

1

2 Bioinformatic identification of B-cell epitope for designing Multitope Peptide-Based Vaccine
3 against Covid-19
4

5 Zahra Rashidian¹, Mohammadreza Nassiri¹, Mohammad Hadi Sekhavati¹, Alireza Haghparast², Ali Javadmanesh^{1*}

6 ¹ Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran.

7 ² Department of Pathobiology, Faculty of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran.

8

9

Corresponding Author: javadmanesh@um.ac.ir

10

11 **Abstract**

12 **Introduction:** During the global outbreak of COVID-19, the crucial role of vaccination in enhancing
13 population immunity, limiting viral transmission, and reducing disease severity, hospitalizations, and
14 mortality became evident. The purpose of all vaccines is to expose the body to an antigen that will not cause
15 disease, but will stimulate the immune system. There are different vaccines against the corona virus. An
16 important aspect of COVID-19 vaccine development is the wide variety of technological platforms that
17 have been explored for producing effective immunization strategies, among which recombinant protein-
18 based vaccines are characterized by high protein expression efficiency, favorable long-term stability, and
19 the ability to elicit robust immune responses. The aim of this research was to design the first Multi-epitope
20 Peptide Vaccine (MPV) candidate targeting SARS-CoV-2. **Materials and Methods:** Protein sequences
21 were retrieved from public databases, and key epitope-rich regions were linked using rigid peptide linkers.
22 The physicochemical properties of the construct, including antigenicity, allergenicity, toxicity, stability,
23 half-life, and molecular features, were evaluated using bioinformatics tools. **Results:** Comprehensive in
24 silico analyses confirmed the construct's promising vaccine profile: it was predicted to be highly antigenic
25 (VaxiJen score up to 1.698), non-allergenic, non-toxic, and stable (instability index: 37.59). Structural
26 modeling and validation yielded a high-quality 3D model with 96.1% of residues in favored Ramachandran
27 regions and a ProSA Z-score of -2.46, indicative of native-like folding. **Conclusion:** These computational
28 results suggest that the designed multi-epitope construct holds strong potential to have a protective immune
29 response against COVID-19. Nevertheless, further laboratory and in-vivo investigations will be required
30 for validating its safety, immunogenicity, and protective potential.

31 **Keywords:** Bioinformatic Design, Multitope Peptide, Recombinant Protein, SARS-CoV-2, Vaccine
32 Development.
33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

1. Introduction

SARS-CoV-2, first reported in late 2019 and provisionally named 2019-nCoV, rapidly expanded across international borders, ultimately posing a substantial global public health risk [1]. Although numerous speculations have been made regarding how this disease began, its exact origin remains unknown. In a very short period, COVID-19 became a global pandemic and was identified as the most recent threat to public health since the influenza pandemic of 1918 [2]. Findings suggested that the initial transmission dynamics of 2019-nCoV between humans paralleled those documented during the 2002 SARS-CoV emergence. Mitigating the spread of the virus depends heavily on comprehensive international collaboration and public health coordination [3].

44 According to the World Health Organization, the most effective way to end this crisis is through a safe
45 and effective vaccine. The genetic sequence of SARS-CoV-2, the coronavirus responsible for COVID-19,
46 was published on January 11, 2020, marking a significant milestone for global research and development
47 efforts to produce a vaccine against this disease [4].

48 These viruses are single-stranded, enveloped, and contain a single-stranded RNA consisting of 29,891
49 nucleotides and 12 genes for the synthesis of non-structural and structural proteins of the virus [5]. The
50 virus has a diameter of 80-160 nm and is classified into four groups: Alpha, Beta, Delta, and Gamma
51 coronaviruses [6]. It also contains several non-structural proteins (NSPs) such as NSP13, NSP3, NSP12,
52 and NSP5, which are critical for their lifespan and pathogenicity [4]. The coronavirus has four main
53 structural proteins known as membrane protein (M), spike protein (S), envelope protein (E), and
54 nucleocapsid protein (N). The spike protein enables the virus to penetrate host cells and cause infection [7].
55 The nucleocapsid (N) protein encapsulates the viral RNA genome to form the ribonucleocapsid structure
56 [8]. Antibody levels targeting the nucleocapsid (N) protein - known to be the most abundant structural
57 protein of the virus - have been shown to rise. Although it is unlikely that N antibodies neutralize the virus,
58 a study has shown that they protect mice against hepatitis virus and coronavirus [6]. The formation and
59 assembly of the virus are carried out by the M protein, and the role of the E protein is to create pores for
60 ion transport and the formation of virions [9].

61 Coronaviruses are classified into four major evolutionary groups known as Alpha, Beta, Gamma, and
62 Delta. Only the Alpha and Beta branches contain strains that have adapted to infect humans and are linked
63 with respiratory illnesses. In contrast, viruses within the Gamma and Delta groups are largely associated
64 with animal hosts and seldom involve human infection. SARS-CoV-2 belongs to the Betacoronavirus genus
65 and is associated with respiratory, gastrointestinal, and neurological manifestations in humans, collectively
66 referred to as Coronavirus Disease 2019 (COVID-19) [10].

67 In addition to these proteins, five other proteins, including ORF7, ORF6, ORF8, ORF3a, and ORF9, are
68 also present in the structure of the virus. These proteins assist in interactions with host cells, creating
69 favorable conditions for viral replication, modifying host gene induction, and neutralizing the host's
70 antiviral defense system [1].

71 The earliest COVID-19 vaccine candidate advanced rapidly into human clinical trials on March 16, 2020.
72 By April 8, 2020, a total of 115 vaccine candidates had been registered worldwide, of which 78 were under
73 active development and 37 had been discontinued. The reported efficacy of the authorized vaccines ranged
74 from approximately 70% to 90%. Consequently, only a small proportion of vaccinated individuals exhibited
75 breakthrough infections, while mortality rates were reduced by nearly 10%. These findings underscore the
76 crucial role of vaccination in mitigating the world-wide impact of COVID-19 [11].

77 A wide range of genetic factors were evaluated for the production of COVID-19 vaccines, including
78 nucleic acid-based vaccines (DNA or RNA), virus-like particles, viral vector systems (both non-replicating
79 and replicating), peptide vaccines, recombinant protein formulations, as well as live-attenuated and
80 inactivated virus platforms. In general, four strategies have been used for vaccine production: Based on
81 traditional models or second the use of weakened or inactivated viruses. Eight vaccines have been
82 developed using this approach, including the Iranian vaccine (COV-Iran Barekat), the Indian vaccine
83 (Bharat), and the Chinese vaccine (Sinopharm) [11].

84 The second vaccination strategy employs viral vector platforms, through which sixteen vaccines have
85 been developed worldwide, including several from China. The third approach is based on nucleic acid
86 technologies (DNA or RNA), with approximately sixteen vaccines designed using this model—most
87 notably the Moderna and Pfizer vaccines developed in the United States. The fourth strategy utilizes viral

88 protein subunits or recombinant proteins. Vaccines from France, the Iran–Cuba Pasteur Institute (Sanofi),
89 and the United States (Novavax) follow this approach. Several FDA-approved vaccines require strict
90 temperature regulation; factors such as improper cold-chain management, transportation issues, or
91 inadequate storage conditions can compromise vaccine efficacy and result in the loss of immunization
92 capacity [12].

93 While traditional vaccine platforms including inactivated viruses, viral vectors, and protein subunits have
94 been instrumental in pandemic control, they often face challenges related to complex manufacturing, cold-
95 chain requirements, and a focus on single antigenic targets that may be susceptible to immune evasion by
96 viral variants [7, 12]. In contrast, the multi-epitope peptide vaccine (MPV) strategy represents a next-
97 generation, rational design approach. MPVs are engineered by bioinformatic selection of immunodominant
98 epitopes, which are then chemically synthesized or recombinantly expressed as a single polypeptide. This
99 strategy offers distinct advantages: it bypasses the need for culturing live pathogens, allows for precise
100 targeting of conserved regions across multiple viral proteins, and can be rapidly adapted in response to
101 emerging variants. Critically, by presenting the immune system with a curated pool of key antigenic
102 signatures without the pathogen's replication machinery, an MPV can initiate a focused and accelerated
103 immune response, effectively "outpacing" the natural infection timeline to establish protective immunity
104 before the pathogen can establish a significant stand [13, 14].

105 The efficacy of a vaccine hinges on its ability to stimulate both arms of the adaptive immune system:
106 humoral immunity, mediated by B cells and neutralizing antibodies, and cellular immunity, driven by helper
107 (CD4+) and cytotoxic (CD8+) T cells [15]. The MPV platform is uniquely suited for this coordinated
108 activation through targeted immunogen design. By integrating well-characterized B-cell and T-cell epitopes
109 into a single construct, an MPV can be engineered to simultaneously elicit robust antibody production for
110 viral neutralization and potent T-cell responses for the clearance of infected host cells. This targeted
111 approach minimizes the inclusion of non-essential or immunosuppressive viral sequences, thereby
112 enhancing vaccine safety by reducing the risk of adverse reactions. Furthermore, the selection of conserved,
113 immunodominant epitopes aims to provide broad and durable protection, potentially overcoming the
114 limitations of strain-specific immunity. Therefore, a rationally designed MPV that synergistically activates
115 humoral and cellular pathways represents a promising strategy for developing safe, effective, and broadly
116 protective vaccines against pathogens like SARS-CoV-2 [8, 13].

117 At present the bioinformatic approach has been demonstrated to be effective in designing vaccines,
118 including for COVID-19 vaccines and other infectious diseases [16]. This study utilized computational
119 analyses to develop a multi-epitope peptide vaccine (MPV) candidate against SARS-CoV-2. The design
120 strategically incorporated T- and B-cell epitopes derived from the virus's nucleocapsid (N), spike (S), and
121 membrane (M) proteins, aiming to maximize immune response potential.

122

123 **2- Materials and Methods**

124

125 **2.1. Acquisition of Protein Sequences**

126 The sequences of the SARS-CoV-2 structural proteins (in FASTA format), including the Spike (S)
127 glycoprotein (RefSeq: A0A679G9E9), Nucleocapsid (N) protein (RefSeq: A0A6C0T6Z7), Membrane (M)
128 protein (RefSeq: P0DTC5), and Envelope (E) protein (RefSeq: P0DTC4), were retrieved from the National
129 Center for Biotechnology Information (NCBI) and UniProt databases (<https://www.uniprot.org>). Epitopic
130 regions of B cells were selected using online software. To preserve structural integrity, glycine–proline (GP)
131 linkers were applied, and the three-dimensional (3D) configuration of the final construct was generated and
132 validated using bioinformatics modeling tools (Figure 1).

133

134
135
136
137
138
139
140
141

Figure 1: Nucleotide sequence of the designed structure. Green: nucleotide sequence from Spike protein. Purple: nucleotide sequence of membrane protein. Red: nucleotide sequence of the coating protein. Aqueous: nucleotide sequence of nucleocapsid protein. Black: linkers.

```
VRFPNITNLCPFGEVFNGPSPTKLNDLCFTNVYGP
TESNKKFLPFQQFGRDIAGPVTLACFVLAAYRIN
GPLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYS
RVKNLNSSRGPiRQGTDYKHWPQIAQFAPSAAFFG
MSRIGMEVTPSGTWLTYTGAIKL
```

142
143
144
145
146
147
148
149
150
151
152
153
154

Figure 2: The amino acid sequence of the designed construct: Green: Selected epitope from the spike protein. Purple: Selected epitope from the membrane protein. Red: Selected epitope from the envelope protein. Blue: Selected epitope from the nucleocapsid protein. Black: Linkers.

2.2. B-cell epitope prediction

155
156
157
158
159
160
161
162
163
164
165
166
167
168

An indispensable step in the design of a multiepitopic protein for vaccine production or the construction of a serological test is the prediction of Linear B cell epitopes. B-cell epitopes represent distinct antigenic regions accessible on the surface of a pathogen, which can be specifically recognized and bound by antibodies, ultimately leading to the activation of the host immune system [17]. To predict linear epitopes for the four main structural proteins of the coronavirus, the ABCPred (<http://www.imtech.res.in/raghava/abcpred>) and IEDB (<http://tools.iedb.org/conservancy>) databases were used. The highest predictive performance in ABCPred was achieved using a 16-residue peptide window during model configuration. This online software utilizes a neural network algorithm (with a threshold of 0.51) for predicting linear epitopes.

2.3. T-cell Epitope Prediction

169
170
171
172
173
174
175
176
177
178

The analysis employed a combined methodology, based on established computational epitope prediction strategies, utilizing the IEDB epitope analysis resource (https://www.iedb.org/home_v3.php) for MHC class I and II binding predictions. Specifically, CD8+CTL epitopes from the DENV-4 antigenic components were predicted using the NetMHCpanEL method, constrained by the HLA-A*11:01 allele. These results were subsequently filtered using the SMM method based on an IC50 threshold of 500 nM and a percentile rank ≤0.2 to prioritize CTL candidates. Concurrently, CD4+HTL epitopes (15-mer) were predicted using NetMHCIIpanEL for the HLA-DRB1*04:01 allele. Finally, all predicted MHC classes were screened for antigenicity, non-allergenicity, non-toxicity, high conservancy, and appropriate transmembrane topology.

2.4 Protein Construct Design

179
180
181

Th GTTAGATTTCCTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAACGGTCCATCTCCTAC 1e-proline
TAAATTAATGATCTCTGCTTTACTAATGTCTATGGTCCAACCTGAGTCTAACAAAAAGTTTCTG
(GP CCTTTCCAACAATTTGGCAGAGACATTGCTGGTCCAGTAACTTTAGCTTGTTTTGTGCTTGCTG
CTGTTTACAGAATAAATGGTCCATTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATTGTGT

182
183
184
185
186

2.5 A f TAAAAATCTGAATTCTTCTAGAGGTCCAATCAGACAAGGAAGTATTACAAACATTGGCCGCA 1g the 3D
conf AATTGCACAATTTGCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGT 1n the field
CACACCTTCGGGAACGTGGTAAAAGCTT

of computational structural biology. In this study, for the 3D structure of the construct, the online server GalaxyTEM (template-based modeling) (<https://galaxy.seoklab.org>) was used. This site uses template-

187 based models of homologous proteins as templates to predict the structure of the target protein. Finally, the
188 quality and reliability of the generated model were assessed using the online software PROCHECK
189 (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK>) and ProSA
190 (<https://prosa.services.came.sbg.ac.at/prosa.php>). The ProSA server indicates the quality of the generated
191 model by calculating the Z-score, which is a metric that reflects the overall quality of the model. This
192 platform evaluates the performance of a given model by comparing its score against those derived from all
193 reference structures available in the Protein Data Bank (PDB). Additionally, the quality of the generated
194 models was further assessed using the Ramachandran plot through the PROCHECK server [18]. The
195 Ramachandran diagram is a two-dimensional plot used to visualize the bond angles in terms of energy and
196 stability, represented by the values of the ω and ϕ angles in the protein structure [19].
197

198 **2.6. Evaluation of the Antigenicity of the Selected Epitope**

199 The antigenicity of the selected epitope is assessed to determine its potential to trigger an immune
200 response. This is done using computational tools and algorithms that predict how likely a peptide is to be
201 recognized by the immune system. For predicting the antigenicity of the epitope subunits, the online server
202 VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) was used. Default parameters
203 of this server were applied to determine the antigenic protein. According to studies, the accuracy of this
204 software for viral models has been reported to be 87% [20].
205

206 **2.7. Determination of the Allergenicity of the Selected Epitope:**

207 The allergenicity of a selected epitope is an essential factor to consider when designing vaccines, as it
208 helps ensure that the epitope does not provoke allergic reactions in individuals. The AllerTop v.2.0 online
209 server (<https://www.ddg-pharmfac.net/AllerTOP/data.html>) was used to predict the allergenicity of the
210 proposed epitope for the design of an immunogenic construct. This server is the first free server capable of
211 predicting the allergenic potential of target antigens [21]. The multi-epitope sequences that exhibited the
212 maximum antigenicity among the different peptide positioning combinations were selected.
213

214 **2.8. Evaluation of the Toxicity of the Predicted Epitopes**

215 Assessing the toxicity of predicted epitopes is an essential step in vaccine design to ensure that the selected
216 epitopes do not pose harmful effects to the host. Several computational tools and databases are available to
217 predict the toxicity of peptides. To evaluate the toxicity of the epitopes, the ToxinPred server
218 (<http://crdd.osdd.net/raghava/toxinpred>) was utilized. This server is designed to identify whether short
219 epitopes are toxic or non-toxic. This tool predicts the toxic potential of peptides by considering factors such
220 as amino acid composition, structure, and sequence similarity to known toxins [22].
221

222 **2.9. Prediction of Stability and Half-Life**

223 To assess the stability and half-life of epitopes, computational tools and bioinformatics resources are used
224 to estimate their properties under different biological conditions. The ProtParam tool predicts the half-life
225 of peptides in different biological systems, including: *In vitro* (mammalian cells): Typical half-life
226 prediction is based on N-terminal rule degradation. *In vivo* (bacteria or yeast): Predicted half-life is based
227 on known experimental data from similar sequences [23]. The stability coefficient and half-life under *in*
228 *vivo* and *in vitro* conditions were assessed using the ProtParam tool from the ExpASy on-line server
229 (<https://web.expasy.org/protparam>).
230

231 **2.10. Evaluation of Physicochemical Properties**

232 The physicochemical parameters of the vaccine construct were evaluated using the PEPCALC tool
233 (<https://pepcalc.com>) and ProtParam (<https://web.expasy.org/protparam/>). Properties such as amino acid
234 composition and charge, molecular weight, polarity, half-life, extinction coefficient, instability index, and
235 water solubility were analysed [24].
236

237 **3-Results**

238
239
240
241
242
243
244
245

3.1. Analysis of Predicted Linear B-Cell Epitopes

Based on ABCPred analyses, linear epitopes were predicted. The nucleotide sequence of each epitope, sequence length, and structural protein associated with each sequence are presented in Table 1. Additionally, the amino acid sequences, start and end positions of each epitope, along with the corresponding structural protein, are shown in Table 2. In the final stage, a multi-epitope protein was engineered by joining the selected epitopic regions with suitable linkers. GP linkers were employed to connect these regions, resulting in the assembly of the complete multi-epitope structure.

246

247

248 Table 1: Nucleotide Sequences of the Designed Construct from Different Regions of the SARS-CoV

Nucleotide Sequence of Epitopes	Length of Sequence (bp)	Structural Protein
GTTAGATTTCTTAATATTACAAACTTGTGCCCTTTTGGTGAAGTTTTTAAC	51	Spike
TCTCCTACTAAATTAATGATCTCTGCTTTACTAATGTCTAT	42	Spike
ACTGAGTCTAACAAAAAGTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGCT	54	Spike
GTAACCTTTAGCTTGTTTTGTGCTTGCTGCTGTTTACAGAATAAAT	45	Membrane
TTGCTAGTTACTAGCCATCCTTACTGCGCTTCGATTGTGTGCGTACTGCTGCAATATTGTT	129	Envelope
AACGTGAGTCTTGTA AACCTTCTTTTTACGTTTACTCTCGTGTTAAAAATCTGAATCTTCTAGA		
ATCAGACAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTTGCCCCAGCGCT	126	Nucleocapsid
TCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTCGGGAACGTGGTAAAAGCTT		

249

250

251 Table 2: Selected B-Cell Epitopes from Studied Proteins Predicted by ABCPred

Amino Acid Sequence of Epitopes	Start Position	End Position	Structural Protein
VRFPNITNLCPPGGEVFN	327	343	Spike
SPTKLNLCFTNVY	382	396	Spike
TESNKKFLPFQFGRDIA	553	569	Spike
VTLACFVLAAYVRIN	59	74	Membrane
LLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKLNLSRR	27	69	Envelope
IRQGTDYKHWPOIAQFAPSAAFFGMSRIGMEVTPSGTWLTYTGAIKL	292	339	Nucleocapsid

252

253

3.2. T-cell Epitope Prediction and Prioritization

The analysis identified potential T-cell epitopes across the selected viral components. The final prioritized list of predicted epitopes for both MHC Class I (CD8+) and MHC Class II (CD4+) binders is presented in Table 3. The initial prediction utilized established computational epitope analysis resources, including the IEDB and CTLPred servers, constrained by the specific HLA alleles (HLA-A*11:01 and HLA-DRB1*04:01). Subsequent prioritization involved stringent, in-house filtering criteria applied to all predicted candidates. These selection rules included: favorable IEDB scores, high predicted conservancy, strong binding affinity ($IC_{50} \leq 500$ nM and percentile rank ≤ 0.2), minimum lengths (≥ 9 -mer for MHC I and 15-mer for MHC II), documented antigenicity/immunogenicity, and favorable topological accessibility. Following these rigorous specifications, a total of 6 epitopes were shortlisted for MHC I binders and 6 epitopes for MHC II binders across the analyzed viral components for the multi-epitope vaccine design.

264

265
266
267
268
269
270
271
272

Table 3: Predicted T-Cell Epitopes from Studied Proteins Predicted by IEDB epitope analysis

Peptide	MHC Class	Region	Start	End	IC50 (nM)	HLA Alleles
GPVRFNIT	I	N-terminal	1	9	25	HLA-A*03:01
VRFPNITNL	I	N-terminal	3	11	28	HLA-B*57:01
YGPTESNKK	I	Central	46	54	12	HLA-A*01:01, B*35:01
LCAVYRINGP	I	Central	69	77	8	HLA-A*02:01, B*07:02
CCNIVNVSL	I	C-terminal	117	125	38	HLA-C*07:01
NIVNVSLVK	I	C-terminal	119	127	41	HLA-B*27:05
PVRFPNITNLCPFG	II	N-terminal	2	17	22	DRB1*01:01
FGEVFNGPSPTKL	II	N-terminal	10	22	24	DRB1*03:01
ACFVLAAYRINGP	II	Central	61	73	18	DRB1*15:01
LCAVYRINGPLL	II	Central	69	81	10	DRB1*04:01, DQB1*03:02
IYSRVKNLNSSRG	II	C-terminal	133	145	35	DRB1*11:01
KPSFYVYSRVKNL	II	C-terminal	137	149	40	DQB1*06:02

273
274

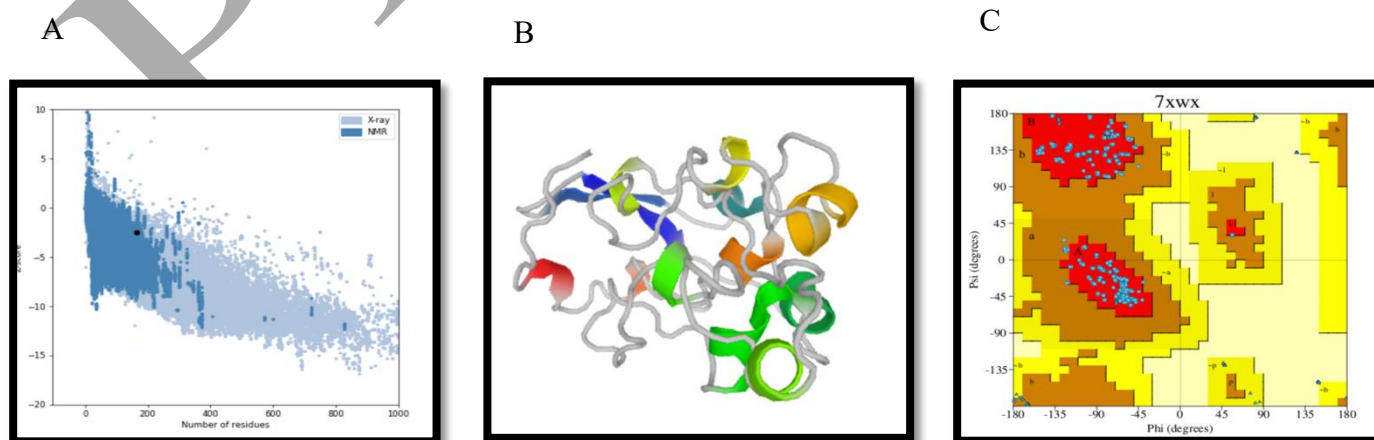
3.3. Generation of three-dimensional structures of epitopes

Five three-dimensional (3D) structures were predicted using the GalaxyTEM server. The quality of the models was then analyzed using the ProSA and Ramachandran Plot servers to evaluate the third model and select the best structure (Figure 3A).

ProSA evaluates the overall quality of a protein model. Positive scores indicate potential errors in the generated 3D structures, whereas the most negative Z-score corresponds to the highest structural integrity of the designed tertiary construct [18]. The highest negative score among the five predicted structures was -2.46 (model 3), which falls within the range of characteristic scores for native crystallized proteins of similar size (dark blue). This result validates the structural models generated by the GalaxyTEM software (Figure 3B).

The Ramachandran plot also revealed that 0%, 3.9%, and 96.1% of the amino acid residues are located in the outlier region, allowed region, and favored region, respectively (Figure 3C).

275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298



299 **Figure 3.** A) The structural integrity of the 3D crude model was evaluated using the ProSA-web server to detect potential and
300 quality errors. The model exhibited a Z-score of -2.46 , which is consistent with the range observed for naturally occurring protein
301 structures. B) The final vaccine construct was refined using the GalaxyRefine server, which produced five three-dimensional
302 models. Among these, the third model, displaying the highest refinement score, was selected for further analysis and visualized
303 using YASARA software. C) Ramachandran plot assessment of the selected 3D model showed that the vast majority of bond angles
304 fall within allowed regions, indicating a reliable stereochemical quality.

305
306

307 **3.4. Antigenicity Analysis of the Selected Epitope**

308 The peptide epitopes were analyzed for their antigenic potential using the VaxiJen v2.0 online platform.
309 Among the predicted epitopes, six with the most desirable antigenicity scores were selected, ranging from
310 a minimum of 0.6241 to the highest score of 1.6982 . The predicted scores indicate that the antigenic nature
311 of the protein sequence has successfully created an immunogenic construct. Therefore, this sequence was
312 used for designing an epitope-based immunogenic vaccine [20].
313

314 **3.5. Allergenicity Evaluation of the Selected Epitope**

315 Each of the predicted epitopes was individually assessed for allergenicity using the AllerTop v.2.0
316 software. Those with an allergic risk were removed, and those lacking allergenic properties were selected
317 for further synthesis to create the immunization construct for subsequent analysis.
318

319 **3.6. Toxicity Evaluation of the Epitopes**

320 The ToxinPred server, one of the best bioinformatics tools, was used to predict the toxicity of the target
321 antigens. With the default parameters of this server, all the predicted epitopes were found to be non-toxic.
322

323 **3.7. Prediction of Stability and Half-life**

324 The resulting sequence, consisting of 166 amino acids and a molecular weight of $18,248.28$ Daltons, was
325 proposed as a multi-epitope vaccine candidate. The half-life of this multi-epitope that predicted by
326 ProtParam was 30 hours in vitro and 10 hours in vivo.
327

328 **3.8. Evaluation of Physical and Chemical Properties**

329 ProtParam was employed to evaluate the physicochemical properties of the engineered protein. The
330 analysis estimated the molecular weight of the construct at ~ 18 kDa and its theoretical isoelectric point (pI)
331 at 9.68 . The calculated pI value indicated that the protein exhibits an overall alkaline character. The
332 construct comprises 166 amino acid residues. The protein exhibited an instability index (II) of 37.59 ,
333 suggesting that it is probably stable under physiological conditions. Furthermore, the aliphatic index and
334 GRAVY value were 88.67 and 0.190 , respectively, reflecting the hydrophilic nature of the construct and its
335 potential for favorable interactions with water molecules.
336

337 **4. Discussion**

338 The COVID-19 pandemic underscored the critical need for next-generation vaccine platforms that can be
339 rapidly deployed, cost-effectively manufactured, and designed to induce broad, sustained protection against
340 emerging variants. The design and development of vaccines involve various stages, from identifying target
341 epitopes to evaluating their immunogenicity, antigenicity, stability, and allergenicity. In this context,
342 immunoinformatics offers a highly streamlined approach, moving beyond traditional vaccine development
343 to target only the most potent immunodominant epitopes [4].

344 In this study, a comprehensive bioinformatics strategy was employed, integrating various computational
345 tools to develop a multi-epitope vaccine targeting SARS-CoV-2, specifically focusing on structural proteins
346 pivotal to viral infectivity and pathogenicity.

347 The designed vaccine comprises amino acid sequences related to the Receptor-Binding Domain (RBD)
348 of SARS-CoV-2, including the S2, S1, E, N, and M regions. The protein sequences of SARS-CoV-2 were
349 obtained from the NCBI database, and the optimal regions were linked together using linkers to construct
350 the appropriate vaccine design. This multi-epitope strategy is inherently superior to first-generation
351 vaccines that focused solely on the Spike protein's receptor-binding domain, which is prone to mutation
352 and subsequent immune evasion [13]. By incorporating conserved epitopes from the N and M proteins, the
353 MPV design promotes the induction of cytotoxic T-helper (Th) and T lymphocytes (CTLs) cells, which are
354 crucial for clearing viral-infected cells and establishing long-term cellular immunity, a feature vital for
355 cross-protection against SARS-CoV-2 [13].

356 Precisely controlling the spatial arrangement and orientation of protein domains to optimize the functional
357 performance of the target protein remains a significant challenge in protein engineering [3]. The structural
358 integrity and potential immunogenicity of the designed MPV construct were rigorously assessed through
359 comprehensive in silico validation methods. In most cases, the direct linkage of functional peptides without
360 a linker can lead to numerous undesirable outcomes, including improper folding of the fusion protein,
361 disruption of biological activity, or low protein production yield [25]. Linkers, in addition to their essential
362 role in connecting functional domains (such as flexible linkers) or releasing functional domains in vivo
363 (such as cleavable linkers), may offer many other advantages for the production of recombinant proteins.
364 These include improving biological activity, enhancing expression yield, and achieving desirable
365 pharmacokinetic profiles [14]. In this study, the selected regions of linear B-cell epitopes were fused using
366 GP linkers, resulting in the construction of a novel multi-epitope structure. B-cell epitopes are regions on
367 the surface of an antigen, to which specific antibodies recognize and bind, triggering the immune response,
368 this interaction is at the core of the adaptive immune system [15].

369 This structural optimization was confirmed by the predicted tertiary model's quality. Analysis of the
370 Ramachandran plot demonstrated that 96.1% of the residues were located in the most favored regions,
371 which is a strong indicator of a stable, folded protein structure similar to a native biological molecule.
372 Complementary analysis using the ProSA server yielded a Z-score of -2.46, falling well within the
373 acceptable range for experimental proteins, further validating the thermodynamic stability and quality of
374 the predicted 3D structure. Such robust bioinformatics evidence provides a strong foundation for the
375 candidate's stability, which is a key advantage for recombinant protein platforms over more labile vaccine
376 formats [5]. The pI value of the construct was approximately 18 kDa, indicating its alkaline nature.
377 Additionally, a high aliphatic index of 88.67 suggested that the designed protein is thermostable, while the
378 instability index of 37.59 indicated that the construct is stable. Based on the GRAVY score, the construct
379 was hydrophilic, demonstrating its ability to interact easily with water molecules.

380 The selected epitopes were analyzed for their antigenicity using the VaxiJen 2.0 server, which confirmed
381 their potential to trigger an immune response. Six epitopes with the highest antigenicity scores, ranging
382 from 0.6241 to 1.6982, were chosen for further studies. These scores demonstrated that the designed protein
383 construct possesses an immunogenic nature, making it a viable candidate for vaccine development.
384

385 5. Conclusion

386 In fact, by examining the physicochemical properties, it was determined that the designed construct is
387 hydrophilic, heat-resistant, alkaline, has a long half-life, and is stable, making it suitable for use in
388 preventing COVID-19. Despite the promising outcomes obtained from in silico analyses, additional in vivo
389 and in vitro investigations are required to validate the safety, effectiveness, and immunogenic potential of
390 the vaccine against SARS-CoV-2. The successful in silico validation now provides a clear pathway for
391 moving to the next stage of preclinical investigation, involving gene synthesis, expression in a suitable host
392 system, and comprehensive in vivo immunogenicity and challenge studies to confirm the protective efficacy
393 of this promising vaccine candidate.

394 **Acknowledgments**

395 This study was supported by Ferdowsi University of Mashhad, grant number 3/59004.

396 **Ethics**

397 Not Applicable.

398 **Conflict of Interest**

399 The authors declare that there is no conflict of interest.

400 **Data Availability**

401 All data and supplementary information would be available upon request from the corresponding
402 author.

403 **Authors' Contribution**

404 ZR: Drafting of the manuscript, Analysis and interpretation of data. MN, MHS and AH:
405 Administrative, technical, and material support. AJ: Study concept and design, Study
406 supervision, review, and approval of the manuscript.

407 **Funding**

408 This research funded by Ferdowsi University of Mashhad.

409 **References**

- 410 1- Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2
411 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges.
412 International journal of antimicrobial agents. 2020 Mar 1;55(3):105924. DOI:
413 10.1016/j.ijantimicag.2020.105924
- 414 2- Imai N, Gaythorpe KAM, Abbott S, et al. Adoption and impact of non-pharmaceutical interventions
415 for COVID-19. Wellcome Open Res. 2020;5:59. DOI: 10.12688/wellcomeopenres.15808.1
- 416 3- Riou J, Althaus CL. Pattern of early human-to-human transmission of Wuhan 2019 novel
417 coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance. 2020 Jan
418 30;25(4):2000058. DOI: 10.2807/1560-7917.ES.2020.25.4.2000058
- 419 4- Yang L, Liu S, Liu J, Zhang Z, Wan X, Huang B, et al. COVID-19: immunopathogenesis and
420 Immunotherapeutics. Signal transduction and targeted therapy. 2020 Jul 25;5(1):128. DOI:
421 10.1038/s41392-020-00243-2
- 422 5- Rabi FA, Al Zoubi MS, Kasasbeh GA, Salameh DM, Al-Nasser AD. SARS-CoV-2 and coronavirus
423 disease 2019: what we know so far. Pathogens. 2020 Mar 20;9(3):231. DOI:
424 10.3390/pathogens9030231.
- 425 6- Abdoli A, Alizadeh R, Aminianfar H, Kianmehr Z, Teimoori A, Azimi E, et al., Safety and potency
426 of BIV1-CovIran inactivated vaccine candidate for SARS-CoV-2: A preclinical study. Reviews in
427 medical virology. 2022 May;32(3):e2305. DOI: 10.1002/rmv.2305

- 428 7- Pollet J, Chen WH, Strych U. Recombinant protein vaccines, a proven approach against
429 coronavirus pandemics. *Advanced drug delivery reviews*. 2021 Mar 1;170:71-82. DOI:
430 10.1016/j.addr.2021.01.001
- 431 8- Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. Immunoinformatic identification of B
432 cell and T cell epitopes in the SARS-CoV-2 proteome. *Scientific Reports*. 2020 Aug
433 25;10(1):14179. DOI: 10.1038/s41598-020-70864-8
- 434 9- Basiri N, Koushki M. Study of Vaccine Production Abroad and Scientific and Research Challenges
435 of COVID-19 Vaccine Production in Iran. *Annals of the Romanian Society for Cell Biology*. 2021
436 Apr 1;25(4). <http://annalsofrscb.ro/index.php/journal/article/view/7528>
- 437 10- Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. *Viruses*. 2020 Jan
438 24;12(2):135. DOI: 10.3390/v12020135
- 439 11- Abdulla ZA, Al-Bashir SM, Al-Salih NS, Aldamen AA, Abdulazeez MZ. A summary of the SARS-
440 CoV-2 vaccines and technologies available or under development. *Pathogens*. 2021 Jun
441 22;10(7):788. DOI: 10.3390/pathogens10070788
- 442 12- Callaway E. The race for coronavirus vaccines: a graphical guide. *Nature*. 2020;580(7805):576-
443 577. DOI: 10.1038/d41586-020-01221-y
- 444 13- Khan MS, Khan IM, Ahmad SU, Rahman I, Khan MZ, Khan MS, et al. Immunoinformatics design
445 of B and T-cell epitope-based SARS-CoV-2 peptide vaccination. *Frontiers in Immunology*. 2023
446 Jan 4;13:1001430. DOI: 10.3389/fimmu.2022.1001430
- 447 14- Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. *Advanced*
448 *drug delivery reviews*. 2013 Oct 15;65(10):1357-69. DOI: 10.1016/j.addr.2012.09.039
- 449 15- Boehm T. Design principles of adaptive immune systems. *Nature Reviews Immunology*. 2011
450 May;11(5):307-17. DOI: 10.1038/nri2944
- 451 16- Alabboud M, Javadmanesh A. In silico study of various antiviral drugs, vitamins, and natural
452 substances as potential binding compounds with SARS-CoV-2 main protease. *DYSONA-Life*
453 *Science*. 2020 Jul 1;1(2):44-63. DOI: 10.30493/DLS.2020.225404
- 454 17- Van Regenmortel MH. What is a B-cell epitope?. In *Epitope Mapping Protocols: Second Edition*
455 2009 Feb 24 (pp. 3-20). Totowa, NJ: Humana Press. DOI: 10.1007/978-1-59745-450-6
- 456 18- Sadeghpour S, Ghasemnejad-Berenji H, Malekzadegan Y, Rezaei Arablouydareh S, Fathi J,
457 Sabaghan M, et al. Design and Bioinformatic Evaluation of Multi-Epitope Protein to Diagnose
458 Strongyloidiasis Infection. *Journal of Mazandaran University of Medical Sciences*. 2023 Dec
459 10;33(2):222-31. <http://jmums.mazums.ac.ir/article-1-19673-en.html>
- 460 19- Hashemi Attar, M., Nassiry, M. R., Javadmanesh, A., Nayeri Fasaei, B. Structural analysis of
461 aspartate kinase enzyme in *Corynebacterium* species to find the best enzyme structure for high
462 industrial production of lysine amino acid. *Iranian Journal of Animal Science Research*, 2019;
463 11(4): 501-512. DOI: 10.22067/ijasr.v11i4.79758
- 464 20- Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour
465 antigens and subunit vaccines. *BMC bioinformatics*. 2007 Jan 5;8(1):4. DOI: 10.1186/1471-2105-
466 8-4
- 467 21- Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. *BMC*
468 *bioinformatics*. 2013 Apr 17;14(Suppl 6):S4. DOI: 10.1186/1471-2105-14-S6-S4
- 469 22- Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R. Open Source Drug Discovery Consortium,
470 Raghava GP. In silico approach for predicting toxicity of peptides and proteins. *PloS one*. 2013 Sep
471 13;8(9):e73957. DOI: 10.1371/journal.pone.0073957
- 472 23- Gonda DK, Bachmair A, Wüning I, Tobias JW, Lane WS, Varshavsky A. Universality and
473 structure of the N-end rule. *Journal of Biological Chemistry*. 1989 Oct 5;264(28):16700-12.
474 <https://pubmed.ncbi.nlm.nih.gov/2506181/>
- 475 24- Mousavi, Z., Sekhavati, M. H., Farzaneh, M., Javadmanesh, A. Investigating of the binding energy
476 of DEP-A and DEP-B enzymes with DON mycotoxin chemotype by molecular docking. *Veterinary*
477 *Research & Biological Products*, 2024; 37(1): 15-26. DOI: 10.22092/vj.2023.362064.2062

478 25- Ehsasatvatan M, Baghban Kohnehrouz B. Effect of linker's length and sequence on the structure
479 and stability of mGLP-1-DARPin fusion protein for treatment of type 2 diabetes: A computational
480 study. Yafteh. 2023 Aug 10;25(2):26-47. <http://yafte.lums.ac.ir/article-1-3625-en.html>

Preprint