Identification and Comparison of Possible Epitope – Designed Targets Using in-Silico Techniques for Corona Virus

[1]M. Nithya, [2]Dr. Horne Iona Averal
[1]Research Scholar, Department of Zoology, Holy Cross College (Autonomous), Tiruchirappalli-620 002
[2]Associate Professor and IQAC Coordinator, Department of Zoology, Holy Cross College (Autonomous), Tiruchirappalli-620 002

Abstract **- The SARS Coronavirus-2 (SARS-CoV-2) epidemic has become a global issue that has raised concerns for the scientific community to design and find a way to combat this deadly virus. To date, the epidemic has claimed hundreds of thousands of lives as a result of infection and spread. Growing evidence suggests that T cells may play a key role in the fight against acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Therefore, COVID-19 vaccines that can obtain a strong T cell response may be very important. The design, development and evaluation of vaccine trials help to understand the T cell epitopes of SARS-CoV-2, which is less well known. Because of the challenges of diagnosing epitopes by experimentation, many studies have suggested the use of in-silico methods. Here, we present of the in-silico methods used to predict SARS-CoV-2 T cell epitopes. These methods use a different set of technical methods, which often focus on machine learning. Functional comparisons are based on the diagnostic power of a specific set of immunogenic epitopes determined by experiments targeted T cells in recovering COVID-19 patients, highlighting the relative functional relevance of the various methods adopted by in - Silico studies. The investigation also prioritizes ideas for future research guidelines.**

Index Terms **- Coronavirus, COVID-19, SARS-CoV-2, Epitope based techniques, In- silico techniques, MHC prediction, Bioinformatics.**

INTRODUCTION

There are extensive studies being conducted around the world to develop appropriate therapies to control the effects of severe acute respiratory syndrome coronavirus 2 (SARS-cov-2) that causes COVID-19 in humans. The first patient infected with SARS-cov-2 was diagnosed in December 2019 in Wuhan, China. Later, the virus spread to 187 countries and regions due to its high quality and infected people. Coronaviruses belong to the Family, Coronaviridae, and a branch of Nidovirales is a big family presented with positive strains. Coronaviruses are zoonotic pathogens, which can cause diseases of the intestines, liver, respiratory system and central nervous system.

The World Health Organization (WHO) has declared the outbreak of SARS-cov-2 as a global public health emergency of international concern (PHEIC) on January 30, 2020, and the epidemic on March 11, 2020.The unavailability of drugs or medically proven vaccines is a major concern of the COVID-19 epidemic. Therefore, effective measures such as early detection of people with the virus, isolation from society, the use of masks and hand sanitizers, etc (Zhou et al., 2020)In the Sars-cov-2 Genome, the 5'-end Polyprotein (Pp1ab) codes and then cleaves into 15-protein Structures (Nsp-1-a, NSP-10, NSP-12-NSP-16), And 3'- for the latter codes Four Structural Proteins, including, Spike (S), Coating (E), Membrane Surface (M) and Nucleocapsid (N) Protein, as well as six additional proteins (3a, 3b) 6, (7a, 7b), 8a, 9, and Orf 14) (Zheng et al., 2020).

The virus controls the cell by initially binding to the angiotensin converting enzyme II (ACE2) in the alveoli of the lungs. As a result, the infected person suffers from respiratory problems. Traditional methods of vaccination based on laboratory tests in the event of an outbreak could not meet emergency needs, and many medical agents are being investigated. Approaches are used to develop vaccines or recombinant vaccines. Traditional methods are based on inactivated or live attenuated computer threats that can be used for vaccine development, but it has been noted that these methods have some limitations, such

as Labour intensity, as well as problems in producing many proteins and pathogens (Dangi et al.,2018).

The use of silicon preparation method is very important because it will allow predicting antigenicity, epitope regions on T cells, and other parameters such as peptide, subcellular localization, and salicylic acid in the target protein (Dangi et al., 2018; May et al., 2020). For this analysis, the unifying force between the predicted epitopes of the selected MHC-I and MHC-II genes is an important part of obtaining a silicon approach (Dangi et al.,2018). At present, insufficient knowledge of latency and SARS-CoV-2 infection increases the uncertainty of viral persistence. Some antiretroviral drugs are currently not available.

Symptoms of COVID-19 can range from mild to severe including cough, fever and shortness of breath. Most people have asymptomatic. Symptoms may appear two to fourteen days after exposure. About 20% progress to serious illness, such as respiratory failure, pneumonia and death in some cases. Many patients appear to have less illness (Gu et al., 2020; Kooraki et al., 2020). Current information has shown that it spreads from person to person among those close to a distance of 6 meters, or 2 meters. The virus is spread through respiratory droplets when an infected person coughs or sneezes (Peeri et al., 2020).

Silicon analysis and its use for analysing the influence of physicochemical properties, antigenicity, subcellular localization of viral proteins, predicting the t-cell epitope, predicting the MHC-I and MHC-II Epitopes, predicting the effect of solvent on the location of epitopes, finally, the effects of changes often occur, and structural protein that have been predicted to have only one peptide per antigenicity. Epitopes have been studied for comparison in various organisms such as Mouse and Human.

TOOLS AND SOFTWARE

There are many tools involved in these studies.FASTA stands for FAST homology search All sequences. The protein sequencing alignment system was created by Pearsin and Lipman in 1988.The basic idea is to add a quick pre-screen step to get the most similar segments between two sequences, and then extend these same segments into local alignment using more robust algorithms like Smith-Waterman. FASTA can be very specific when identifying long regions of low similarity especially in very different sequences.

PREDICTION OF T-CELL EPITOPES WITH HLA

Peptide prediction of potential T-cell epitopes schedule-MHC I selected the best proximity binds with confidence 1/0. 89, and lest ic50 for each allele, and in the same MHCpred used an additive method to predict forced major histocompatibility complex (MHC) Class I (HLA- (A*) And II (DRB*) Molecules and Transporter Associated with Processing (TAP). The allele-specific quantitative structure-activity relationship (QSAR) model was found, with partial least squares (pls) method (Kobayashi H *et al*., 2000).

PREDICTION OF B CELL EPITOPES

B Epitopes plays an important role in the development of epitope-based vaccines and the surrounding research area. the dominant B- cell line epitope, which can be used in the treatment of autoimmune diseases, in which the goal is a neutralizing antibody response it can also induce antibodies that cross-react with the parent's protein (Saha *et al*.,2006).

PREDICTION OF MHC-I AND MHC-II EPITOPES

B-cell Epitopes are part of an antigen for induced immunoglobulin or antibody, and in order to activate single-cells to give the immune system a response (Sanchez-trincado, *et al*., 2017). T cells recognize epitopes when these are presented to them bound to MHC molecules. Therefore, epitopes can be predicted by computing Their MHC-binding Profile. because of the differences in the molecular interactions between epitopes and MHC I And II Complexes, the prediction of epitopes binding to MHC I is more accurate than to MHC II. for both types, we used IEDB tools as detailed in materials and methods (Sanchez-trincado, *et al*., 2017).

These silico methods and tools are often developed independently and, in many cases, have been trained using viral or other virus-related data sets, making it difficult to understand the epitope-related functionality of SARS-CoV-2. To help clarify these questions, this review provides a prediction comparison of 61 SARS-CoV-2 in silico studies, showing similarities and differences between certain SARS-CoV-2 epitopes predicted in different ways. Also, it was evaluated and compared, predictions using the emerging data from nine experimental studies that identified SARS-CoV-2 T cell epitopes targeted to COVID-19 recovering patients.

The main objectives is to

- Search for ligand using in-silico tools that would elicit response to form antibodies and hence protect individual from the infection.

Check various parameters of ligand - targets which will be most possibly used to produce a vaccine.

- Identify and compare the parameters that will give the possible target. Thus, there is a scope by epitope-designing and recombinant DNA technology to study and to produce vaccine models.

MATERIALS AND METHODS

COLLECTION OF TARGET PROTEIN SEQUENCE

Collecting Protein Sequences of these Possible Ligands from Uniprot KB Database to Provide the Scientific Community with a comprehensive, High-quality and freely Accessible Resource of Protein Sequence and Functional Information. https://www.Uniprot.org/

PREDICTION OF T-CELL EPITOPES WITH HLA

MHCpred (Peptide Prediction to find Possible T-cell Epitopes with HLA MHCI best affinity binding alleles with confidence 1/0.89 and one with least IC50 value for each allele were selected and listed. http://www.ddg-pharmfac.net/mhcpred/MHCPred/

PREDICTION OF B CELL EPITOPES.

Linear B Cell epitopes of the reference genome structural proteins, variant proteins, and the proteins that have a signal peptide were predicted by Abcpred http://crdd.osdd.net/raghava/abcpred/

PREDICTION OF MHC I AND MHC II EPITOPESE

Prediction of MHC-I and MHC-II epitopes of the reference genome structural proteins, Variant proteins, and the proteins that have a Signal Peptide were analysed By IEDB http://tools.iedb.org/bcell/

RESULT AND DISCUSSION

FASTA SEQUENCE

(i) Mus musculus (Mouse)

>tr|Q9D836|Q9D836_MOUSE Angiotensin-converting enzyme 2 (Fragment) OS=Mus musculus X=10090 GN=Ace2 PE=2 SV=2

XCDISNSTEAGQKLLKMLSLGNSEPWTKALEN
VVGARNMDVKPLLNYFQPLFDWLKEQNRNSF
VGWNTEWSPYADQSIKVRISLKSALGANAYEW
TNNEMFLFRSSVAYAMRKYFSIIKNQTVPFLEE
DVRVSDLKPRVSFYFFVTSPQNVSDVIPRSEVE
DAIRMSRGRINDVFGLNDNSLEFLGIHPTLEPPY
QPPVTIWLIIFGVVMALVVVGIIILIVTGIKGRKK
KNETKREEYDSMDIGKGESNAGFQNSDDAQTS
F

(ii) Homo sapiens (Human)

>AAQ89076.1 ACE2 [Homo sapiens]

MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFN
HEAEDLFYQSSLASWNYNTNITEENVQNMNNA
GDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQL
QALQQNGSSVLSEDKSKRLNTILNTMSTIYSTG
KVCNPDNPQECLLLEPGLNEIMANSLDYNERL
WAWESWRSEVGKQLRPLYEEYVVLKNEMAR
ANHYEDYGDYWRGDYEVNGVDGYDYSRGQL
IEDVEHTFEEIKPLYEHLHAYVRAKLMNAYPSY
ISPIGCLPAHLLGDMWGRFWTNLYSLTVPFGQ
KPNIDVTDAMVDQAWDAQRIFKEAEKFFVSVG
LPNMTQGFWENSMLTDPGNVQKAVCHPTAW
DLGKGDFRILMCTKVTMDDFLTAHHEMGHIQ
YDMAYAAQPFLLRNGANEGFHEAVGEIMSLSA
ATPKHLKSIGLLSPDFQEDNETEINFLLKQALTI
VGTLPFTYMLEKWRWMVFKGEIPKDQWMKK
WWEMKREIVGVVEPVPHDETYCDPASLFHVSD
DYSFIRYYTRTLYQFQFQEALCQAAKHEGPLH
KCDISNSTEAGQKLL

PREDICTION OF T-CELL EPITOPES WITH HLA

(i) Mus musculus (Mouse)

>tr|Q9D836|Q9D836_MOUSE Angiotensin-converting enzyme 2 (Fragment) OS=Mus musculus X=10090 GN=Ace2 PE=2 SV=2

The HLA allele used in the test is: A0201

The query sequence

CDISNSTEAGQKLLKMLSLGNSEPWTKALENVVGARNMDVKPL
LNYFQPLFDWLKEQNRNSFVGWNTEWSPYADQSIKVRISLKSAL
GANAYEWTNNEMFLFRSSVAYAMRKYFSIIKNQTVPFLEEDVRV
SDLKPRVSFYFFVTSPQNVSDVIPRSEVEDAIRMSRGRINDVFGL
NDNSLEFLGIHPTLEPPYQPPVTIWLIIFGVVMALVVVGIIILIVTG
IKGRKKKNETKREENSYDSMDIGKGESNAGFQNSDDAQTSF

Amino acid groups	Predicted -logIC$_{50}$ (M)	Predicted IC$_{50}$ Value (nM)	Confidence of prediction (Max = 1)
YAMRKYFSI	7.53	29.51	1.00
IIHPTLEPPY	7.504	31.33	0.78
PRVSFYFFV	7.437	36.56	0.89
YFFVTSPQN	7.409	38.99	0.78
LLNYFQPLF	7.386	41.11	1.00
LEPPYQPPV	7.381	41.59	0.89
KLLKMLSLG	7.311	48.87	0.89
YEWTNNEMF	7.273	53.33	0.89
KNQTVPFLE	7.229	59.02	0.78
EMFLFRSSV	7.218	60.53	0.89
IIFGVVMAL	7.216	60.81	1.00
LEFLGIHPT	7.21	61.66	0.89
YFQPLFDWL	7.186	65.16	0.89
IKNQTVPFL	7.179	66.22	0.89
TIWLIIFGV	7.169	67.76	1.00
VSFYFFVTS	7.075	84.14	0.89
EFLGIHPTL	7.032	92.90	0.78
PLLNYFQPL	6.999	100.23	1.00
RNMDVKPLL	6.999	100.23	0.89
VMALVVVGI	6.956	110.66	1.00
NDNSLEFLG	6.911	122.74	0.78
PLFDWLKEQ	6.897	126.77	0.89
KMLSLGNSE	6.89	128.82	0.89
FYFFVTSPQ	6.878	132.43	0.78

KALENVVGA	6.844	143.22	1.00
KYFSIIKNQ	6.812	154.17	0.78
MDIGKGESN	6.803	157.40	0.78
LIVTGIKGR	6.79	162.18	1.00
VTIWLIIFG	6.727	187.50	0.89
GRINDVFGL	6.708	195.88	0.89
VVVGIIILI	6.708	195.88	1.00
FLGIIHPTLE	6.666	215.77	0.89
IFGVVMALV	6.666	215.77	0.89
SLGNSEPWT	6.659	219.28	1.00
SFYFFVTSP	6.657	220.29	0.78
WLIIFGVVM	6.651	223.36	1.00
NYFQPLFDW	6.626	236.59	0.78
IIKNQTVPF	6.58	263.03	1.00
QTVPFLEED	6.563	273.53	0.89
QKLLKMLSL	6.543	286.42	0.89
ISNSTEAGQ	6.528	296.48	0.78
GIIILIVTG	6.523	299.92	0.89
FLFRSSVAY	6.513	306.90	0.89
GRKKKNETK	6.504	313.33	0.78
LLKMLSLGN	6.503	314.05	0.89
LSLGNSEPW	6.489	324.34	0.78
QSIKVRISL	6.458	348.34	0.89
NMDVKPLLN	6.457	349.14	0.89
DSMDIGKGE	6.449	355.63	0.78
EAGQKLLKM	6.447	357.27	0.89
INDVFGLND	6.438	364.75	0.78
AMRKYFSII	6.427	374.11	1.00
YADQSIKVR	6.416	383.71	1.00
LNDNSLEFL	6.396	401.79	0.89
VRVSDLKPR	6.394	403.65	0.89
SLKSALGAN	6.389	408.32	0.89
VVMALVVVG	6.375	421.70	0.89
FSIIKNQTV	6.371	425.60	0.89
NSLEFLGIH	6.358	438.53	0.78
VKPLLNYFQ	6.35	446.68	0.78
LFDWLKEQN	6.333	464.52	0.78
PFLEEDVRV	6.31	489.78	0.89
FVGWNTEWS	6.304	496.59	1.00
EWTNNEMFL	6.304	496.59	0.78
ALVVVGIII	6.304	496.59	1.00
FQPLFDWLK	5.602	2500.35	0.89
SSVAYAMRK	5.601	2506.11	0.78
NSYDSMDIG	5.598	2523.48	0.78
SIIKNQTVP	5.564	2728.98	0.89
FLEEDVRVS	5.561	2747.89	1.00
VFGLNDNSL	5.556	2779.71	0.89
TLEPPYQPP	5.55	2818.38	0.89
RKKKNETKR	5.549	2824.88	0.89
NSDDAQTSF	5.546	2844.46	0.89
WSPYADQSI	5.484	3280.95	0.89
NSTEAGQKL	5.459	3475.36	0.89
VEDAIRMSR	5.453	3523.71	0.89
PYQPPVTIW	5.442	3614.10	0.78
WTNNEMFLF	5.426	3749.73	1.00
KNETKRIEN	5.385	4120.98	0.78
SRGRINDAF	5.359	4375.22	0.89
VAYAMRKYF	5.348	4487.45	1.00
NNEMFLFRS	5.344	4528.98	0.89
GFQNSDDAQ	5.342	4549.88	0.78
VSDLKPRVS	5.333	4645.15	0.89
ANAYEWTNN	5.33	4677.35	0.78
LKPRVSFYF	5.33	4677.35	0.89
AGQKLLKML	5.326	4720.63	1.00
TSPQNVSDV	5.324	4742.42	0.89
DVIPRSEVE	5.324	4742.42	0.89
SVAYAMRKY	5.313	4864.07	0.89
LKSALGANA	5.312	4875.28	0.89
ENVVGARNM	5.31	4897.79	0.78
SEPWTKALE	5.31	4897.79	0.78
VVGARNMDV	5.303	4977.37	1.00
GNSEPWTKA	-	-	-
WTKALENVV	-	-	-
LKMLSLGNS	-	-	-
TKALENVVG	-	-	-
DVRVSDLKP	-	-	-
LGIIHPTLEP	-	-	-
IRMSRGRIN	-	-	-
GARNMDVKP	-	-	-
RISLKSALG	-	-	-
SNSTEAGQK	-	-	-
KALENVVGA	6.844	143.22	1.00
KYFSIIKNQ	6.812	154.17	0.78
MDIGKGESN	6.803	157.40	0.78
LIVTGIKGR	6.79	162.18	1.00
VTIWLIIFG	6.727	187.50	0.89
GRINDVFGL	6.708	195.88	0.89
VVVGIIILI	6.708	195.88	1.00
FLGIIHPTLE	6.666	215.77	0.89
IFGVVMALV	6.666	215.77	0.89
SLGNSEPWT	6.659	219.28	1.00
SFYFFVTSP	6.657	220.29	0.78
WLIIFGVVM	6.651	223.36	1.00
NYFQPLFDW	6.626	236.59	0.78
IIKNQTVPF	6.58	263.03	1.00
QTVPFLEED	6.563	273.53	0.89
QKLLKMLSL	6.543	286.42	0.89
ISNSTEAGQ	6.528	296.48	0.78
GIIILIVTG	6.523	299.92	0.89
FLFRSSVAY	6.513	306.90	0.89
GRKKKNETK	6.504	313.33	0.78
LLKMLSLGN	6.503	314.05	0.89
LSLGNSEPW	6.489	324.34	0.78
QSIKVRISL	6.458	348.34	0.89
NMDVKPLLN	6.457	349.14	0.89
DSMDIGKGE	6.449	355.63	0.78
EAGQKLLKM	6.447	357.27	0.89
INDVFGLND	6.438	364.75	0.78
AMRKYFSII	6.427	374.11	1.00
YADQSIKVR	6.416	383.71	1.00
LNDNSLEFL	6.396	401.79	0.89
VRVSDLKPR	6.394	403.65	0.89
SLKSALGAN	6.389	408.32	0.89
VVMALVVVG	6.375	421.70	0.89
FSIIKNQTV	6.371	425.60	0.89
NSLEFLGIH	6.358	438.53	0.78
VKPLLNYFQ	6.35	446.68	0.78
LFDWLKEQN	6.333	464.52	0.78
PFLEEDVRV	6.31	489.78	0.89
FVGWNTEWS	6.304	496.59	1.00
EWTNNEMFL	6.304	496.59	0.78
ALVVVGIII	6.304	496.59	1.00

GLNDNSLEF	5.985	1035.14	1.00
SFVGWNTEW	5.963	1088.93	0.78
NAYEWTNNE	5.956	1106.62	0.89
RINDVFGLN	5.954	1111.73	0.89
SDLKPRVSF	5.918	1207.81	0.89
ALGANAYEW	5.901	1256.03	0.89
TNNEMFLFR	5.885	1303.17	0.89
ADQSIKVRI	5.879	1321.30	0.89
SNAGFQNSD	5.861	1377.21	0.78
PPYQPPVTI	5.85	1412.54	1.00
VTSPQNVSD	5.834	1465.55	0.89
IWLIIFGVV	5.822	1506.61	0.89
MALVVVGII	5.81	1548.82	1.00
EWSPYADQS	5.809	1552.39	0.78
RKYFSIIKN	5.8	1584.89	0.78
QNRNSFVGW	5.798	1592.21	0.78
LIIFGVVMA	5.785	1640.59	1.00
PRSEVEDAI	5.772	1690.44	0.89
WLKEQNRNS	5.763	1725.84	1.00
DNSLEFLGI	5.761	1733.80	0.89
DVKPLLNYF	5.76	1737.80	1.00
SIKVRISLK	5.755	1757.92	0.89
VIPRSEVED	5.728	1870.68	0.89
VGWNTEWSP	5.711	1945.36	0.89
CDISNSTEA	5.7	1995.26	0.89
EPWTKALEN	5.694	2023.02	0.78
KVRISLKSA	5.688	2051.16	1.00
IILIVTGIK	5.684	2070.14	0.89
PYADQSIKV	5.683	2074.91	0.89
LFRSSVAYA	5.671	2133.04	0.89
SEVEDAIRM	5.662	2177.71	0.89
EDVRVSDLK	5.656	2208.00	0.67
NSFVGWNTE	5.652	2228.44	0.78
YDSMDIGKG	5.649	2243.88	0.78
KKNETKREE	5.641	2285.60	0.78
FGLNDNSLE	5.629	2349.63	0.89
FQNSDDAQT	5.622	2387.81	1.00
NSEPWTKAL	5.617	2415.46	0.89
NEMFLFRSS	5.617	2415.46	0.89
DISNSTEAG	5.614	2432.20	0.89
TEAGQKLLK	5.606	2477.42	0.78

The HLA allele used in the test is: A0201

The query sequence

MREAGWDKGGRILMCTKVTMDDFLTAHHEMGHIQYDMAYAAQ
PFLLRNGANEGFHEAVGEIMSLSAATPKHLKSIGLLSPDFQEDNE
TEINFLLKQALTIVGTLPFTYMLEKWRWMVFKGEIPKDQWMKK
WWEMKREIVGVVEPVPHDETYCDPASLFHVSNDYSFIRYYTRTL
YQFQFQEALCQAAKHEGPLHKCDISNSTEAGQKLFNMLRLGKSE
PWTLALENVVGAKNMNVRPLLNYFEPLFTWLKDQNKNSFVGWS
TDWSPYADQSIKVRISLKSALGDKAYEWNDNEMYLFRSSVAYA
MRQYFLKVKNQMILFGEEDVRVANLKPRISFNFFVTAPKNVSDII
PRTEVEKAIRMSRSRINDAFRLNDNSLEFLGIQPTLGPPNQPPVSI
WLIVFGVVMGVIVVGIVILIFTGIRDRKKKNKARSGENPYASIDIS
KGENNPGFQNTDDVQTSF

Amino acid groups	Predicted -logIC_{50} (M)	Predicted IC_{50} Value (nM)	Confidence of prediction (Max = 1)
LLNYFEPLF	7.94	11.48	1.00
TLYQFQFQE	7.638	23.01	0.89
PRISFNFFV	7.575	26.61	0.89
IVFGVVMGV	7.552	28.05	1.00
PLLNYFEPL	7.495	31.99	1.00
EMYLFRSSV	7.428	37.33	0.89
AMRQYFLKV	7.411	38.82	1.00
MVFKGEIPK	7.403	39.54	0.89
KNMNVRPLL	7.337	46.03	0.89
ISFNFFVTA	7.281	52.36	0.89
GLLSPDFQE	7.265	54.33	0.89
YAMRQYFLK	7.261	54.83	0.89
KSIGLLSPD	7.256	55.46	0.78
SIWLIVFGV	7.242	57.28	1.00
EIVGVVEPV	7.215	60.95	0.89
YDMAYAAQP	7.204	62.52	0.78
YMLEKWRWM	7.199	63.24	1.00
FLLKQALTI	7.181	65.92	1.00
KLFNMLRLG	7.177	66.53	0.89
VFKGEIPKD	7.153	70.31	0.78

MGVIVVGIV	-	-	-
EPLFTWLKD	-	-	-
FGEEDVRVA	-	-	-
SPYADQSIK	-	-	-
DYSFIRYYT	-	-	-
EKAIRMSRS	-	-	-
TAHHEMGHI	-	-	-
AQPFLLRNG	-	-	-
STEAGQKLF	-	-	-
TPKHLKSIG	-	-	-
GGRILMCTK	-	-	-
SPDFQEDNE	-	-	-
HKCDISNST	-	-	-
FQEALCQAA	-	-	-
KGENNPGFQ	-	-	-
MGHIQYDMA	-	-	-
VGAKNMNVR	-	-	-
VGEIMSLSA	-	-	-
KPRISFNFF	-	-	-
EGPLHKCDI	-	-	-
NDAFRLNDN	-	-	-
KGEIPKDQW	-	-	-
NPGFQNTDD	-	-	-
TGIRDRKKK	-	-	-
RTEVEKAIR	-	-	-
DQSIKVRIS	-	-	-

VSIWLIVFG	7.143	71.94	0.78
FNFFVTAPK	7.114	76.91	0.78
NDYSFIRYY	7.095	80.35	0.78
EMKREIVGV	7.091	81.10	0.89
VVMGVIVVG	7.079	83.37	0.89
MAYAAQPFL	7.049	89.33	1.00
YEWNDNEMY	7.032	92.90	0.78
VKNQMILFG	7.027	93.97	0.78
WLIVFGVVM	7.025	94.41	1.00
YCDPASLFH	7.021	95.28	0.78
NYFEPLFTW	7.019	95.72	0.78
QMILFGEED	6.982	104.23	0.89
AAQPFLLRN	6.978	105.20	0.89
LLKQALTIV	6.976	105.68	1.00
EFLGIQPTL	6.963	108.89	0.78
GTLPFTYML	6.948	112.72	1.00
PLFTWLKDQ	6.942	114.29	0.89
NDNSLEFLG	6.911	122.74	0.78
FIRYYTRTL	6.909	123.31	1.00
FEPLFTWLK	6.907	123.88	0.78
DFLTAHHEM	6.89	128.82	0.89
IVGTLPFTY	6.88	131.83	0.89
YAAQPFLLR	6.873	133.97	1.00
KAYEWNDNE	6.856	139.32	0.89
IVGVVEPVP	6.847	142.23	0.89
FFVTAPKNV	6.844	143.22	0.89
GHIQYDMAY	6.816	152.76	0.78
VMGVIVVGI	6.799	158.85	1.00
FRLNDNSLE	6.792	161.44	0.78
SKGENNPGF	6.779	166.34	0.89
MRQYFLKVK	6.743	180.72	0.78
KDQNKNSFV	6.734	184.50	0.89
KNVSDIIPR	6.734	184.50	0.89
AAKHEGPLH	6.715	192.75	0.89
KHLKSIGLL	6.707	196.34	0.89
YLFRSSVAY	6.703	198.15	0.89
YQFQFQEAL	6.689	204.64	1.00
FLLRNGANE	6.687	205.59	0.89
NMLRLGKSE	6.683	207.49	0.89
ILFGEEDVR	6.67	213.80	1.00
RLGKSEPWT	6.642	228.03	1.00

IQYDMAYAA	5.414	3854.78	1.00
TEVEKAIRM	5.412	3872.58	0.89
SLEFLGIQP	5.403	3953.67	0.89
PPNQPPVSI	5.392	4055.09	1.00
TLPFTYMLE	5.391	4064.43	0.89
AATPKHLKS	5.391	4064.43	1.00
DNEMYLFRS	5.375	4216.97	0.89
DPASLFHVS	5.368	4285.49	1.00
DNETEINFL	5.366	4305.27	0.89
GFQNTDDVQ	5.361	4355.12	0.78
PHDETYCDP	5.346	4508.17	0.78
LSAATPKHL	5.345	4518.56	0.89
EEDVRVANL	5.333	4645.15	0.78
PTLGPPNQP	5.324	4742.42	0.89
LKDQNKNSF	5.322	4764.31	0.89
NNPGFQNTD	5.321	4775.29	0.78
IILAVGEIMS	5.319	4797.33	0.89
VGIVILIFT	5.31	4897.79	1.00
FHEAVGEIM	-	-	-
PGFQNTDDV	-	-	-
STDWSPYAD	-	-	-
AYAMRQYFL	-	-	-
DISKGENNP	-	-	-
RTLYQFQFQ	-	-	-
PDFQEDNET	-	-	-
AFRLNDNSL	-	-	-
IIIQYDMAYA	-	-	-
LLRNGANEG	-	-	-
RNGANEGFH	-	-	-
SFVGWSTDW	-	-	-
RISLKSALG	-	-	-

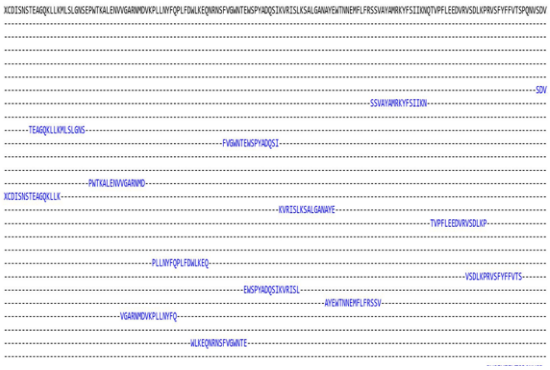
PREDICTION OF B CELL EPITOPES.

>tr|Q9D836|Q9D836_MOUSE Angiotensin-converting enzyme 2 (Fragment) OS=Mus musculus X=10090 GN=Ace2 PE=2 SV=2

Sequence name	
Length of the sequence	265
Number of 16mers from the input sequence	250
Threshold setting (Default value is 0.5)	0.51

Rank	Sequence	Start position	Score
1	IILIVTGIKGRKKKNE	218	0.94
2	DSMDIGKGESNAGFQN	242	0.93
3	EFLGIHPTLEPPYQPP	183	0.91
3	MSRGRINDVFGLNDNS	166	0.91
4	SEVEDAIRMSRGRIND	158	0.90
5	SDVIPRSEVEDAIRMS	152	0.88
6	SSVAYAMRKYFSIIKN	105	0.86
7	KREENSYDSMDIGKGE	235	0.85
8	TEAGQKLLKMLSLGNS	8	0.82
9	FVGWNTEWSPYADQSI	63	0.81
10	KKKNETKREENSYDSM	229	0.79
10	PVTIWLIIFGVVMALV	198	0.79
11	PWTKALENVVGARNMD	25	0.77
12	XCDISNSTEAGQKLLK	1	0.76
13	KVRISLKSALGANAYE	79	0.75
14	TVPFLEEDVRVSDLKP	122	0.74
15	KGESNAGFQNSDDAQT	248	0.72
16	FGLNDNSLEFLGIHPT	175	0.71
17	PLLNYFQPLFDWLKEQ	43	0.70
17	VSDLKPRVSFYFFVTS	132	0.70
18	EWSPYADQSIKVRISL	69	0.68
19	AYEWTNNEMFLFRSSV	92	0.67
20	VGARNMDVKPLLNYFQ	34	0.66
21	FGVVMALVVGIIILI	206	0.64
22	WLKEQNRNSFVGWNTE	54	0.63
22	EPPYQPPVTIWLIIFG	192	0.63
23	RVSFYFFVTSPQNVSD	138	0.52

OVERLAP DISPLAY



>AAQ89076.1 ACE2 [Homo sapiens]

INPUT INFORMATION

Sequence name	
Length of the sequence	555
Number of 16mers from the input sequence	540
Threshold setting (Default value is 0.5)	0.51

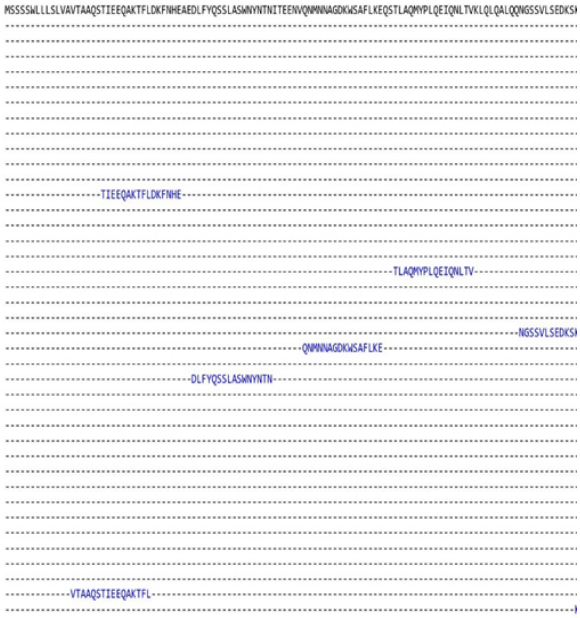
TABULAR RESULT

Predicted B-cell epitope

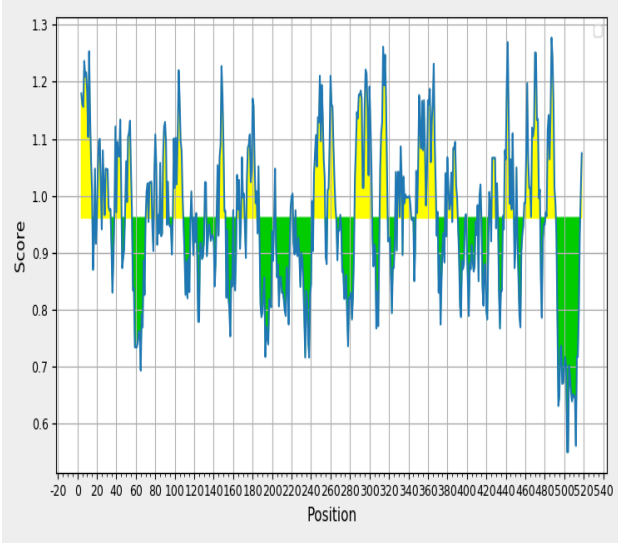
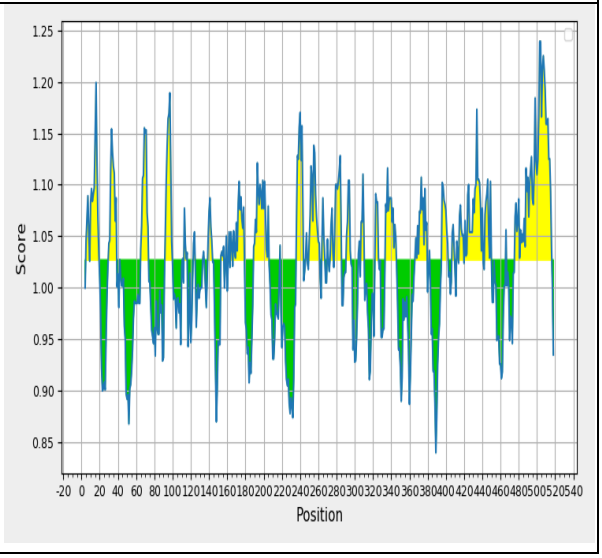
**The predicted B cell epitopes are ranked according to their score obtained by trained recurrent neural network.
Higher score of the peptide means the higher probability to be as epitope.
All the peptides shown here are above the threshold value chosen.**

Rank	Sequence	Start position	Score
1	MSTIYSTGKVCNPDNP	123	0.96
2	IVGVVEPVPHDETYCD	484	0.93
2	YEDYGDYWRGDYEVNG	196	0.93
3	RILMCTKVTMDDFLTA	357	0.90
3	HLLGDMWGRFWTNLYS	265	0.90
4	SMLTDPGNVQKAVCHP	331	0.88
4	TFEEIKPLYEHLHAYV	229	0.88
5	KKWWEMKREIVGVVEP	475	0.87
5	TIVGTLPFTYMLEKWR	445	0.87
6	HEMGHIQYDMAYAAQP	374	0.86
7	LGKGDFRILMCTKVTM	351	0.85
7	TIEEQAKTFLDKFNHE	20	0.85
8	KGEIPKDQWMKKWWEM	465	0.84
8	QKAVCHPTAWDLGKGD	340	0.84
8	PDNPQECLLLEPGLNE	135	0.84
9	LFHVSDDYSFIRYYTR	503	0.83
10	TLAQMYPLQEIQNLTV	78	0.82
10	KWRNMVFKGEIPKDQW	458	0.82
10	FWTNLYSLTVPFGQKP	274	0.82
10	RLWAWESWRSEVGKQL	161	0.82
10	NGSSVLSEDKSKRLNT	103	0.82
11	QNMNNAGDKWSAFLKE	60	0.81
11	DETYCDPASLFHVSDD	494	0.81
11	DLFYQSSLASWNYNTN	38	0.81
12	VGEIMSLSAATPKHLK	404	0.80
12	DDFLTAHHEMGHIQYD	367	0.80
12	DQAWDAQRIFKEAEKF	299	0.80
12	VGKQLRPLYEEYVVLK	172	0.80
13	SLTVPFGQKPNIDVTD	280	0.79
13	GQLIEDVEHTFEEIKP	220	0.79
14	EVNGVDGYDYSRGQLI	208	0.78
15	KPNIDVTDAMVDQAWD	288	0.77
16	KSIGLLSPDFQEDNET	419	0.76
16	YDMAYAAQPFLLRNGA	381	0.76
16	QGFWENSMLTDPGNVQ	325	0.76
17	YVRAKLMNAYPSYISP	243	0.75
17	NEIMANSLDYNERLWA	149	0.75
18	VTAAQSTIEEQAKTFL	14	0.73
18	KRLNTILNTMSTIYST	114	0.73
19	TGKVCNPDNPQECLLL	129	0.72
20	AAKHEGPLHKCDISNS	532	0.71
20	TNITEENVQNMNNAGD	52	0.71
21	DYSFIRYYTRTLYQFQ	509	0.70
22	YEHLHAYVRAKLMNAY	237	0.69
23	YEEYVVLKNEMARANH	180	0.68
24	ISPIGCLPAHLLGDMW	256	0.66
25	PDFQEDNETEINFLLK	426	0.65
26	GDKWSAFLKEQSTLAQ	66	0.64
26	LHKCDISNSTEAGQKL	539	0.64
27	YTRTLYQFQFQEALCQ	516	0.60
27	EAEKFFVSVGLPNMTQ	310	0.60
27	FLDKFNHEAEDLFYQS	28	0.60
28	VGLPNMTQGFWENSML	318	0.59
29	SAATPKHLKSIGLLSP	411	0.57
30	LQEIQNLTVKLQLQAL	85	0.56
30	FLLRNGANEGFHEAVG	390	0.56

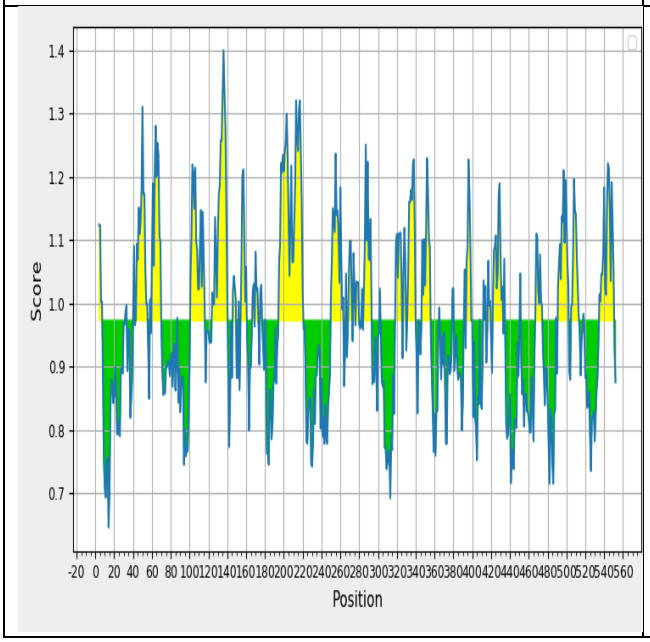
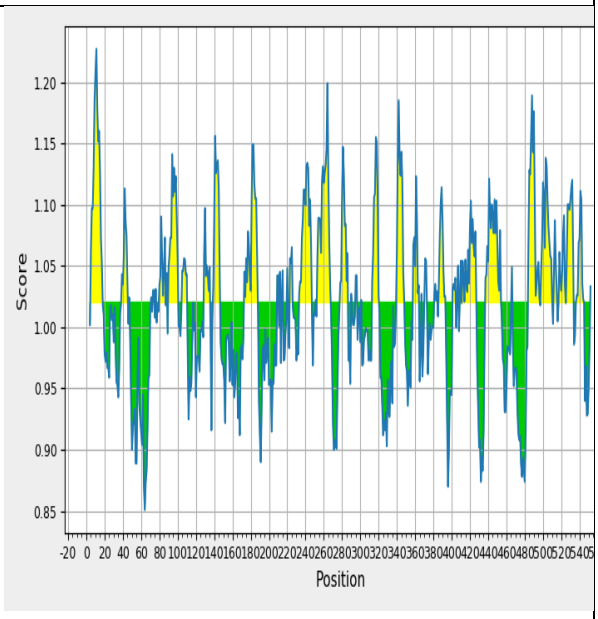
OVERLAP DISPLAY



PREDICTION OF MHC I AND MHC II EPITOPESE
>tr|Q9D836|Q9D836_MOUSE Angiotensin-converting enzyme 2 (Fragment) OS=Mus musculus OX=10090 GN=Ace2 PE=2 SV=2

Chou & Fasman Beta-Turn Prediction Results	Kolaskar and Tongaonkar
	

>AAQ89076.1 ACE2 [Homo sapiens]

Chou & Fasman Beta-Turn Prediction Results	Kolaskar and Tongaonkar
	

SARS-COV-2, The causative agent of respiratory distress syndrome, has infected more than 10,000 people worldwide, leading to numerous deaths. First detected in Wuhan, Hubei Province, China, COVID-19 spread uncontrollably, eventually becoming a global threat. Scientists around the world are working to find a solution to this evil outbreak (manojit bhattacharya et al., 2019). Based on the World Health Organization (WHO) website (https://who.sprinklr.com/) 23 July, 2021, there have been 192,284,207 confirmed cases of covid-19, including 4,136,518 deaths, reported to WHO. As of 24 july 2021, a total of 3,605,386,928 vaccine doses have been administered (WHO). Today, researchers are exploring ways to develop subunit vaccines from a complete gene / protein of pathogens. Epitope assessment for antibodies has become more important with the development of computational tools for vaccine design (Dubey et al., 2018). There is a subdivision in the field of bioinformatics, which includes many tools and databases. Immunological dataset prediction and in silico analysis is done with

the help of tools. Advances in tools and the availability of a variety of data, such as genetic, proteomic, and various algorithms, have made it more effective for scientists to accurately estimate the epitopes that are most effective in the development of subunits of vaccines (De Gregorio De et al., 2012).

Prediction Of T-Cell Epitopes with HLA using the MHC prediction server. Program results are shown in a three-column table. The first column shows the peptide sequence, the second and third columns show the IC50 and IC50 values inverted by the IC, respectively. If the IC50 value is above 5000, the peptide will not bind to MHC atoms. Arrays of peptides are sorted with IC50 values. Peptides with lower IC50 values (or higher IC50 values than IC50 values) are listed first and non-binders are given at the bottom of the Table. (http://www.ddgpharmfac.net/mhcpred/MHCPred/)

The future development of MHC Pred will improve both the scope and use of the server and the sub-system. First, we expect to significantly increase the number of allele models, with increased focus on both human Class II and HLA-B and HLA-C loci, as well as non-human allele, particularly murine, i -bovine and primate. Although the peptide binding data binding to class I alleles of length outside 9 is limited, we will also seek to produce peptide binding models of lengths 8, 10 and 11. We are also looking at technological advances aimed at the automatic excavation of epitope genomes. Similarly, by combining a user-defined set of allele models, we will be able to address the problem of identifying contaminated peptides that can bind several different MHC alleles. Second, the additive method used in MHC Pred, itself, relies on the presence of certain amino acids in specific areas within the set of training peptides to reliably predict the effect of that residue on that condition in any experimental peptides.

MHC Class I predictor episodes: The MHC Pred server predicts peptides binding to different alleles of Class I MHC. Selection of binding peptides is made on the basis of binding points. Peptides selected as epitopes have a threshold score of 0.8 or more. Such peptides showed a predictable binding correlation of less than 10nM. The server uses a default threshold score of 0.5 however, a higher score value has been used to detect epitopes with high binding relationships of MHC alleles. Predicted episodes and their schools. Higher numbers of epitopes were predicted by protein sequence (Van Regenmortel et al., 1993).

MHCPred is one of the most effective among the available binding methods for class I MHC. Therefore, this method was chosen to make predictions. There have been highly contaminated epitopes predicted in the study. The immoral nature of epitopes is a desirable asset as one epitope can bind to different alleles. Only spike proteins have shown the presence of four contaminated epitopes. All peptides showing high binding may not be able to activate Tc cells. In order for the peptide to act as a Tc cell epitope it must be approved for proteasome processing, indicating the correlation to be mediated by TAP in addition to the Class I MHC affinity. The server predicts CTL epitopes among a group of binding peptides in Class I MHC.

The combined score greater than 1 was used as the selective condition for CTL epitopes (although the default limit value is 0.75). As MHC Pred is trained in human data it should therefore provide better functioning of the human proteasome and TAP (Larsen et al., 2004). Therefore, peptides complementing both conditions (MHCPred score of 0.8 or more and CTL scores over 1) were selected as epitopes. These epitopes have a high binding affinity for MHC molecules and are capable of processing the cytosolic pathway, transported by TAP with high potential to act as epitopes. We analyzed five different SARS-COV-2 proteins in the current study (due to their availability on the NCBI-GenBank website and their role in the structural role in SARS-COV-2 and finally revealed T-Cell epitopes that could be used for wet laboratory observations In the most recent study, different episodes of SARS-COV-2 were discovered, based on In-silico methods and focus, but in our study there are many differences as we analyze a group of proteins from SARS-COV-2 to classify Ts -Cell epitopes with short lengths straight to MHC I and MHC II.

Prediction of B-Cell Epitopes using ABC prediction server. Server for estimating linear B cell epitope regions in antigen sequence using artificial neural network. This server assists in selecting synthetic vaccine candidates, identifying epitope areas that can be used in diagnosis and allergy research. High score of the peptide means higher probability to as epitope. All the peptide shown here threshold value chosen. (http://crdd.osdd.net/raghava/abcpred/).The threshold

applied to both servers was 0.5 and the peptides had high score they were considered epitopes. The common epitopes predicted by both servers were used. ACE2 prediction could be used as ABC Pred server can process predictions for less than 6000 protein residues of amino acids. B epitopes of B cells fall into two categories - linear and continuous. For the lack of all SARS-CoV-2 protein components, in our study we have predicted only for episodes. The B cell epitopes through two servers that strengthen the chances of detection by immune system.Prognosis of B-cell epitopes in antigen sequence is an important and complex problem. Although, most antigenic protein selections do not persist, it is possible to mimic epitopes with synthetic peptides (Van Regenmortel et al., 1993;1994). One of the major problems faced in developing B-cell epitope predictors is the variable epitope length. Performance is much better than random, despite the fact that B-cell epitope prediction is a complex problem. It is therefore advisable to use an ABC pred server to detect B-cell epitopes in the antigen.

The immune epitope site (IEDB) (Peters et al., 2005) is probably the most complete site for B-cell and T-cell epitopes tested. The IEDB provides users with access to epitope-related analyzes and predictive tools including: (i) a few ways to predict accurate and consistent B-cell epitopes; (ii) a visual tool for predicted conformational epitopes in a 3D antigen structure; (iii) several epitope data analysis tools (e.g., computerized epitope preservation and epitope population inclusion). The IEDB allows users to obtain both internal biochemical information and external epitope-based information (Peters et al., 2005). This makes it possible to easily integrate customized data sets (e.g., a set of security data. In addition, several researchers have used the IEDB to perform meta-pathogens analysis of interest, thereby improving the use of the IEDB in the analysis and prediction of B-cell epitopes.

Chou and Fasman's method is based on arithmetic the potential for the expanded remains to form part of the second curve structure of B turn. First of the two experimental defensive epitopes were predicted with seven to six lists therethe window size was seven and nine, respectively. Kolaskar and Tongaonkar calculated the frequency of occurrences of each type of amino acid (fAg) in 156 experimental epithets from 34 different proteins. Then, using the Parker's scale, moderate levels of hydrophilicity, accessibility, and variability were calculated across heptapeptide dispersed at 156 epitopes, and the frequency of amino acids increased. (fs) was calculated. The antigenic propensity (Ap) value of each amino acid is calculated. About 75% of epitopes have been correctly predicted using an antigenic propensity scale for epitopes for which the scale has been developed. In HA1, the route predicted epitope 91–108, while the second protective episode was not predicted.

Reverse vaccinology plays an important role in the development of recombinant vaccines by allowing the in silico analysis of the viral genome. In-silico analysis it helps to identify the most antigenic and hidden proteins that are important for vaccine development before the start of a wet laboratory study (Dangi et al., 2018). Using this approach, the current study aims to identify potentially antigenic proteins and epitope regions targeted by both the B and T cell flexible immune systems to improve vaccine or sero diagnostic testing as described.The recent global pandemic of SARS-CoV-2 has claimed hundreds of precious lives in various parts of the world and weakened the economies of many countries. A fully effective vaccine against the approved drug or SARS-CoV-2 has not yet been reported. In this study, there was a successful attempt to make ACE2 against SARS-COV-2. The current study describes ACE2 as a potential candidate for vaccine production. However, current research is the result of an integrated vaccine omics approach. Therefore, more experimental research by future is needed to demonstrate the efficacy of the developed vaccine.

REFERENCE

[1] Boni, M. F. 2020. The Evolutionary Origin of Sars-cov-2, The Sarbecovirus Lineage Behind The COVID-19 Pandemic. Microbiol.

[2] Chen, H., Tang, L., Yu, X., Zhou, J., Chang, Y. and Wu, X. 2020. Bioinformatics analysis of epitope-based vaccine design against the novel SARS-CoV-2. Infectious Diseases of Poverty, 9(1): 1–10.

[3] Dangi, M., Kumari, R., Singh B. and Chhillar, A. K. 2018. Advanced In Silicon Tools for The Design of The Antigenic Epitope, As A Potential Candidate Vaccines Against the Coronavirus. In

Bioinformatics: Sequence, Structure, Phylogeny. 329-357.

[4] Flower, D.R.2003. Towards in silico prediction of immunogenic epitopes. Trends Immunol.24:667.

[5] Guo, Y.R. and Cao, Q.D. 2020. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status, Mil. Med. Res. 7: 11.

[6] Haste Andersen, P., Nielsen, M. and Lund, O. 2006.Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 15: 2558-2570.

[7] Kobayashi, H., Wood, M., Song, Y., Appella, E. and Celis, E. 2000. Promiscuous Helper T-cell Epitopes of the Class II MHC Tumor Antigen HER2/Neu. Cancer Res. 60:5228-5236.

[8] Peters, B., Sidney, J., Bourne, P., Huynh-Hoa, B. and Buus ,S. 2005. The immune epitope database and analysis resource: From vision to blueprint. PLoS Biol. 3 (3): e91-10.1371/journal.pbio.0030091.

[9] Saha, A., Sharma, A. R., Bhattacharya, M., Sharma, G., Lee, S.-S. and Chakraborty, C. 2020. Tocilizumab: A therapeutic option for the treatment of cytokine storm syndrome in COVID-19. Archives of Medical Research, 51, 595–597.

[10] Sanchez-Trincado, J.L., Gomez-Perosanz, M. and Reche, P.A. 2017. Fundamentals and Methods for T- and B-Cell Epitope Prediction. J Immunol Res. 2017:2680160.

[11] Van Regenmortel, M.H. 1993. Synthetic peptides versus natural antigens in immunoassays. Ann Biol Clin (Paris). 51:39– 41.

[12] Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L., and Zou, Q. 2020. Progress and prospects on vaccine development against SARS-CoV-2. Vaccines, 8:153.