Research Article

Identification of T-Cell Epitopes in Proteins of Novel Human Coronavirus, SARS-Cov-2 for Vaccine Development

Biswajit Sahoo[1], Krishna Kant[1], Neeraj K. Rai[2*], Dharmendra Kumar Chaudhary[1*]

[1]Department of Molecular Medicine & Biotechnology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, 226014, India

[2]Department of Biotechnology, Central University of South Bihar, Gaya, 824236, India

*Corresponding Author(s): Dharmendra Kumar Chaudhary, Department of Molecular Medicine & Biotechnology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, 226014, India; E-mail: chaudharydk12@gmail.com; Neeraj K. Rai, Department of Biotechnology, Central University of South Bihar, Gaya, 824236, India, E-mail: neerajrai2601@gmail.com

Abstract

Recently, a novel human coronavirus, SARS-CoV-2 led to a worldwide serious health concern, causing severe respiratory tract infections in humans. It is the third highly pathogenic and transmissible coronavirus after severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in humans. The source of origin, transmission to humans and mechanisms associated with the pathogenicity of SARS-CoV-2 are not clear yet, however, its resemblance with SARS-CoV and several other bat coronaviruses was recently confirmed through genome sequencing related studies. It has been an emergent need to develop a potent and adequate number of drugs and vaccines to control the spread of coronavirus. We have screened the specific proteins such as ORF1ab polyprotein, surface glycoprotein, membrane glycoprotein and nucleocapsid phosphoprotein of SARS-CoV-2 for identification of T-cell epitopes using immunoinformatics tools. In this study we used different bioinformatics tools for analysis of genome and proteome. We retrieved gene sequence from NCBI. The expected molecular weight and isoelectric point (pI) values were also verified using Generunner and ExPaSy. These epitopes have showed the highest binding affinity with major histocompatibility complex (MHC) class I and II molecules. These findings may be useful as an immunodiagnostic tool for the development of peptide based novel vaccines.

Keywords: SARS-CoV-2; Epitopes; MHC; Immunodiagnostic; Vaccine

Introduction

Coronaviruses (CoV) belong to the family Coronaviridae [1]. Recombination rates of CoVs are very high because of constantly developing transcription errors and RNA Dependent RNA Polymerase (RdRP) jumps [2]. Initially CoVs were not considered pathogenic for humans until the severe acute respiratory syndrome (SARS) was found in the Guangdong state of China (2002 – 2003). two types of CoV such as CoV OC43, CoV 229E that have mostly caused mild infections in people with a responsive immune system [3, 4]. Approximately ten years after SARS, another highly pathogenic CoV, Middle East Respiratory Syndrome Coronavirus (MERS-CoV) had emerged in the Middle East countries [5]. In December 2019, the novel Coronavirus (nCoV), which caused another public health problem, has emerged in the Huanan Seafood Market, where livestock animals are generally traded in Wuhan State of Hubei Province in China and has been the focus of global attention due to a pneumonia epidemic of unknown etiology [6]. Primarily, an unknown pneumonia case was detected on December 12, 2019. Aftermath, Chinese authorities announced on January 7, 2020 that a new type of Coronavirus (novel Coronavirus, nCoV) was isolated [7]. This virus was named as 2019-nCoV by WHO on January 12 and COVID-19 on 11 February 2020. According to WHO, after quick spread in china in very short time span; globally 1,056,159 peoples are Corona positive with 57,206 deaths across 207 countries up to 4th April 2020. The probable initial infection was transmitted by zoonotic agent (from animal to human). The increase in the number of cases in Wuhan city of China and across world has showed a second transmission from human-to-human and no COVID-19 vaccine has been successfully developed yet. In the meantime, WHO declared it as pandemic on 11th March 2020.

Corona virus's genome structure is best among all RNA viruses. Two-thirds of RNA encodes viral polymerase (RdRp), RNA synthesis materials, and two large non-structural polyproteins, that are not reported involved in host response modulation (ORF1a-ORF1b). The other one-third of the genome encodes four structural proteins (spike (S), envelope (E), membrane (M) nucleocapsid (N)), and the other helper proteins. Although the length of the CoV genome shows high variability for ORF1a/ORF1b with four structural proteins, it is mostly associated with the number and size of accessory proteins [8, 9]. The first step in virus infection is the interaction of Spike Protein with human cells. The coronavirus spike protein is a multifunctional molecular machine that mediates coronavirus entry into host cells. It first binds to a receptor on the host cell surface through its S1 subunit and then fuses viral and host membranes through its S2 subunit. It's been reported that 2019-nCoV can infect the human respiratory epithelial cells through interaction with the human ACE2 receptor [10]. Genome encoding occurs after entering to the cell and facilitates the expression of the genes that encodes useful accessory proteins, which advances the adaptation of CoVs to their human host [9].

Immunoinformatics is a branch of bioinformatics primarily concerned with *in silico* analysis and modelling of immunological data and problems. It stresses the research mostly on the design and study of algorithms for mapping potential B- and T-cell epitopes, which shorten the time and lowers the cost needed for laboratory analysis of pathogen gene products. Several immunoinformatics tools are available for prediction and mapping of antigenic epitopes in protein sequence. It assists in designing subunit vaccines that starts from prediction of antigenic epitope through *in silico* techniques from

protein sequence of pathogens independent of their abundance [11, 12]. Evaluation of synthetic peptides as potential vaccine candidate for flavivirus has been investigated. Using the computational tools for prediction of epitopes and synthetic peptides from E glycoprotein of Murray Valley encephalitis (MVE) and DEN 2 viruses were prepared and their immunogenicity was evaluated in mice [13]. The identification of significant T-cells epitopes from secretory and cell surface proteins virulent proteins of M. tuberculosis H37Rv strain was done. The promiscuous nanomer candidate epitopes from HTL and CTL were recognized [14]. T-cell analyses of synthetic peptides to other viruses have correlated the association between T- and B-cell responses [15]. A new approach for vaccine design in immunology and the development of bioinformatics tools for T cell epitope prediction from primary protein sequences is essential. The primary focus of present study is to identify and map of the specific epitopes from five different proteins of SARS-CoV-2.

Material and Methods

Retrieve Target Sequence: The FASTA formatted amino acid sequences of SARS-CoV-2 were retrieved from the NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/) in this study. We have screened the specific proteins such as ORF1ab polyprotein, surface glycoprotein, membrane glycoprotein and nucleocapsid phosphoprotein nucleoprotein sequences were primarily selected for antigenicity prediction.

Physical properties of Protein Identification: To determine physical protein of SARS-CoV-2, the FASTA formatted amino acid sequences of total structural proteins were submitted to Generunner and ExPaSy (http://www.expasy.org). The expected

molecular weight and isoelectric point (pI) values were calculated.

Identification of T-Cell Epitope: The T-cell epitopes are typically peptide fragments which are immunodominant and can elicit specific immune responses, important for epitope-based peptide vaccine design. Due to the importance of T-cell epitopes, we used Propred (https://webs.iiitd.edu.in/raghava/propred/index.html) and Propred1 (https://webs.iiitd.edu.in/raghava/propred1/index.html) immunoinformatics tools which are available for prediction of epitopes in the protein primary sequences.

The server uses to read the input sequence, thus it can accept most commonly used standard sequence formats (FASTA). The sequence can be uploaded, from a file by using the cut and paste option. Users can customize these servers by selecting single/multiple allele, threshold and other parameter in order to achieve desirable results. The server analyses sequence data and generates output as text or graphics. These tools cover maximum number of human leukocyte antigen (HLA) comparison to other epitopes prediction tools. We have considered the parameters during epitopes prediction such as 3% threshold with maximum binding score to HLA molecules [16, 17].

Results

In the present study, four proteins of novel human coronavirus, SARS-CoV-2 were used for the physicochemical analysis such as molecular weight, isoelectric point (pI value) and antigenic nature. Nucleocapsid proteins showed highest molecular weight with 13.13 kDa and the lowest molecular weight with 3.93 kDa of surface glycoprotein. Isoelectric point of proteins was ranged between 4.46

to 9.98. The physicochemical properties of four proteins were given (Table 1). The pI value of protein is indicated the stability of protein at that particular pI.

Discussion

In recent years, many diseases have emerged due to the occurrence of several outbreaks through the different types of newer viruses. So, vaccine development against these emerging diseases within a short time is very crucial to protecting the people from the rising viral attacks. Vaccines are the pharmacological products which can provide the finest cost-benefit ratio in the prevention or treatment of diseases. However, an effective vaccine progression and production are costly and can take years to be completed. So, the researchers have tried for many years to minimize the cost and time for the development of vaccines. At this time, there are different strategies available for the design and development of effective and safe new-generation vaccines, based on the Bioinformatics approaches [18, 19]. The next-generation sequencing and progressive genomics and proteomics technologies have brought about a great change in computational immunology. However, the advancement of newer immunoinformatic tools has made a broader way in developing the vaccine or vaccine candidates through the satisfactory understanding of the immune response of the human body against an organism within a short time [20–22].

Table 1: Physicochemical properties of different proteins of SARS-CoV-2

Protein	Accession Number	Expected Molecular Weight (Da)	pI Value
ORF1ab polyprotein, partial	MN938386.1	10.64 kDa	9.87
Surface glycoprotein, partial	MN975266.1	3.93 kDa	4.46
Membrane glycoprotein, partial	MT008022.1	11.97 kDa	9.42
Nucleocapsid phosphoprotein, partial	LC523807.1	13.13 kDa	9.98

The epitope is recognizable by the immune system as a part of the antigen, and in particular by antibodies, B cells or T cells. The epitopes may belong to both foreign and self-proteins, and they can be categorized as conformational or linear, depending on their structure and integration with the paratope [23]. T-cell epitopes are presented on the surface of an antigen presenting cell (APC), where they are bound to major histocompatibility (MHC) molecules in order to induce immune response [24]. MHC class I molecules usually present peptides between 8 to 11 amino acids in length, whereas the peptides binding to MHC class II may have length from 12 to 25 amino acids [25]. If sufficient quantities of the epitope are presented, the T cell may trigger an adaptive immune response specific for the pathogen. Class II MHCs are expressed on specialized cell types, including professional APCs such as B cells, macrophages and dendritic cells, whereas class I MHCs are found on every nucleated cell of the body [26]. The recognition of epitopes by T cells and the induction of immune response have a key role for the individual's immune system. Even the slightest deviation from the normal functioning can have a grave impact on the organism. Knowledge about the peptide's epitopes has a key role for manufacturing epitope-based vaccines. One of the key issues in T-cell epitope prediction is the prediction of MHC binding, as it is considered a prerequisite for T

cell recognition. All T-cell epitopes are good MHC binders, but all good MHC binders are not T-cell epitopes. In the present study, an immunoinformatic-driven approach was incorporated to screen emergent immunogen against SARS-CoV-2 proteome. The results revealed that the SARS-CoV-2 of total 36 epitopes were predicted for class I MHC and 25 epitopes for class II MHC molecules in these proteins. Till date, no effective immunoinformatics study for the SARS-CoV-2 polyprotein has been performed for the identification of a potential vaccine target. However, we identified the potential T-cell epitopes from the all antigenic protein of SARS-CoV-2, as they play a key role in the creation of a defensive immune response against different pathogenic infections [27]. Various successful studies have been performed for the epitope based peptide vaccine design against West Nile virus [28], Zika virus [29], dengue virus [30], Chikungunya virus [31], Rift valley fever virus [32], shigellosis [33], and so on.

In this study, putative epitopes of ORF1ab polypeptide, surface glycoprotein, membrane glycoprotein and ucleocapsid phosphoprotein in SARS-CoV-2 were identified. Total 36 epitopes were predicted for class I MHC and 25 epitopes for class II MHC molecules in these proteins (Table 2).

Conclusion

Immunoinformatics is a newer strategy for identification and mapping of epitopes in the protein sequences of novel human coronavirus, SARS-CoV-2 exclusive of the virus culture. The predicted SARS-CoV-2 nanomer epitopes for T-cell is recognized against MHC Class II and MHC class I may be useful for development of sensitive, rapid and cost effective diagnosis. Further, these epitopes of SARS-CoV-2

may be served as vaccine candidates for the prevention of disease.

Funding

Conflict Of Interest

The authors report no conflicts of interest in this work.

Table 2: Most potential 36 T-cell epitopes with interacting MHC-I alleles and 25 T-cell epitopes with interacting MHC-II alleles epitope of SARS-CoV-2

Protein Name	T-cell epitopes	Amino acid position	No. of MHC Class II binding alleles	T-cell epitopes	Amino acid position	No. of MHC Class I binding alleles
ORF1ab polyprotein	VVIGTSKFY	67	05	FYGGWHNML	75	04
	YAISAKNRA	26	07	FAYTKRNVI	09	08
	MNLKYAISA	22	04	IPTITQMNL	17	08
	LFAYTKRNV	07	11	IAATRGATV	60	03
	WHNMLKTVY	78	03	SAKNRARTV	30	03
				TQMNLKYAI	21	03
				YEDQDALFA	02	03
				KFYGGWHNM	74	03
				TQMNLKYAI	21	03
Surface glycoprotein	IRGDEVRQI	08	20	KIADYNYKL	24	09
				FVIRGDEVR	07	03
				NVYADSFVI	01	05
				IAPGQTGKI	17	06
				SFVIRGDEV	06	03
				APGQTGKIA	18	05
				YADSFVIRG	03	03
Membrane glycoprotein	FVLAAVYRI	03	31	FIASFRLFA	35	04
	LVIGAVILR	76	20	FVLAAVYRI	04	03
	LRGHLRIAG	83	16	SYFIASFRL	33	03
	YRINWITGG	09	17	SFNPETNIL	50	03
	ILLNVPLHG	56	13	HLRIAGHHL	87	03
	FRLFARTRS	38	28	TRPLLESEL	69	03
	LRIAGHHLG	87	12	AAVYRINWI	07	04
	FIASFRLFA	34	05	IAIAMACLV	19	03
	FARTRSMWS	41	05	RPLLESELV	70	04
	FNPETNILL	50	04	WLSYFIASF	31	03
	VILRGHLRI	81	12	FNPETNILL	51	03
	YFIASFRLF	33	06	YFIASFRLF	34	03
	WLSYFIASF	30	08	IAMACLVGL	21	03
	ILTRPLLES	66	06			
Nucleocapsid phosphoprotein	YRRATRRIR	63	16	YYRRATRRI	63	03
	YYRRATRRI	62	15	SPRWYFYYL	82	09
	FYYLGTGPE	86	06	LGTGPEAGL	90	03
	WYFYYLGTG	84	03	FPRGQGVPI	43	09
	IGYYRRATR	60	07	SPDDQIGYY	56	03
				YGANKDGII	100	07
				LPNNTASWF	22	04

References:

1. Woo PC, Huang Y, Lau SK, et al. Coronavirus genomics and bioinformatics analysis. Viruses 2 (2010): 1804-1820.

2. Drexler JF, Gloza-Rausch F, Glende J, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. Journal of Virology 84 (2010): 11336-11349.

3. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology 23 (2018): 130-137.

4. Peiris JS, Lai ST, Poon LL, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. The Lancet 361 (2003): 1319-1325.

5. Zaki AM, Van Boheemen S, Bestebroer TM, et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. New England Journal of Medicine 367 (2012): 1814-1820.

6. Seven days in medicine. BMJ 368 (2020).

7. Imperial College London. Report 2: estimating the potential total number of novel coronavirus cases in Wuhan City, China. Jan 2020.
   https://www.imperial.ac.uk/mrc-globalinfectiousdisease-analysis/news--wuhan-coronavirus.

8. Luk HK, Li X, Fung J, et al. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infection, Genetics and Evolution (2019)

9. Coronavirinae in ViralZone. Available online: https://viralzone. expasy.org/785 (accessed on 05 February 2019).

10. Li F. Structure, function, and evolution of coronavirus spike proteins. Annual Review of Virology 3 (2016): 237-261.

11. Groot AS, Rappuoli R. Genome-derived vaccines. Expert review of vaccines 3 (2004): 59-76.

12. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 19 (2001): 2688-2691.

13. Gao XM, Liew FY, Tite JP. A dominant Th epitope in influenza nucleoprotein. Analysis of the fine specificity and functional repertoire of T cells recognizing a single determinant. The Journal of Immunology 144 (1990): 2730-2737.

14. Somvanshi P, Singh V, Seth PK. In silico prediction of epitopes in virulence proteins of Mycobacterium tuberculosis H37Rv for diagnostic and subunit vaccine design. J Proteomics Bioinform 1 (2008): 143-153.

15. Hu GJ, Wang RY, Han DS, et al. Characterization of the humoral and cellular immune responses against hepatitis C virus core induced by DNA-based immunization. Vaccine 17 (1999): 3160-3170.

16. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. Bioinformatics 17 (2001): 1236-1237.

17. Singh H, Raghava GP. ProPred1: prediction of promiscuous MHC Class-I binding sites. Bioinformatics 19 (2003): 1009-1014.

18. María RR, Arturo CJ, Alicia JA, et al. The impact of bioinformatics on vaccine design and development. InTech, Rijeka, Croatia; (2017).

19. Seib KL, Zhao X, Rappuoli R. Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clinical Microbiology and Infection 18 (2012): 109-116.

20. Groot AS, Rappuoli R. Genome-derived vaccines. Expert review of Vaccines 3 (2004): 59-76.

21. Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. PLoS Computational Biology 2 (2006).

22. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nature reviews Drug discovery 6 (2007): 404-414.

23. Huang J, Honda W. CED: a conformational epitope database. BMC Immunology 7 (2006): 7.

24. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annual Review of Immunology 13 (1995): 587-622.

25. Jardetzky TS, Brown JH, Gorga JC, et al. Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. Proceedings of the National Academy of Sciences 93 (1996): 734-738.

26. Janeway CA, Travers P, Walport M, et al. Immunobiology: the immune system in health and disease, 5th edn. Garland Science. New York (2001).

27. Esser MT, Marchese RD, Kierstead LS, et al. Memory T cells and vaccines. Vaccine 21 (2003): 419-430.

28. Larsen MV, Lelic A, Parsons R, et al. Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL. PloS one 5 (2010).

29. Alam A, Ali S, Ahamad S, et al. From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. Immunology 149 (2016): 386-99.

30. Chakraborty S, Chakravorty R, Ahmed M, et al. A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. In Silico Biology 10 (2010): 235-246.

31. Hasan MA, Khan MA, Datta A, et al. A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment. Molecular Immunology 65 (2015): 189-204.

32. Adhikari UK, Rahman MM. Overlapping CD8+ and CD4+ T-cell epitopes identification for the progression of epitope-based peptide vaccine from nucleocapsid and glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. Infection, Genetics and Evolution 56 (2017): 75-91.

33. Oany AR, Pervin T, Mia M, et al. Vaccinomics approach for designing potential peptide vaccine by targeting Shigella spp. serine protease autotransporter subfamily protein SigA. Journal of Immunology Research (2017).

Citation: Biswajit Sahoo, Krishna Kant, Neeraj K. Rai, Dharmendra Kumar Chaudhary. Identification of T-Cell Epitopes in Proteins of Novel Human Coronavirus, SARS-Cov-2 for Vaccine Development. International Journal of Applied Biology and Pharmaceutical Technology 11 (2020): 37-45.