

IDENTIFICATION OF CYTOTOXIC T-CELL AND B-CELL EPITOPES IN THE NUCLEOCAPSID PHOSPHOPROTEIN OF SARS-COV-2 USING IMMUNOINFORMATICS

Ali Adel Dawood

*Department of Anatomy, College of Medicine, University of Mosul, Mosul, Iraq,
Almajmuaa Str., Mosul, 41002, Nineveh, Iraq
e-mail: aad@uomosul.edu.iq*

*Last December, a novel coronavirus emerged in Wuhan city, China. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) causes a high intense acute respiratory syndrome with elevation mortality. Nucleocapsid phosphoprotein (NP) is one of the most structural proteins of the virus. NP possesses active immunogenicity for T-cell response. Because NP considered as a potential vaccine target, our study goal was to identify the cytotoxic T-cell (CTL) and B-cell epitopes inside NP peptides. **Methods.** We used a series of popular immunoinformatics and algorithm tools such as FASTA-NCBI, CLUSTAL-OMGA, T-COFFEE, SWISS-MODEL, CTLPred and its branches. **Results.** Homology modeling and alignment of SARS-CoV-2 NP showed high conserved residues compared with related sequences. Different types of the major histocompatibility complex (MHC) alleles were identified, specifically human leukocyte antigens (HLA-A) affinity for NP. We also demonstrate six B-cell epitopes with a high score above the threshold. **Conclusions.** We recorded high binder HLA-A*02:01 alleles matched between the novel coronavirus SARS-CoV-2 NP and the Bat coronavirus SARS-Bat-CoV NP. Identification of CTL response and B-cell predictions will be helpful in reverse immunogenetic approaches, hence in the strategy process of the plausible design of the vaccine.*

Keywords: Nucleocapsid phosphoprotein, SARS-CoV, Immunoinformatics, MHC epitope, Cytotoxic T-cells, Affinity.

SARS-CoV 2003 and MERS-CoV 2012 have caused two large-scale pandemic respiratory diseases [1]. Before efficacious antiviral remedies or vaccines are developed for SARS-CoV or MERS-CoV, COVID-19 has emerged in Wuhan since December 2019 [2–4]. Structurally, coronavirus is a group of enveloped positive single-strand (ssRNA) virus, which comprises four main structural proteins including envelope, nucleocapsid phosphoprotein (NP), membrane protein and spike glycoprotein. These structural proteins possess higher immunogenic activity than the non-structural proteins for T-cell responses. The most important surface spike protein binds to the receptor of the host cell and can fusion through the cell membrane [5–7].

Few researchers have concentrated on the NP of coronavirus. The NP forms from three highly distinct conserved domains, two dependent structural and independent folded structural regions. Two of these regions are called the N-terminal 1 and C-terminal 3 domain. The third jumbled region in the center (RNA-binding domain 2) separates the NTD1 and CTD3 [8, 9]. NP

is a highly phosphorylated protein responsible for the construction of the ribonucleoprotein complex by reacting with the viral genome and serves as a robust immunogen by inducing cellular and humoral immunity. It also regulates the synthesis of viral RNA and protein. One study revealed that NP has a role in viral pathogenesis since anti-N monoclonal antibodies maintain mice from a fatal infection. In the host cells, NP has been shown causing deregulation of the cell cycle [10]. SARS-CoV NP plays a critical role in the antagonism of IFN induction by inhibiting the synthesis process [11]. Moreover, NP plays a crucial role in enhancing the efficiency of viral RNA transcription as well as viral replication. NP is one of the main antigens of all strains SARS-CoV and there are abundant antigenic sites predicted in this protein. It has been suggested that NP is important for the survival of the virus. More importantly, NP was considered as a potential vaccine target in both SARS-CoV and MERS-CoV [12]. The greatest urgency is to be able to detect in a reliable and viable manner the presence of antibodies against coronavirus in the blood of potentially infected people. Because of

the above reasons about the importance of NP, **our study aimed** to find out the immune epitopes of T- and B-cells that could bind with NP through the cell infection. Because of the rapid identification of immune epitopes which is decisive importance at the time of an impending pandemic, our study intended to identify and to detect the signs of cytotoxic T-cell (CTL) and B-cell epitopes in the NP of SARS-CoV-2 (COVID-19) using immunoinformatics and algorithm tools.

Methods. In the current study, we have used a series of presentation measures, popular bioinformatics tools and statistical analysis to evaluate the predictive execution and to achieve the goal of study.

Multiple sequence alignment

14 different amino acid sequences of NP were selected from the National Centre of Biotechnology Information of different sources. We applied sequence alignment using two servers: CLUSTAL-OMEGA from EMBL-EBI-United Kingdom (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and T-COFFEE- Spine (<http://tcoffee.crg.cat/apps/tcoffee/contacts.html>). Eight of them related to the coronavirus family while others related to other families (Table 1). Amino acid sequences retrieved in FASTA format database. Because of the novel nature of COVID-19, we compared the NPs' alignment to ensure sequence similarity in CLUSTAL- OMEGA from EMBL-EBI-UK

pairwise sequence aligner. The sequence alignment was confirmed using T-COFFEE. A phylogenetic tree was extracted using CLUSTAL-OMEGA (EMBL-EBI-UK). Compared sequence alignment from different sources of NP selected had been closest between SARS-Bat-CoV and SARS-CoV-2 around ~ 70% (pink and yellow IDs) using T-COFFEE server rather than other types of viruses (green IDs) with the total score 61 (Table 1).

To confirm the closest correlation of amino acid sequences, more comparisons have done for NP QIM47464.1 SARS-CoV-2, ABN10873.1 Bat-CoV, and YP_002308510.1 Munia-CoV HKU13-3514. SWISS-MODEL and Ramachandran plot was applied for three sequences.

Identification of CTL epitopes

NetCTL 1.2 server (Department of Bio and Health Informatics- Denmark, 2019, <http://www.cbs.dtu.dk/services/NetCTL/>) was used to identify CTL epitopes within the NP sequence. Next, CTLPred (India, <http://crdd.osdd.net/raghava/ctlpred/contact.html>), NetCTL- NetMHC (Department of Bio and Health Informatics- Denmark, 2019, <http://www.cbs.dtu.dk/services/NetCTL/>) server's artificial neural networks (ANNs) were used to predict major histocompatibility complex (MHC) binding sites. In this way, the MHC epitopes specific for NP. NetChop-3.0 and NetChop-2.0-(Denmark, 2020, <http://www.cbs.dtu.dk/services/NetChop/>) were used to predict C-terminal cleavages and a weight

Table 1

Amino acid sequence alignment for different source of NP using CLUSTAL-OMEGA and T-COFFEE

Accession no. NCBI	NP Origin	CLUSTAL-OMEGA similarity %	T-COFFEE similarity %	Release year
QIQ50109.1	SARS-CoV-2	100	72	26-MAR-2020
QIM47464.1	SARS-CoV-2	98.33	73	18-MAR-2020
QIK50415.1	SARS-CoV-2	100	70	16-MAR-2020
ABN10873.1	Bat-CoV HKU4-4	49.87	70	7-FEB-2007
QIM47474.1	SARS-CoV-2	98.06	77	18-MAR-2020
YP_008798235.1	Porcine torovirus	15.82	18	13-AUG-2018
YP_337910.1	Breda virus	15.43	36	13-AUG-2018
YP_009380540.1	Goat Torovirus	15.43	39	13-AUG-2018
YP_001552240.1	Rhinolophus Bat-CoV-HKU2	28.41	52	13-AUG-2018
YP_173243.1	Human CoV-HKU1	18.65	15	13-AUG-2018
YP_009361864.1	Bat-CoV	49.62	68	10-MAR-2019
YP_009724397.2	SARS-CoV-2	98.57	73	30-MAR-2020
YP_001039960.1	Tylonycteris Bat-CoV-HKU4]	50.13	70	13-AUG-2018
YP_002308510.1	Munia-CoV -HKU13-3514	26.23	58	13-AUG-2018

Notes: Accession no. – ID number of each NP sequences obtained from NCBI. According to the highlight colour, the high similarity between sequences showed as: **BAD** **AVG** **GOOD**. The similarity between sequences showed different percentages using CLUSTAL-OMEGA and T-COFFEE.

matrix to estimate TAP transport efficiency to generate prognosis for NP.

Identification of B-cell epitopes

Consecutive B-cell epitopes within the NP peptides were recognized using BepiPred 2.0 (Denmark, 2020, <http://www.cbs.dtu.dk/services/BepiPred/>) server. The residues were scored above the default threshold of 0.5 were further assured in the ABCPred prediction server (India, 2018, <http://crdd.osdd.net/raghava/abcpred/>) with a default threshold of 0.51.

Results. We performed an alignment for NP amino acid sequences of related coronaviruses with the high identity to other viruses, to predict percentage a clear cross-reactivity as follows. The match percentage between NP amino acid sequences illustrated in the Table 1 using two types of alignment server – CLUSTAL-OMEGA and T-COFFEE. Partial amino acid sequence alignment in Fig. 1 showed high conservatism between related coronavirus family but not with other lines on the top of the Fig.1. The same correlation is shown in

the phylogenetic tree in Fig. 2.

Three types of NP sequences have selected for homology modeling using SWISS-MODEL version 11 (Expasy, Switzerland, <https://swiss-model.expasy.org>) server. Favored residues for NP sequences for SARS-Bat-CoV-HKU4-4 and SARS-CoV-2 are matched with 97.75 % and 94.69 % respectively compared with the Munia-CoV- HKU13-3514 with 70.63 % (Fig. 3). The white circles show the perfect locus for CTL epitope binding according to the result extracted from CTLPred, and NetCTL-NetMHC servers.

The next sets of plots clarify predicted local similarity to the target sequence with the number of residues. Predicted local similarity sequences and comparison with the non-redundant set were observed between SARS-Bat-CoV-HKU4-4 and SARS-CoV-2 rather than Munia-CoV- HKU13-3514, the first set plots in Fig. 4. The second set of plots shows comparison with a non-redundant set of a Protein Data Bank (PDB) structure. Red star refers to the closest model with the normalized QMEAN 4 score.

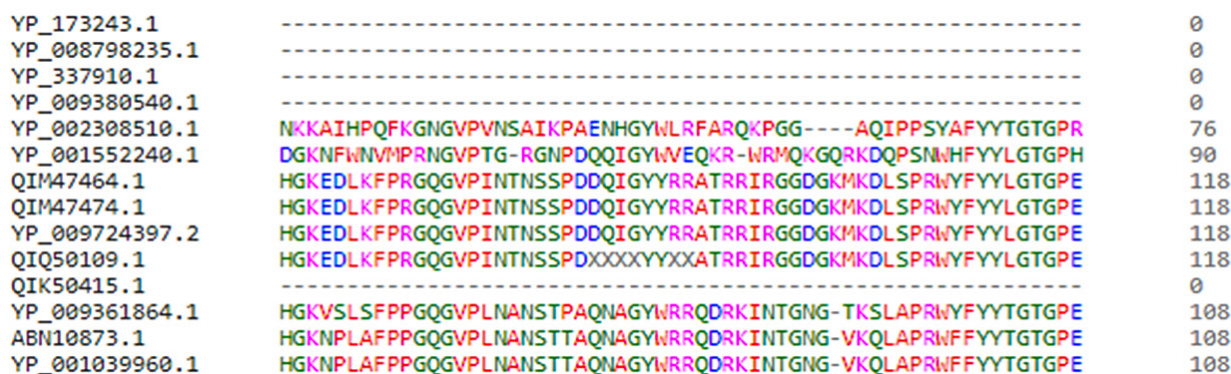


Fig. 1. Partial sequences alignment of 14 types of NP aligned using CLUSTAL-OMEGA taken from the Table 1

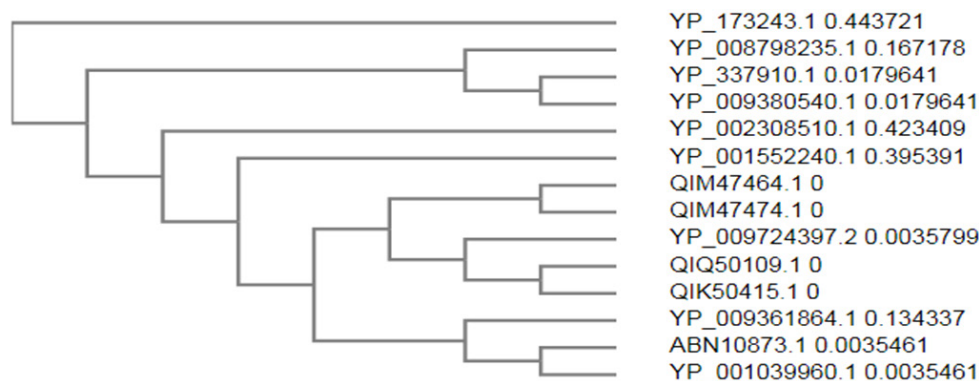


Fig. 2. Phylogenetic tree of ancestor NP extracted from EMBL-EBI through CLUSTAL-OMEGA. It is noticeable that all the sequences of the NP of coronavirus from different sources are related, that is, of one origin in the phylogenetic tree.

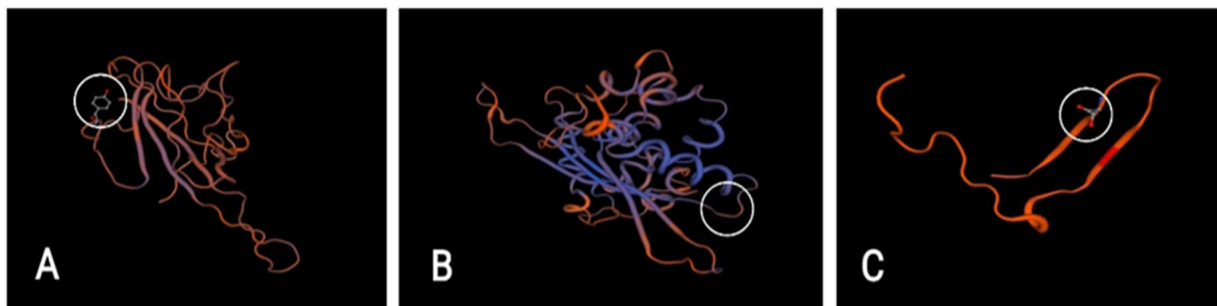


Fig. 3. 3D structure of nucleocapsid protein of related coronaviruses: **A:** SARS-CoV-2(QIM47474.1) 94.69 %; **B:** NP of Bat-CoV- HKU4-4 (ABN10873.1) 97.75 %; **C:** NP of Munia-CoV-HKU13-3514 (YP_002308510.1) 70.63 %. The white circles show the perfect locus for CTL epitope binding according to the result extracted from CTLPred and NetCTL-NetMHC servers.

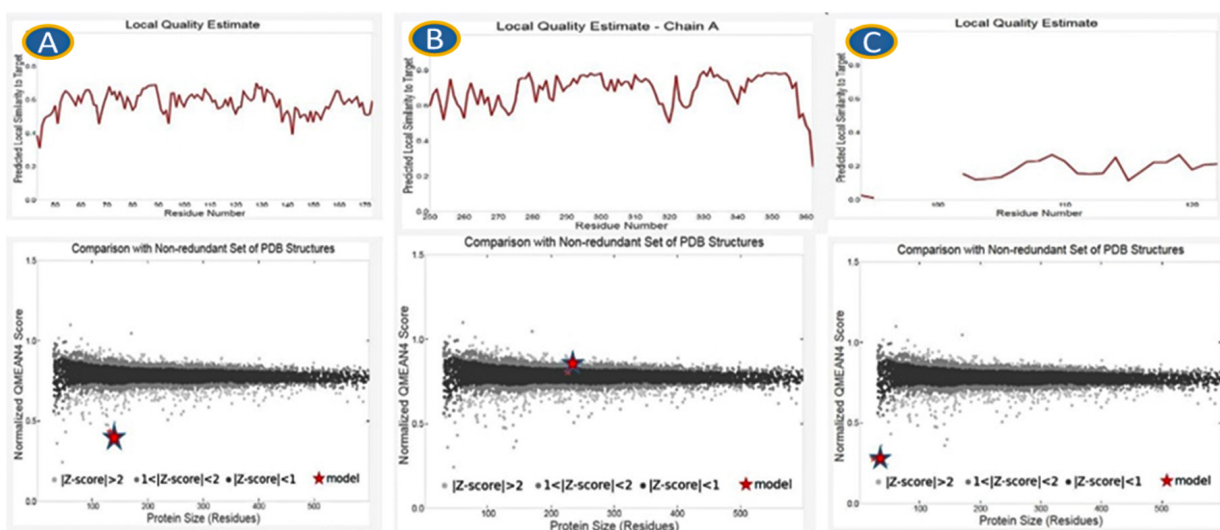


Fig. 4. The local quality estimation between 3 NP sequences: **A:** SARS-CoV-2(QIM47474.1); **B:** NP of Bat-CoV-HKU4-4 (ABN10873.1); **C:** Munia-CoV-HKU13-3514 (YP_002308510.1). The first set shows high similarity between A and B rather than in C. The second set of plots shows comparison with a non-redundant set in PDB structure. Red star refers to the closest model with the normalized QMEAN 4 standard score.

Ramachandran plot sets are high proportion of residues located within the selected sequences and allowed regions in red dots between sequences (Fig. 5). According to the Ramachandran plot, which specified a high attribution of residues to be located within the preferable and allowed regions for three NP sequences under study. Red dots refer to the closest sites between residues location.

NetCTL-1.2 server predicted the MHC supertype A1. We identified nine sites sequences for CTL epitopes inside NP SARS-CoV-2 (QIM47464_1) and eight sequences for CTL epitopes for SARS-Bat-CoV (ABN10873.1) in Table 2. The rank threshold for strong binding peptides is 0.50 while the rank threshold for weak binding peptides is 2.0. We found a similar match on position 351 and 352 between SARS-Bat-CoV-

HKU4-4 and SARS-CoV-2 respectively (Table 2).

CTLPred and NetMHC servers of ANNs were used to predict MHC binding sites specific for the NP sequence (Table 3). The cutoff score is 0.51. Different types of MHC binding restriction alleles were identified for specific three epitopes sequences of NP SARS-CoV-2 (Table 3).

The highly specific MHC epitopes predicted for NP sequences using the NetMHCpan 4.0 server are shown in Table 4. The rank threshold for strong binding peptides is 0.500 while for weak binding peptides is 2.000. The score of the three prediction epitopes is merged into a single score sensitivity and specificity for the epitope-MHC-I binding event (Table 4). We predicted MHC-peptide-binding for the following MHC class 1 supertypes: HLA-A*02:01, HLA-C*04:01, H-2-Dd, HLA-B*

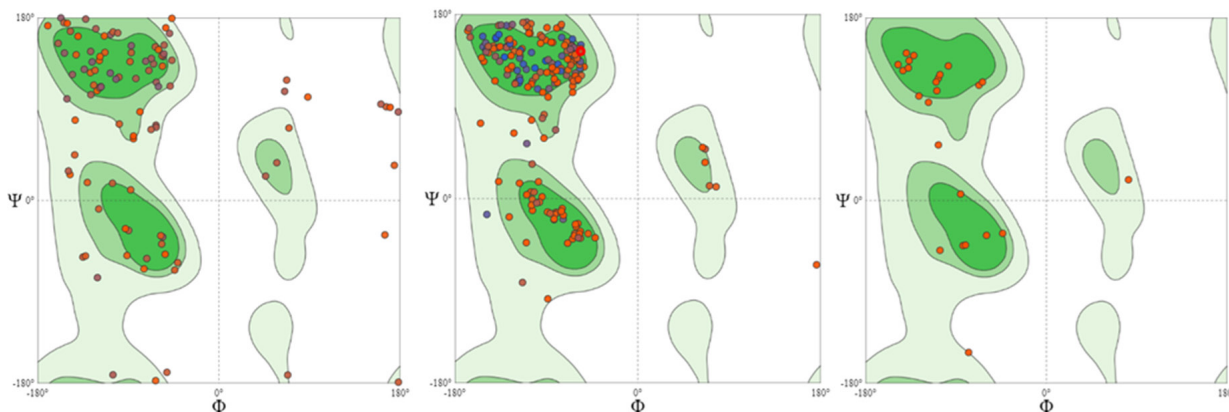


Fig. 5. Ramachandran plots: A: NP of SARS-CoV-2(QIM47474.1) 94.69%; B: NP of Bat-CoV-HKU4-4 (ABN10873.1) 97.75 %; C: NP of Muna-CoV-HKU13-3514 (YP_002308510.1) 70.63 %. Red dots refer to the closest sites between residues location. The similarity residues between A and B is higher than in C.

Table 2

Identifying of CTL epitopes within the NP sequence for two types of coronavirus (SARS-CoV-2 QIM47464_1 and SARS-Bat-CoV ABN10873.1)

NP ID	Pos.	Epitope Sequence	Rank %
QIM47464_1	48	NTASWFTAL	0.952
	78	SSPDDQIGY	1.45
	79	SPDDQIGYY	1.14
	101	MKDLSRWY	0.912
	103	DLSRWYFY	1.944
	104	LSPRWYFYY	2.34
	164	GTTLPKGIFY	1.684
	295	GTDYKHWPQ	1.205
ABN10873.1	352	LLNKHIDAY	1.386
	69	STTAQNAGY	2.7053
	70	TTAQNAGYW	0.7509
	93	QLAPRWFFY	1.9356
	94	LAPRWFFYY	1.1467
	116	VKDGIVVY	1.6694
	337	KLDPKNPNY	1.8579
	351	LLESNIDAY	2.1380
	380	SSASQMEDV	0.8334

Notes: NP ID – nucleocapsid phosphoprotein identification number taken from NCBI. Pos. – position of the epitope in the sequence. Epitope sequence – letters of abbreviation of amino acids. Rank – all selected epitopes are in the in acceptable position sequence of each epitope.

Table 3

Different types of MHC epitopes sequences specific for NP SARS-CoV-2 (QIM47464_1)

Rank	Pos.	Epitope seq.	Score	MHC restriction allele
1	115	TGPEAGLPY	1.000	HLA-A1, HLA-Cw*0401, H2-Db, H2-Kb, H2-Dd, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-B*2703
2	222	LLLDRLNQL	1.000	HLA-A2, HLA-A*0201, HLA-A*0206, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01
3	339	LDDKDPNFK	1.000	HLA-B*3501, HLA-Cw*0401, H2-Db, H2-Dd, H2-Kb, H2-Kd, H2-Ld, HLA-G, H-2Qa, Mamu-A*01, HLA-B*2902, HLA-B44

Notes: Pos. – the position of epitope in the amino acid sequence. Score – we selected only the high score to match with MHC1. MHC restriction allele – types of different alleles that give the high score relate to the epitope.

35:01, HLA-B*44:02 and H-2-Ld. The highest-scoring epitopes were elected for analysis of binding alliance to their identical representative supertypes for each supertype in NetMHCpan. From Table 4, we reported that HLA-A*02:01 was presented in positions 217 and 222 of NP SARS-Bat-CoV-HKU4-4 and SARS-CoV-2 respectively.

The sequential portend B-cell epitopes are assorted according to their score gained by trained frequent neural networks using BepiPred-2.0

(Table 5). All the observed peptides are above the chosen threshold value. NetChop 3.1 and NetChop 2.0 servers were used to identify the cleavage sites in NP SARS-CoV-2. Number of cleavage sites for NetChop 3.1 is 130 compared with 167 for NetChop 2.0 (Table 5). We identified six sequential B-cell epitopes for both NP SARS-CoV-2 and NP SARS-Bat-CoV-HKU4-4 with a high score between them (0.94–0.89) (Table 5). All the epitopes peptides detected are above the cycle threshold chosen. Further identification of cleavage

Table 4
Highly closest of MHC alleles for sequences of both related NP of SARS-CoV-2 QIM47464_1 and SARS-Bat-CoV ABN10873.1

NP Ass.	Pos.	MHC allele	Peptide Sequence	Rank
QIM47464_1	222	HLA-A*02:01	LLLDRLNQL	0.0423
	316	HLA-A*02:01	GMSRIGMEV	0.2558
	314	HLA-C*04:01	FFGMSRIGM	0.4556
	338	HLA-C*04:01	KLDDKDPNF	0.0859
	104	H-2-Dd	LSPRWYFYLL	0.4862
	149	H-2-Dd	RNPANNAAI	0.072
	78	HLA-B*35:01	SSPDDQIGYY	0.295
	355	HLA-B*44:02	KHIDAYKTF	1.3986
	308	H-2-Ld	APSASAFFGM	0.3273
	ABN10873.1	329	HLA-A*02:01	FLSYSGAIKL
	154	HLA-A*02:01	TLIPKNFHI	0.040
	211	HLA-A*02:01	SLAALPLLL	0.072
	217	HLA-A*02:01	LLLDLQKRL	0.2409
	307	HLA-A*02:01	FMSMSQFKL	0.3912

Notes: Pos. – the position of each allele specified in the MHC I. Rank – all selected alleles are within the high rank match with the amino acid sequence of NP.

Table 5
Predict B-cell epitopes and scores, cleavage sites for SARS-CoV-2. Threshold: 0.5.
Comparison between two types of server to detect the B-cell epitopes.

NP Ass.	Rank	Pos.	Epitope B-Cell	Score	NetChop 3.1	Pos.	NetChop 2.0	Pos.
QIM47464_1	1	91	TRRIRGGDGKMKDLSP	0.94	M S	411	M	411
	2	249	KSAAEASKKPRQKRTA	0.93	S	413	S S	413
	2	136	EGALNTPKDHIGTRNP	0.93	A	414	A	414
	3	354	NKHIDAYKTFPTEPK	0.91	D	415	D S	415
	3	24	TGSNQNGERSGARSKQ	0.91	T	417	T	417
	3	127	KDGIIWVATEGALNTP	0.91	Q	419	A S	419
ABN10873.1	1	436	AGSITMRSGSSPALQD	0.93	–	–	–	–
	2	60	RGRGRNPKPRPAPNNT	0.92	–	–	–	–
	3	161	EGATDAPSVFGTRNPA	0.91	–	–	–	–
	4	42	PRTISFADNNDNQPNQ	0.89	–	–	–	–
	4	352	HQSNDDKGDPIYFLSY	0.89	–	–	–	–
	4	322	FGTEDPRWPQMAELAP	0.89	–	–	–	–

Notes: Pos. – site position of the epitope inside the amino acid sequence. Score – we selected high identity score for epitopes which gives the similarity between two types of NP under study. Abbreviation of amino acids: M – methionine; S – serine; A – alanine; D – aspartic acid; T – threonine; Q – glutamine. It has been noted that there is differences between amino acid sequences of epitopes between SARS-CoV-2 QIM47464_1 and SARS-Bat-CoV ABN10873.1.

C-terminal sites residues of NP SARS-CoV-2 was done by two servers. NetChop 3.1 is detected 130 number of cleavage sites compared with NetChop 2.0 with 167 sites.

Discussion. Previous studies revealed that novel SARS-CoV-2 (COVID-19) may originate from SARS-Bat-CoV emerged in 2003. SARS-CoV-2 has been mutated in the spike glycoprotein (S) and NP [13]. NP is a highly phosphorylated protein that has several functions. Our study aimed to focus on NP because of the low variable rate with regards to its size advocates its importance to virus survival [14].

In the current study, results of sequence alignment are agreed with a previous study that recommended identity 80 % of the complete genome between SARS-Bat-CoV and SARS-CoV-2 [15–17]. On the contrary, we demonstrated a difference between NP SARS-CoV-2 and SARS-Bat-CoV using CLUSTSL-OMEGA server.

Dependable prognosis of immunogenic peptides may minimize the experimental effort prerequested to identify new epitopes. It was evaluated that only 1/200 peptides will bind a specific MHC class 1 allele with adequate strength to evolve CTL response. The restrictive step involved in antigen presentation is the binding to MHC class 1 [18, 19]. The CTL epitopes are typically used as an antigen-presenting which we observed to bind with the peptide-binding groove of MHC1. This may be relevant for specific MHC alleles. Our records are agreed with a previous study revealed that residue contacts between epitopes and HLA chains were disclosed in a certain map matrix and the salt bridge-forming residues of the surface glycoprotein SARS-CoV [20]. Combining predictions of TAP transport efficiency with the HLA-A*0201 affinity predictions has previously been done and was shown improving identification of CTL epitopes [21].

The predicted B-cell epitopes are assorted upon on their score acquired by trained frequent neural networks. A higher registered score of peptide means a higher eventuality to be an epitope. Different methods have been developed for predicting the proteasomal cleavage pattern of proteins [22, 23]. In designing the optimal prediction method, we tested two versions of

the NetChop cleavage predictor. Because of outperforms, the different identities between two types, and changing peptide residues, we do not recommend using them for low specificity and sensitivity.

In this study, we applied a completely integrative approach to improving CTL epitope identification, unified predictions of MHC class 1 binding alliance and C-terminal proteasomal cleavage sites. In our opinion, afflicted identification of eligible peptides of evolving CTL response and B-cell prognostic will be helpful in reverse immunogenetic accessions, hence in the process of vaccine designing.

Our study has few limitations including difficulty predicting CTL and B-cell epitopes *in vitro* compared to practical work *in vivo*. Because of novel COVID-19 emerged as a mutation of SARS-CoV and MERS-CoV, virus RNA replication is error-prone and reported a mutation rate of 4×10^{-4} substitutions. In our point of view, it is feasible, as an emergency, to use the NP from SARS-CoV as antigens to detect the new SARS-CoV-2, as it is the detection and not discrimination at the moment.

Conclusions. Sequence alignment results of NP from several sources are different closest between CLUSTAL-OMEGA and T-COFFEE. Combining HLA-A*0201 affinity predictions with the prognosis of target sequence efficiency was shown to improve the identification of CTL epitopes. NetChop 3.1 and 2.0 servers are less accurate for predicting cleavage sites of peptides. We record high binder HLA-A*02:01 alleles matched between NP SARS-CoV-2 and NP SARS-Bat-CoV. Furthermore, we demonstrate six B-cell epitopes with a high score above the threshold.

Acknowledgments

The author is very grateful to the University of Mosul, College of Medicine for improving the quality of this work.

Competing Interests

The author has declared that is no competing interest exists.

Funding

This paper didn't receive any specific grant from any institution.

ВИКОРИСТАННЯ ІМУНОІНФОРМАТИКИ ДЛЯ ІДЕНТИФІКАЦІЇ ЕПІТОПІВ НУКЛЕОКАПСИДНИХ ФОСФОПРОТЕЇНІВ SARS-COV-2 ДЛЯ РОЗПІЗНАВАННЯ ЦИТОТОКСИЧНИМИ Т-ЛІМФОЦИТАМИ ТА В-КЛІТИНАМИ

Алі Адель Давуд

*Кафедра анатомії, Медичний коледж,
Університет Мосула, Мосул, Ірак,
вул. Алмаджмуа, Мосул, 41002, Ніневія, Ірак*

Резюме

У грудні минулого року в місті Ухань (Китай) з'явився новий коронавірус. Вірус SARS-COV-2 викликає гострий респіраторний синдром високої інтенсивності з підвищеним рівнем смертності. Нуклеокапсидний фосфопротеїн (НФ) є одним з найбільших структурних білків вірусу, він має імуногенну активність і здатен викликати Т-клітинну імунну відповідь. Оскільки НФ розглядається як потенційна основа для створення вакцини, **метою** нашого дослідження було виявити епітопи всередині пептидів НФ, які можуть розпізнаватися цитотоксичними Т-лімфоцитами (ЦТЛ) та В-клітинами. **Методи.** Ми використували низку загальнозживаних засобів та алгоритмів імуноінформатики, таких як FASTA-NCBI, CLUSTAL-OMGA, T-COFFEE, SWISS-MODEL, STLPred та їх різновиди. **Результати.** Гомологічне моделювання та суміщення НФ SARS-CoV-2 показали наявність високої кількості консервативних залишків порівняно із суміжними послідовностями. Були виявлені різні типи алелів основного комплексу гістосумісності (МНС), зокрема афінність лейкоцитарних антигенів людини (HLA-A) до НФ. Ми також показали наявність шести епітопів, які здатні розпізнаватися В-клітинами на рівні вище порогового значення. **Висновки.** Ми встановили співпадіння у високому рівні зв'язування алелей HLA-A*02:01 між НФ нового коронавірусу SARS-CoV-2 та НФ коронавірусу кажанів SARS-Bat-CoV. Ідентифікація відповіді ЦТЛ та прогнозування відповіді В-клітин буде корисною при зворотному імуногенетичному підході, отже, у процесі стратегічного планування правдоподібного дизайну вакцини.

Ключові слова: нуклеокапсидний фосфопротеїн, SARS-CoV, імуноінформатика, епітоп МНС, цитотоксичні Т-лімфоцити, афінність.

1. Thomson G. COVID-19: social distancing, ACE 2 receptors, protease inhibitors and beyond. *Int J Clin Prac.* 2020; March. <http://doi.org/10.1111/ijcp.13503>.
2. Cheepsattayakorn A, Cheepsattayakorn R. Proximal Origin and Phylogenetic Analysis of COVID-19 (2019-nCoV or SARS-CoV-2). *EC Mic.* 2020;9–12. <https://www.nature.com/articles/s41591-020-0820-9>.
3. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication and pathogenesis. *J Med Virol.* 2020; 92:418–423. <https://doi.org/10.1002/jmv.25681>.
4. Dawood A. Mutated COVID-19, May Foretells Mankind in a Great Risk in The Future. *N Mic N Inf.* 2020. <https://doi.org/10.1016/j.nmni.2020.100673>.
5. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang L, Sheng A, Quan L, Xia Z, Tan W, Cheng G, Jiang T. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Ce Ho & Mic.* 2020; <http://doi.org/10.1016/j.chom.2020.02.001>.
6. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Tan KS, Wang DY, Yan Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak –an update on the status. *Mil Med Res.* 2020; 7:11 <https://doi.org/10.1186/s40779-020-00240-0>.
7. Li S, Wu H, Yan H, Ma H, Wang S, Zhang L, Tang M, Temperton X, Weiss N, Brenchley R, et al. T cell responses to whole SARS coronavirus in humans. *J Immunol.* 2008; 181:5490–5500.
8. Kim JM, Chung YS, Jo HJ, Lee NJ, Kim MS, Woo SH, Park S, Kim JW, Kim HM, Han MG. Identification of Coronavirus Isolated from a Patient in Korea with COVID-19. *Os Pub He Res Per.* 2020; 11(1):3–7. <https://doi.org/10.24171/j.phrp.2020.11.1.02>.
9. McBride R, Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses.* 2014; 6:2991–3018. <http://doi: 10.3390/v6082991>.
10. Wurm T, Chen H, Hodgson T, Britton P, Brooks G, Hiscox JA. Localization to the nucleolus is a common feature of coronavirus nucleopro-

- teins, and the protein may disrupt host cell division. *J Virol.* 2001; 75:9345–9356.
11. You J, Dove BK, Enjuanes L, DeDiego ML, Alvarez E, Howell G, Heinen P, Zambon M, Hiscox JA. Subcellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein. *J Gen Virol.* 2005; 86(Pt 12):3303–3310.
 12. Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J Med Virol.* 2020; 92:495–500. <http://DOI: 10.1002/jmv.25698>.
 13. Lana R, Coellaho FC, Gomes MF, Cruz OG, Bastos LS, Villela, Codeco CT. The novel coronavirus (SARS-CoV-2) emergency and the role of timely and effective national health surveillance. *Rep in Pub He.* 2020; <http://doi: 10.1590/0102-311X00019620>.
 14. Dawood A, Alnori H. Tunicamycin Anticancer Drug May Reliable to Treat Coronavirus Disease-19. *OAMJMS.* 2020; 8(T1):129–133. <https://doi.org/10.3889/oamjms.2020.4954>.
 15. Dawood A. Glycosylation, Ligand Binding Sites and Antigenic Variations Between Membrane Glycoprotein of COVID-19 and Related Coronaviruses. *Vacunas.* 2020. <https://doi.org/10.1016/j.vacun.2020.09.005>.
 16. Baglivo M, Baronio M, Natalini G, Beccari T, Fuulcheri PC, Petralia P, Michelini S, Fiorentini G, Miggiano GA, Morresi A, Tonini G, Bertelli M. Natural small molecules as inhibitors of coronavirus lipid dependent attachment to host cells: a possible strategy for reducing SARS-COV-2 infectivity? *Acta Biomed.* 2020; 91(1):161–164. <http://doi:10.23750/abm.v91i1.9402>.
 17. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Arxiv.* 2020; 10965v1 [q-bio.GN].
 18. Leung D, Tam F, Ma CH, Chan PK, Cheung JL, Niu H, Tam JS, Lim PL. Antibody Response of Patients with Severe Acute Respiratory Syndrome (SARS) Targets Viral Nucleocapsid. *T J Inf Dis.* 2004; 190:379–86. <https://doi:10.1086/422040>.
 19. Peters B, Bulik S, Tampa R, Endert PM, Holzhtutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol.* 2003; 171:1741–1749.
 20. Chen J. Pathogenicity and transmissibility of 2019-nCoV — a quick overview and comparison with other emerging viruses. *Microb & Inf.* 2020; 22:69–71.
 21. Zhang L, Zhang F, Yu W, He T, Yu J, Yi CE, Ba L, Li W, Farzan M, Chen Z, Yuen KY, Ho D. Antibody responses against SARS coronavirus are correlated with the disease outcome of Infected Individuals. *J Med Virol.* 2006; 78:1–8. <https://doi:10.1002/jmv.20499>.
 22. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency and proteasomal cleavage predictions. *Eur J Immunol.* 2005; 35:2295–2303. <https://doi:10.1002/eji.200425811>.
 23. Nielsen M, Lundegaard C, Lund O, Kasmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogen.* 2005; 57:33–41. <https://doi:10.1007/s00251-005-0781-7>.

Received 24.08.2020