5 282.4	-
FLLVTLAIL	QLLVTLGILKLLGD /18-31/ mitochondrial inner membrane protease subunit 1 isoforms X2 LLLT LAILFASDA /18-331/ desmocollin-2 isoform Dsc2a preproprotein LLLVTLIILRKKHL /836-849/ receptor-type tyrosine-protein phosphatase O isoform	<50 <50 28.6	-
SVLLFLAFV	SVLLFLAWVCFLFY /5-18/ GPI ethanolamine phosphate transferase 3 isoforms TLLLFLAFLSSR /11-24/ ADAMTS-like protein 1 isoform 4 precursor	34.6 -	-
VLLFLAFVV	VLLFLAWVCFLFYA /6-19/ GPI ethanolamine phosphate transferase 3 isoforms	34.6	-
TVYSHLLL	TVYSHLFDHV VNRV /426-439/ unconventional myosin-VI isoforms	13.6	-
ALVYFLQSI	RLVFFLQSNRLQS /237-250/ kelch-like protein 26 isoform HLIYFLQRQDDPES /507-520/ cilia- and flagella-associated protein 54 isoform QLVYFLPSLPADLL /620-633/ testis-expressed protein 10 isoform 1,2	>50 >50 29.2	-
FLPFFSNV	FLPFFSNPETSDF /185-198/ OTU domain-containing protein 6A LLPFFNVLSQVFNS /2397-2410/ protein dopey-1 isoform a DLPFFNVSLLDWIN /300-313/ RB1-inducible coiled-coil protein 1 isoform 1 FLPFFSNLVTHLDSF /82-95/ kinetochore protein Nuf2	>50 >50 >50 48.1	-
ALLAVFQSA	ALLAAFQSESFKAG /1160-1173/ protein dopey-2 SLLATFQSAVAILG /74-87/ sodium-dependent multivitamin transporter	>50 11.5	-
FIAGLIAIV	VIGGLIAVVIFILL /1242-1255/ contactin-associated protein-like 4 isoforms	39.6	-
SIIAYTMSL	GHIADYTMSLLPNR /293-306/ N-acetylaspartylglutamate synthase A	7.2	-
LLFNKVT LA	LLFKVTLGLIRKLL /198-211/ exosome complex component RRP40 isoform 1 LLFYKVTEEPLEEE /4428-4441/ ryanodine receptor 3 isoform 1 LLFNKPTHYNNNGNY /363-376/ NT-3 growth factor receptor isoform a precursor DLFNKVKEHINSVS /1491-1505/ cytoplasmic dynein 1 heavy chain 1 KLFNKVFRCPMSFF /898-911/ ATP-binding cassette C member 11 isoform a LLFNKRKFFICVLH /947-960/ probable phospholipid-transporting ATPase IM LLFNKREFFICIAQ /957-970/ phospholipid-transporting ATPase ID isoform e	13.6 - - - - - -	-
VLWAHGFEL	GCWAHGFESLGVFT /107-120/ junctophilin-4 SLWAHGFQDRVASV /138-151/ beta-crystallin B3	4362 45.6	-
TLMNVLT LV	TLSDVLSRLVETVG /1951-1964/ protein furry homolog-like isoform X1 VLLNVLT LNRNLS /576-589/ glutathione hydrolase 7 isoform 1	30.3 448.9	-
NVFAFPFTI	FVFSFPFQTYLLD /84-97/ C-C chemokine receptor type 8 WVFSFPFCGPNEIN /162-175/ olfactory receptor 10A3 VGHDFPFTINNGTG /548-561/ cadherin EGF LAG seven-pass G-type receptor 2 SVFTFPFLPFSPMC /955-968/ zinc finger CCHC domain-containing protein 14	21.9 393.1 7247.8 275.7	-
YLDAYNMMI	нет	<50	+
FLAHIQWMV	нет	<50	+
YLYALVYFL	нет	<50	+
ILFTRFFYV	нет	<50	+
FLLNKEMYL	нет	<50	+
FQFCNDPFL ⁴	нет	<50	+
FLEYHDVRV ⁴	нет	<50	+
YIDIGNYTV ⁴	YIIGNYTSSLGSQV /189-202/ sushi domain-containing protein 1 isoforms	>50	+
KLFIRQEEV ⁴	ALFRRQEEDFLLPE /190-203/ zinc finger and SCAN domain-containing protein 5C SLWIRQEERAEIGA /449-462/ KRAB-A domain-containing protein 2	>50	+

Продолжение таблицы 3

<i>FHLVDFQVTI</i> ⁴	NERVDFEVTIPGEG /108-121/ protein argonaute-1 isoforms	>50	+
<i>KLLEEWNLV</i> ⁴	RLQEEWNLQEAERD /10-23/ DNA-dependent metalloprotease SPRTN isoforms	<50	+
	SYIEEWDLVLAASA /308-321/ nuclear pore complex protein Nup214 isoform 2		
<i>FVLAAYYRI</i> ⁴	FVIATVYRTRKLHT /69-82/ 5-hydroxytryptamine receptor 1B	<50	+
<i>SLVKPSFYV</i> ⁴	SLVKPSENGGVTF /4466-4479/ low-density lipoprotein receptor-related protein 2	<50	+
<i>WLIVGVALL</i> ⁴	WLIVGAALFKGIKS /247-260/ sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	>50	+
<i>VLNDIFSRL</i> ⁴	QLNDFFSRFAWAQRE /331-344/ BEN domain-containing protein 3	>50	+
	LLNDIFQRNCMKRK /320-333/ alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase-like protein MGAT4D isoform X1		
<i>VVFLHVITYV</i> ⁴	VVFPVHSYKHLRQM /780-793/ HEAT repeat-containing protein	>50	+
<i>FLHVITYVPA</i> ⁴	FLHVTVPRKIKSND /24-37/ disintegrin and metalloproteinase domain-containing protein 18 isoform 1 preproprotein	>50	+
<i>FTISVTTEI</i> ⁴	FAISVTTECFLLAT /94-107/ olfactory receptor 5H	>50	+
	ISVTTEEIRETVFS /33440 - 33453/ titin isoform IC		
	FAISVTVILILII /826-839/ leucine-rich repeat-containing protein 37B isoform 2		
	ISVSTEGANTSSST /156-169/ pro-neuregulin-1, membrane-bound isoform isoform HRG-alpha proprotein		
TMADLVYAL	NMADLIYKDKLQKTR /323-336/ disintegrin and metalloproteinase domain-containing protein 22 isoform	285.8	+
FVNEFYAYL	FLNEFYAHPRLQDY /434-447/ potassium channel subfamily T member 1 isoforms	70.1	+
RLQSLQTYV ³	LLQALQTYSDDILS /442-455/ LIM domain only protein 7 isoforms	1425	+
MLDMYSVML	FLDMFSVMSEMAPR /91-103/ calcium and integrin-binding family member 3 isoforms	176.9	+
	LSMYSVMKGRVATV /264-277/ olfactory receptor 1L4	162.9	
FLAFVVFL	PFLFVVFLLTIYFIS /23-36/ olfactory receptor 12D1	131.6	+
FLPRVFSVAV	EFPRVFSSVMKPEN /219-232/ coiled-coil domain-containing protein R3HCC1L isoforms	13973	+
	LQGEVFGAVLRQET /260-273/ antigen peptide transporter 1 isoform 1	376.7	
YLTNDVSFL ³	DLTNDVSSLDPTLP /486-499/ sorting nexin-19 isoform X1	15303.1	+
	YLTNDIGSSSLTNL /2036-2049/ ankyrin-3 isoform 1	809.7	
YLQPRTELL	NLQRTFLLTSSKE /334-347/ tax1-binding protein 1 isoform 1	1009.7	+
FGMSRIGMEV	KGMSRLGEEVNGEA /295-308/ exonuclease 3'-5' domain-containing protein 2 isof.	306.4	+

Примечание: «+» – пептид подходит для дальнейшего исследования; «-» – пептид не подходит для дальнейшего исследования; ³ – штамм вируса MZ 314977.1; ⁴ – штамм вируса OL 677199.1.

Таблица 4. Прогноз включения пептидов в пептид связывающую бороздку HLA-A*02:01

HLA-A*02:01	Мир	Европа
Прогнозируемое количество населения с анализируемым HLA-A	39,08%	47%
Среднее количество пептидных эпитопов вируса, распознаваемых анализируемым HLA-A	10,55	12,69
Минимальное количество пептидных эпитопов вируса, распознаваемых 90% населения	4,43	5,09

можно заключить, что тщательно выбранные пептиды могут не обеспечить желаемый 100% иммунитет всей популяции по причине невысокого количества безопасно распознаваемых пептидных эпитопов вируса в популяции конкретного региона. Становится понятно, что необходима информация по фенотипу/гаплотипу HLA белорусов (в том числе отличающихся по региону проживания и/или происхождения) для поиска актуальных пептидов.

Угроза гибели человечества от COVID-19 заставила мировое сообщество и исследовательские институты открыто публиковать все результаты исследований по этому вопросу. Использование огромных объемов данных невозможно без программного обеспечения и соответствующего технического исполнения. Такие возможности в настоящее время есть. Разработка вакцин и вакцинация как объективный фактор защиты населения проводится как с помощью традиционных технологий, так и с помощью инновационных [35-38]. В цепочке производства вакцин представляется возможным в Республике Беларусь использование традиционных методов производства – ослабленные, инактивированные вирусы, пептидные вакцины.

Литература

1. Shankarkumar U. The human leukocyte antigen (HLA) system. *International Journal of Human Genetics* 2004; Т. 4, №2: 91-103.
2. Dyer P., McGilvray R., Robertson V. Status report from 'double agent HLA': health and disease. *Molecular immunology* 2013; Т. 55, №1: 2-7.
3. Neeftjes J., Jongsma M. L., Paul P. et al. Towards a systems understanding of HLA class I and HLA class II antigen presentation. *Nature reviews immunology* 2011; Т. 11, №12: 823-836.
4. Tian C., Hromatka B. S., Kiefer A. K. et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature communications* 2017; Т. 8, №1: 1-13.
5. Holoshitz J. The quest for better understanding of HLA-disease association: scenes from a road less travelled by. *Discovery medicine* 2013; Т. 16, №87: 93.
6. Бубнова Л.Н., Павлова И.Е., Беркос А.С. и др. Особенности распределения групп аллелей HLA-A*, B*, DRB1* среди лиц, перенесших COVID-19. *Медицинская иммунология* 2021; Т. 23, №3: 523-532.
7. Kushniarevich A., Sivitskaya L., Danilenko N. et al. Uniparental genetic heritage of Belarusians: encounter of rare Middle Eastern matrilineages with a Central European mitochondrial DNA pool. *PLoS one* 2013; Т. 8, №6: 664-699.
8. Gonzalez-Galarza F.F., McCabe A., Santos E.J. et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic acids research* 2020; Т. 48, №1: 783-788.

Заключение

Мы представили алгоритм выбора *in silico* безопасных иммунодоминантных пептидов SARS-CoV-2, которые могут формировать иммунитет *in vivo*. Для людей, носителей HLA-A*02:01 безопасными оказались 27 из 154023 проанализированных вирусных пептидов SARS-CoV-2, состоящих из 9 а.о.: YLQPRTFLL, MLDMYSVML, RLQSLQTYV, FLPRVFSAV, TMADLVYAL, FVNEFYAYL, FLAFVVFLL, YLYALVYFL, ILFTRFFYV, FLLNKEMYL, YLDAYNMMI, FLAHIQWMV, YLTNDVSFL, FGMSRIGMEV, YIDIGNYTV, FLEYHDVVRV, KLFIRQEEV, FHLVDFQVTI, KLLLEEWNLV, FVLAAYVRI, SLVKPSFYV, WLIVGVALL, FQFCNDPFL, FTISVTTEI, VLNDIFSRL, VVFLHVTYV, FLHVTYVPA. Причем, 13 из 27 принадлежат штамму OL677199.1.

Знание популяционного распространения HLA-фенотипа в конкретном регионе планеты является неотъемлемым условием создания вакцин для предотвращения биологических угроз как уже существующих, так и возможных в будущем.

Идеальной эффективной стратегией профилактики вирусных инфекций является создание индивидуальной персонифицированной вакцины, в том числе на пептидной основе.

9. Hilton H.G., Parham P. Missing or altered self: human NK cell receptors that recognize HLA-C. *Immunogenetics* 2017; Т. 69, №8: 567-579.
10. Yarzabek B., Zaitouna A.J., Olson E. et al. Variations in HLA-B cell surface expression, half-life and extracellular antigen receptivity. *Elife* 2018; Т. 7: 349-361.
11. STRING [Electronic resource] / STRING consortium, 2021. Mode of access: <https://string-db.org/cgi/network?taskId=bxAfAttIzXpe&sessionId=bjze5RI30eJ>. Date of access: 27.09.2021.
12. Embgenbroich M., Burgdorf S. Current concepts of antigen cross-presentation. *Frontiers in immunology* 2018; Т. 9: 1643.
13. Kuznetsov A., Voronina A., Govorun V. Critical Review of Existing MHC I Immunopeptidome Isolation Methods. *Molecules* 2020; Т. 25, №22: 5409.
14. Taylor H.B., Klaefer S., Clauser K.R. et al. Mass spectrometry-based HLA-II peptidomics combined with multi-omics will aid the development of future immunotherapies. *Molecular & Cellular Proteomics* 2021: 100-116.
15. Weiner 3rd J., Suwalski P., Holtgrewe M., et al. Increased risk of severe clinical course of COVID-19 in carriers of HLA-C* 04:01. *EClinicalMedicine* 2021; Т. 40: 999-1010.
16. Amoroso A., Magistroni P., Vespasiano F. et al. HLA and AB0 polymorphisms may influence SARS-CoV-2 infection and COVID-19 severity. *Transplantation* 2021; Т. 105, №1: 193-200.
17. Francis J.M., Leistriz-Edwards D., Dunn A. et al. Allelic variation in Class I HLA determines pre-existing memory responses to SARS-CoV-2 that shape the CD8+ T cell repertoire upon viral exposure. *bioRxiv*, 2021.

18. Hinojosa S.S., Romero V. Risk HLA alleles in South America and potential new epitopes for SARS-CoV2. *Human immunology* №21: 188-198.
19. Lim H.X., Lim J., Jazayeri S.D. et al. Development of multi-epitope peptide-based vaccines against SARS-CoV-2. *Biomedical Journal* 2021; T. 44, №1: 18-30.
20. Sakuraba A., Haider H., Sato T. Population difference in allele frequency of HLA-C* 05 and its correlation with COVID-19 mortality. *Viruses* 2020; T. 12, №11: 1333.
21. Latif A.A., Mullen J.L., Alkuzweny M. et al. Report Lineage B.1.617.2 [Electronic resource]. The Center for Viral Systems Biology, 2021. Mode of access: <https://outbreak.info/situation-reports?pango=B.1.617.2>. Date of access: 14.09.2021.
22. Kirola L. Genetic emergence of B. 1.617. 2 in COVID-19. *New Microbes and New Infections* 2021; T. 43: 1009-1029.
23. Andreatta M., Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 2016; T. 32, №4: 511-517.
24. Bui H.H., Sidney J., Dinh K. et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics* 2006; T. 7, №1: 1-5.
25. Kim Y., Ponomarenko J., Zhu Z. et al. Immune epitope database analysis resource. *Nucleic acids research* 2012; T. 40, №W1: W525-W530.
26. Tenzer S. Peters B., Bulik S. et al. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences CMLS* 2005; T. 62, №9: 1025-1037.
27. Gupta S., Kapoor P., Chaudhary K. et al. In silico approach for predicting toxicity of peptides and proteins. *PloS one* 2013; T. 8, №9: 739-757.
28. Gupta S., Kapoor P., Chaudhary K. et al. Peptide toxicity prediction. *Computational peptidology* 2015: 143-157.
29. Altschul S.F., Wootton J.C., Gertz E.M. Protein database searches using compositionally adjusted substitution matrices. *The FEBS journal* 2005; T. 272, №20: 5101-5109.
30. Nersisyan S. Zhiyanov A., Shkurnikov M. T-CoV: a comprehensive portal of HLA-peptide interactions affected by SARS-CoV-2 mutations. *Nucleic Acids Research* 2021; №1.
31. Wolfson R.L. Chantranupong L., Wyant G.A. et al. KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. *Nature* 2017; T. 543, №7645: 438-442.
32. Shimizu A. Asakawa S., Sasaki T. et al. A novel giant gene CSMD3 encoding a protein with CUB and sushi multiple domains: a candidate gene for benign adult familial myoclonic epilepsy on human chromosome 8q23. 3-q24. 1. *Biochemical and biophysical research communications* 2003; T. 309, №1: 143-154.
33. Chen H.Z., Tang L.L., Yu X. L. et al. Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. *Infectious diseases of poverty* 2020; T. 9, №1: 1-10.
34. Chukwudozie O.S. Gray C.M., Fagbayi T.A. et al. Immunoinformatics design of a multimeric epitope peptide based vaccine targeting SARS-CoV-2 spike glycoprotein. *Plos one* 2021; T. 16, №3: 248-261.
35. Di Natale C., La Manna S., De Benedictis I. et al. Perspectives in Peptide-Based Vaccination Strategies for Syndrome Coronavirus 2 Pandemic. *Frontiers in pharmacology* 2020; T. 11: 578-582.
36. Ghosh N., Sharma N., Saha I. Immunogenicity and antigenicity based T-cell and B-cell epitopes identification from conserved regions of 10664 SARS-CoV-2 genomes. *Infection, Genetics and Evolution* 2021; T. 92: 104-123.
37. Kesheh M.M., Shavandi S., Hosseini P. et al. Bioinformatic HLA Studies in the Context of SARS-CoV-2 Pandemic and Review on Association of HLA Alleles with Preexisting Medical Conditions. *BioMed research international* 2021; T. 2021: 66-93.
38. Tavasolian F, Rashidi M., Hatam G.R. et al. HLA, Immune Response, and Susceptibility to COVID-19. *Front. Immunol.* 2021; 11: 601886. doi: 10.3389/fimmu.

Сведения об авторах:

Цыганков А.М. – старший преподаватель кафедры военной подготовки и экстремальной медицины. ORCID: <https://orcid.org/0000-0003-1367-7742>.

Адрес для корреспонденции – 210009, Республика Беларусь, г. Витебск, пр. Фрунзе, д. 27, Витебский государственный медицинский университет, кафедра военной подготовки и экстремальной медицины; тел. моб.: +375 29 2800551; e-mail: 87senka@gmail.com; Цыганков Арсений Михайлович.

Янченко В.В. – к.м.н., доцент кафедры клинической иммунологии и аллергологии с курсом ФПК и ПК, Витебский государственный ордена Дружбы народов медицинский университет, ORCID: <https://orcid.org/0000-0002-9355-8534>.

Поступила 28.09.2021 г.