

RESEARCH ARTICLE



# In silico SARS-CoV-2 vaccine development for Omicron strain using reverse vaccinology

Vladimir Li<sup>1</sup> · Chul Lee<sup>1</sup> · DongAhn Yoo<sup>1</sup> · Seoae Cho<sup>3</sup> · Heeбал Kim<sup>1,2,3</sup>

Received: 17 January 2022 / Accepted: 31 March 2022 / Published online: 4 June 2022  
© The Author(s) 2022

## Abstract

**Background** The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic began in 2019 but it remains as a serious threat today. To reduce and prevent spread of the virus, multiple vaccines have been developed. Despite the efforts in developing vaccines, Omicron strain of the virus has recently been designated as a variant of concern (VOC) by the World Health Organization (WHO).

**Objective** To develop a vaccine candidate against Omicron strain (B.1.1.529, BA.1) of the SARS-CoV-19.

**Methods** We applied reverse vaccinology methods for BA.1 and BA.2 as the vaccine target and a control, respectively. First, we predicted MHC I, MHC II and B cell epitopes based on their viral genome sequences. Second, after estimation of antigenicity, allergenicity and toxicity, a vaccine construct was assembled and tested for physicochemical properties and solubility. Third, AlphaFold2, RaptorX and RoseTTAfold servers were used to predict secondary structures and 3D structures of the vaccine construct. Fourth, molecular docking analysis was performed to test binding of our construct with angiotensin converting enzyme 2 (ACE2). Lastly, we compared mutation profiles on the epitopes between BA.1, BA.2, and wild type to estimate the efficacy of the vaccine.

**Results** We collected a total of 10 MHC I, 9 MHC II and 5 B cell epitopes for the final vaccine construct for Omicron strain. All epitopes were predicted to be antigenic, non-allergenic and non-toxic. The construct was estimated to have proper stability and solubility. The best modelled tertiary structures were selected for molecular docking analysis with ACE2 receptor.

**Conclusions** These results suggest the potential efficacy of our newly developed vaccine construct as a novel vaccine candidate against Omicron strain of the coronavirus.

**Keywords** SARS-CoV-2 · Omicron · *In silico* · Reverse vaccinology · Vaccine design

## Introduction

Since its emergence in the late 2019 the COVID-19 resulted in 216 million infections with estimated 4.5 million of lethal cases by August 2021 (Weekly Epidemiological Update on COVID-19–31 August 2021 2021). Currently, five variants of concerns (VOC) of SARS-CoV 2 are circulating in mass according to the World Health Organization (WHO),

including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) and the latest Omicron (B.1.1.529) variants (Tracking SARS-CoV-2 Variants 2021). Omicron variant B.1.1.529 of SARS-CoV-2 was designated as a VOC on November 26, 2021 (Update on Omicron 2022). Currently the information on its transmissibility, severity and vaccine effectiveness is yet to be confirmed. Currently two sublineages of B.1.1.529 exist including BA.1 and BA.2 (Statement on Omicron Sublineage BA.2 2022; Update on Omicron 2022).

Multiple vaccines were developed against the SARS-CoV-2, including Pfizer-BioNTech, Moderna, J&J Janssen, etc. (Self et al. 2021), and mainly used to prevent infections in various countries. However, a new strain, Omicron, was firstly identified on November 11, 2021 in Botswana and on November 14, 2021 in South Africa, and it can avoid neutralizing antibodies in vaccinated people by existing

✉ Heeбал Kim  
heeбал@snu.ac.kr

<sup>1</sup> Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea

<sup>2</sup> Department of Agricultural Biotechnology, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

<sup>3</sup> eGnome, Seoul, Republic of Korea

COVID vaccines (Cao et al. 2021). This situation is already predicted in a recent study among the US veterans shows that the SARS-CoV-2 vaccine effectiveness decreases (Cohn et al. 2021) and requires vaccine developments for the new strain. Standard process of vaccine development may take decades before a successful candidate is developed. To boost the development rate of vaccines, reverse vaccinology (RV) was developed. RV is a computational approach, which allows to discover the most potent vaccine candidates using bioinformatics methods, thus greatly amplifying the rate of vaccine development (Vaccine Research & Development 2022). Vaxign is a RV program, released in 2010, which was used to predict vaccine candidates for different pathogens (He et al. 2010). Vaxign2 and Vaxign-ML were developed to further promote the discovery of novel vaccines (Ong et al. 2021). As an example, reverse vaccinology was successfully used for the development of 4CMenB vaccine against meningococcal group B bacteria (Giuliani et al. 2022).

In this study, we utilized RV and other *in silico* methods to predict and assess a vaccine candidate for BA.1 and BA.2 subvariants of SARS-CoV-2 Omicron (B.1.1.529). For this purpose, we screened surface (S) glycoprotein and identified the most suitable epitopes. We then tested the selected epitopes for antigenicity and allergenicity to ensure vaccine safety. In the final step we performed molecular docking of the vaccine candidate with human ACE2 (Donoghue et al. 2000). As the result of this study, we developed a novel vaccine candidate against Omicron strain of the SARS-CoV-2.

## Materials and methods

### Sequence retrieval, preparation and screening of antigenic factor

Sequence of the spike(S) protein was retrieved from the reference sequence of SARS-CoV-2 in NCBI database (accession number: NC\_045512) in FASTA format and mutations corresponding to SARS-CoV-2 Omicron subvariants BA.1 and BA.2 of B.1.1.529 were introduced to it (SARS-CoV-2 Variants-Stanford Coronavirus Antiviral & Resistance Database (CoVDB) 2022; Wu et al. 2020).

### MHC I and MHC II epitopes prediction

Vaxign 2 (<http://www.violinet.org/vaxign2>) was used to predict the adhesin probability and protective antigenicity of the selected candidate (Ong et al. 2021). The sequence was then submitted to Vaxitop server (<http://www.violinet.org/vaxign2/vaxitop>) to predict MHC I and MHC II epitopes (Ong et al. 2021). We also utilized the Immune Epitope

Database (IEDB) to identify MHC I and MHC II epitopes to assess whether it would be able to produce results similar to Vaxitop (Bui et al. 2005, p. 2, 2005; Karosiene et al. 2012; Nielsen et al. 2003, 2007; Peters and Sette 2005; Sturniolo et al. 1999; Zhang et al. 2009).

### B cell epitopes prediction

ABCpred server (<https://webs.iitd.edu.in/raghava/abcpred/>) was used to predict B cell epitopes (Saha and Raghava 2006). Default threshold and 16-mer length options were selected for the epitope prediction.

### Prediction of allergenicity, antigenicity and toxicity

First, selected epitopes were subjected to allergenicity test using AlgPred 2.0 server (<https://webs.iitd.edu.in/raghava/algpred2/>). The threshold was set to 0.4 (Sharma et al. 2021). Second, antigenicity was predicted through VaxiJen server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) with 0.4 threshold (Doytchinova and Flower 2007). These methods were performed in order to select epitopes, which would most likely trigger a strong immune response, while reducing to minimum the probability of allergenic (hypersensitivity) reactions. Third, the toxicity of MHC I, MHC II and B cell epitopes was predicted using ToxinPred server (<https://webs.iitd.edu.in/raghava/toxinpred/algop.php>) (Gupta et al. 2013).

### Vaccine construction

The final vaccine construct was assembled by joining the adjuvant protein (50 S ribosomal protein L7/L12 (UniProt accession: P9WHE3)) with the selected epitopes through EAAAK linker (Arai et al. 2001; The UniProt Consortium 2019). The schematic representation of the vaccine construct along with epitope alignments between WT (NC\_045512), BA.1 and BA.2 are presented on Fig. 1a; the linkers were omitted for simplicity.

### Physicochemical properties prediction and solubility

The physicochemical properties of the final construct were estimated using ProtParam web-server (<https://web.expasy.org/protparam/>) (Gasteiger et al. 2005). SOLpro (<http://scratch.proteomics.ics.uci.edu/>) was used to evaluate the solubility of the vaccine (Magnan et al. 2009).

### Protein secondary structure prediction

Protein secondary structure and solvent accessibility was predicted using RaptorX property (<http://raptorx.uchic>



[go.edu/StructurePropertyPred/predict/](http://go.edu/StructurePropertyPred/predict/)) server (Wang et al. 2016).

### Protein 3D structure prediction, refinement and structure validation

We utilized three web-servers (AlphaFold2, RaptorX and RoseTTAfold) to predict tertiary structures of our vaccine construct without using close homologs (Baek et al. 2021; Jumper et al. 2021; Mirdita et al. 2022; Peng and Xu 2011). AlphaFold is one of the newest protein folding tools created using artificial intelligence (AI) (Jumper et al. 2021; Mirdita et al. 2022). It demonstrated impeccable performance during 14th Critical Assessment of protein Structure Prediction (CASP) (Jumper et al. 2021). RaptorX server excels at predicting structures without close homologous structures in databases and was ranked the top in contact prediction in CASP12 and CASP13 (Peng and Xu 2011). RoseTTAfold is a tool, which utilizes neural network and simultaneously considers arrangement of amino acids, their interaction and possible tertiary structures they can form (Baek et al. 2021). PyMOL software was used to visualize the resulting protein (The PyMOL Molecular Graphics System, Version 2.0 2022). GalaxyRefine (<http://galaxy.seoklab.org/>) server was used to refine and relax the protein construct (Heo et al. 2013). The server consistently scores among the top algorithms in CASP. ProSA-Web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) was used to validate the resulting protein structure (Wiederstein and Sippl 2007). This server compares the query model to structures in databases obtained through X-ray and NMR methods. PROCHECK server (<https://saves.mbi.ucla.edu/>) was used to generate a Ramachandran plot to identify allowed and disallowed psi and phi angles of amino acids within the model (Laskowski et al. 1993).

### Molecular docking

ACE2 (PDB: 6VW1) was retrieved from the Research Collaboration for Structural Bioinformatics (RCSB) Protein Data and used as a docking partner of the vaccine candidate (Shang et al. 2020). Docking between ACE2 receptor and the vaccine construct was performed in ClusPro 2.0 web-server (<https://cluspro.bu.edu/>) (Kozakov et al. 2017). ClusPro is based on PIPER algorithm, where receptor is presented as a stationary object and a ligand is rotated at different angles and positions with respect to it. The final step of docking is energy minimization using CHARMM. Third, PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) service was utilized to analyze protein-protein interface of the models (Krissinel and Henrick 2007).

## Results

### Epitope screening based on stability estimation for human

First, the protein sequence of Omicron S protein was uploaded to Vaxign 2 server for prediction of antigenic factor and epitope selection. The calculated Vaxign-ML score and adhesin probability were 96.1 and 0.533 correspondingly. The Vaxign-ML recommended threshold score of 90 indicates that the sequence has very high chances to trigger immune response in human organism (Ong et al. 2021). Neutralizing adhesins is important to prevent a pathogen entering host cells. The estimated value of 0.533 indicates adequate adhesin probability in the target protein sequence. Complete lists of epitopes predicted by Vaxitop and IEDB can be found in Supplementary data 8, 9 correspondingly.

Second, B cell epitopes were predicted utilizing ABCpred server. It uses artificial neural network in its algorithm and is suitable for identifying B cell linear epitopes for vaccine candidates (Saha and Raghava 2006). A total of 132 possible B cell epitopes were predicted, which were also sorted based on the scores produced by the server (Supplementary information 10).

We selected 10 MHC I, 9 MHC II and 5 B cell epitopes for vaccine construction based on the Vaxitop output by considering stability in human cells (Table 1). Selected epitopes were also screened for being antigenic, non-toxic and non-allergenic using the above-mentioned methods (Table 2). The antigenicity was calculated by the VaxiJen

**Table 1** Physicochemical properties of the vaccine construct (BA.1)

#	Properties	Results
1	Number of amino acids	567
2	Molecular weight	60 kDa
3	Acidic amino acids	67
4	Basic amino acids	77
5	Chemical formula	C <sub>2676</sub> H <sub>4296</sub> N <sub>726</sub> O <sub>814</sub> S <sub>11</sub>
6	Estimated half-life in <i>E. coli</i>	> 10 h
7	Estimated half-life in mammalian cells	30 h
8	Estimated half-life in yeast	> 20 h
9	Instability index (II)	35.24
10	Aliphatic index	81.55
11	GRAVY	− 0.163
12	Solubility upon overexpression	0.95
13	Alpha-helix (H)	50%
14	Beta-sheet (E)	6%
15	Coil (C)	43%

**Table 2** Selected MHC class I and MHC class II epitopes and their properties

	Epitope	Antigenicity	Start	End	Allergenicity	Toxicity	
MHC I	ISGTNGTKRF	0.5526	68	79	Non-Allergen	Non-toxic	
	FPNITNLCPF	1.3964	329	338	Non-Allergen	Non-toxic	
	KFLPFQQFGR	0.441	558	567	Non-Allergen	Non-toxic	
	KIYSKHTPI	0.7455	202	210	Non-Allergen	Non-toxic	
	KQLSSKFGAI	0.568	964	973	Non-Allergen	Non-toxic	
	LIDLQELGKY	0.7076	1197	1206	Non-Allergen	Non-toxic	
	LPIGINITRF	1.3027	229	238	Non-Allergen	Non-toxic	
	RQIAPGQTGK	1.7893	408	417	Non-Allergen	Non-toxic	
	CYFPLRSYSF	1.5062	488	497	Non-Allergen	Non-toxic	
	DISGINASV	0.4155	1168	1176	Non-Allergen	Non-toxic	
	MHC II	ASYQTQTKSHRRARS	0.7544	672	686	Non-Allergen	Non-toxic
		CTQLKRALTGIAVEQ	0.7763	760	774	Non-Allergen	Non-toxic
		GDEVQRQIAPGQTGKI	0.9741	404	418	Non-Allergen	Non-toxic
		SAIGKIQDSLST	0.5434	929	941	Non-Allergen	Non-toxic
SECVLGQSKRVDFCGKGYHL		0.9179	1030	1049	Non-Allergen	Non-toxic	
CALDPLSETKCTLKSFTVEK		0.5593	291	310	Non-Allergen	Non-toxic	
TISVTTEILPVSMT		1.2621	719	732	Non-Allergen	Non-toxic	
FPLRSYSFRPTYGVG		0.7464	490	504	Non-Allergen	Non-toxic	
DCLGDIAARDLI		0.4525	839	850	Non-Allergen	Non-toxic	
B cell		GVSVITPGTNTSNQVA	0.4651	594	609	Non-Allergen	Non-toxic
	LRSYSFRPTYGVGHQP	0.4532	492	507	Non-Allergen	Non-toxic	
	HRSYLTPGDSSSGWTA	0.6017	245	260	Non-Allergen	Non-toxic	
	TRFQTLALHRSYLTP	0.5115	236	251	Non-Allergen	Non-toxic	
	YEQYIKWPWYIWLGI	0.951	1206	1221	Non-Allergen	Non-toxic	

server, where epitopes' antigenicity was predicted by alignment free approach based on auto cross covariance (ACC) (Doytchinova and Flower 2007). The values above 0.4 were deemed as antigenic (Supplementary information 1). Epitopes' toxicity was predicted by ToxinPred server using support vector machine (SVM) (Supplementary information 1). Non-allergenic epitopes were predicted via AlgPred 2 server by using machine learning approach (Random Forest) and Basic Local Alignment Search Tools (BLAST) and motif-emerging and with classes-identification (MERC) algorithms (Sharma et al. 2021). Selected threshold of 0.4 ensured that epitopes with values below it was assumed as non-allergenic (Supplementary information 1). The majority of selected MHC I and MHC II epitopes from Vaxitop were also identified in IEDB output, and, therefore, were selected for the final vaccine construct (Supplementary information 2). The vaccine constructs of BA.2 subvariant for both Vaxign 2 and IEDB were created by aligning the corresponding epitopes of our BA.1 constructs with the S protein sequence.

### Physicochemical stability prediction

Physicochemical properties of the vaccine candidate were predicted by two independent programs (Table 2). ProtParam estimated molecular weight as 60 kDa and the instability

index (II) as 35.24, indicating a stable protein structure (Gasteiger et al. 2005). The grand average of hydropathicity (GRAVY) index was calculated at  $-0.163$ . GRAVY index measures average hydrophobicity and hydrophilicity of a protein based on the amino acids content. Negative values indicate that a protein is hydrophilic, while positive values indicate its hydrophobicity (Chang and Yang 2013). SOLpro server predicted the construct to be soluble with a probability of 0.95. The complete list of physicochemical properties of all constructed models can be found in Supplementary information 4.

### Protein structure prediction, refinement and validation

The models with the highest proportion of residues in most favored regions and the best Z-score were selected for further refinement. GalaxyRefine server was used to refine the selected tertiary structure (Supplementary information 5). We used ProSA-Web server and Ramachandran plot to validate the refined protein structure. Model produced by RoseTTAfold outperformed models created by the other servers. The calculated Z-score was  $-8.2$ . (Fig. 1c). The Ramachandran plot demonstrated that 93.7% of amino acid residues were located in the most favored regions, (Fig. 1d).

The complete list with all values can be found in Supplementary information 6.

## Molecular docking

ClusPro 2.0 web-server was utilized to simulate receptor-vaccine interaction (Fig. 1e, f). To maximize docking performance, we supplied information about contact residues of ACE2 (PDB: 6VW1) to the server (Kozakov et al. 2017; Shang et al. 2020). PDBePISA database was used to analyze interacting interfaces of the vaccine construct and the receptor. The results indicated that the solvation free energy ( $\Delta^iG$ ) of the interacting partners was equal to  $-11.5$  kcal/mol, indicating that our vaccine construct makes stronger interaction with the ACE2 (Supplementary information 7). In addition, we also performed molecular docking analysis with the top scoring models produced by AlphaFold2 and RaptorX (Supplementary information 6).

## Discussion

Previously vaccine candidates for COVID-19 were developed using computational approaches, however, they were based on the original SARS-CoV-2 protein sequences and thus do not guarantee to develop protective immunity against Omicron strain (Singh et al. 2020; Tahir ul Qamar et al. 2020). In this study we proposed a new vaccine candidate against Omicron strain SARS-CoV-2 S protein using reverse vaccinology approach. We prepared S protein of Omicron strain by introducing the corresponding mutations and deletions to the original SARS-CoV-19 sequence (NC\_045512) (SARS-CoV-2 Variants - Stanford Coronavirus Antiviral & Resistance Database (CoVDB) 2022; Wu et al. 2020). To our knowledge, this is the first vaccine design study developed specifically to reduce the spread of Omicron strain of the coronavirus infection.

In this work we attempted to design vaccine construct based on BA.1 and BA.2 subtypes of Omicron (B.1.1.529) strain of SARS-CoV-2. When designing vaccines, it is important to consider viral subtypes so the vaccines can have wider specificity. Since subtypes BA.1 and BA.2 have various mutations we concluded that comparing vaccine effectiveness would be complicated due to different epitopes profile of the constructs (SARS-CoV-2 Variants - Stanford Coronavirus Antiviral & Resistance Database (CoVDB) 2022). Thus, we used the epitopes of BA.1 predicted by Vaxign2 and IEDB as templates for BA.2 subtype via alignment of the S proteins. The safety of a vaccine is the first thing to consider, thus epitope used in the construct should be non-toxic and non-allergenic. When these criteria are satisfied, the epitopes should demonstrate adequate immunogenic properties to trigger immune response in

the host. We predicted MHC I and MHC II epitopes using Vaxign2, ensuring their antigenicity, non-toxicity and non-allergenicity to construct a vaccine. Vaxign 2 and Vaxign-ML demonstrated the best results in predicting antigens over software such as Vaxijen3 and Antigenic, and thus were selected for this study (Ong et al. 2021). Moreover, an adjuvant was added to the final vaccine construct. It is a component of a vaccine which can increase immune response to an antigen (Guideline on Adjuvants in Vaccines for Human Use 2005). *M. tuberculosis* 50 S ribosomal protein L7/L12 is a toll-like receptor (TLR) 4 agonist, which promote dendritic cell maturation and induce T cell-mediated-cytotoxic response (Lee et al. 2014). SARS-CoV-2 was shown to have a potency to activate TLR4, thus modulating this receptor can increase chances to trigger immune response against the construct. Based on these results we assume that the vaccine construct has high chances of inducing proper immune response.

In order to investigate the potential efficacy of Omicron vaccine against other SARS-CoV-19 strains, we compared the mutation profiles on epitopes based on viral genomic sequences of BA.1, BA.2 and WT (Fig. 1b). Out of total 24 epitopes, 15 epitopes (62.5%) showed perfectly identical peptide sequences across all three strains. The others (37.5%) had 16 and 10 mutations on 7 and 6 epitopes in WT and BA.2, respectively, when compared to the BA.1 subtype as a reference. It suggests BA.1 vaccine can have the efficacy for other strains via the larger proportions of epitopes without mutations. Focusing on the other epitopes with mutations, BA.2 can have a stronger potentiality to trigger protective immunity than WT, because of less number of mutations harboring small number of epitopes. We believe that subtype specific mutations present in the selected epitopes will be sufficient in triggering immune response against BA.1 and epitopes shared between all three instances can also potentially contribute to the protective immunity against various strains of SARS-CoV-19 if they could not accumulate sufficient substitutions on the epitopes to avoid the Omicron vaccine.

Tertiary structures of our vaccine constructs were evaluated by Ramachandran plot and Z-score estimated by ProSA-Web server and the top performing models were selected for further analysis. Overall, models demonstrated high proportion of residues in allowed regions and Z-scores comparable with the range of scores (X-ray, NMR) normally found within native proteins of the similar size (Laskowski et al. 1993; Wiederstein and Sippl 2007). In this study, RaptorX and RoseTTAfold produced better models than AlphaFold2. Both Raptor X and RoseTTAfold demonstrated well-estimated protein folding with adequate proportion of residues in most favorable and disallowed regions with reliable Z-scores. On the other hands, AlphaFold2 generated worse models, although, had the highest number of

residues in favorable regions and the lowest proportion of residues in disallowed regions showed very poor Z-score in a range 3.7–5.4 for all models. This indicates that tertiary structures were not folded properly and thus were removed from assessment. Visual inspection also confirmed that the structures had relatively high proportion of unfolded regions.

The molecular docking indicated positive interaction of the vaccine construct with the ACE2, mimicking virus interaction with the receptor. These results indicate the possibility for eliciting immune response against the viral epitopes presented by the constructs. However, despite the high theoretical chances of our models to be a vaccine candidate, our results are a theoretical estimation and require *in vitro* confirmation of the results to make sure the vaccine is safe and effective before making solid conclusions.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s13258-022-01255-8>.

**Acknowledgements** This study was supported by the Marine Biotechnology Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Ocean and Fisheries (MOF) (No. 20180430), Republic of Korea. This study was supported by eGnome, Seoul, South Korea.

**Author contributions** VL, CL, DY, SC and HK designed the study, collected and analyzed the data. HK supervised the research. VL, CL and DY performed the experimental part. VL and CL wrote a draft of manuscript. VL, CL and DY edited the final version of manuscript. All authors contributed to the research and came to approve the final version of the manuscript.

## Declarations

**Conflict of interest** The authors declares no conflict of interests.

**Ethical approval** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng Des Selection* 14(8):529–532. <https://doi.org/10.1093/protein/14.8.529>
- Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Baker D (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557):871–876. <https://doi.org/10.1126/science.abj8754>
- Bui H-H, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton K-A, Mothé BR, Chisari FV, Watkins DI, Sette A (2005) Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* 57(5):304–314. <https://doi.org/10.1007/s00251-005-0798-y>
- Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, Huang W, Li Q, Wang P, An R, Wang J, Wang Y, Niu X, Yang S, Liang H, Sun H, Li T, Yu Y, Cui Q, Xie XS (2021) Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*. d41586-021-03796-6 <https://doi.org/10.1038/d41586-021-03796-6>
- Chang KY, Yang J-R (2013) Analysis and Prediction of Highly Effective Antiviral Peptides Based on Random Forests. *PLoS ONE* 8(8):e70166. <https://doi.org/10.1371/journal.pone.0070166>
- Cohn BA, Cirillo PM, Murphy CC, Krigbaum NY, Wallace AW (2021) SARS-CoV-2 vaccine protection and deaths among US veterans during 2021. *Science*. <https://doi.org/10.1126/science.abm0620>
- Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, Donovan M, Woolf B, Robison K, Jeyaseelan R, Breitbart RE, Acton S (2000) A novel angiotensin-converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circul Res*. <https://doi.org/10.1161/01.RES.87.5.e1>
- Doytchinova IA, Flower DR (2007) VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8(1):4. <https://doi.org/10.1186/1471-2105-8-4>
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: &nbsp;Walker&nbsp;JM (ed) *The proteomics protocols handbook*; Humana Press, pp 571–607. <https://doi.org/10.1385/1-59259-890-0:571>
- Giuliani MM, Adu-Bobie J, Comanducci M, Arico B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capocchi B, Cartocci E, Ciucchi L, Marcello FD, Ferlicca F, Galli B, Luzzi E, Masignani V, Serruto D, Veggi D, Pizza M (2022) A universal vaccine for serogroup B meningococcus. *Med Sci* 6
- Guideline on adjuvants in vaccines for human use (2005)&nbsp;18
- Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* 8(9):e73957. <https://doi.org/10.1371/journal.pone.0073957>
- He Y, Xiang Z, Mobley HLT (2010)&nbsp;Vaxign:the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol*&nbsp;1–15. <https://doi.org/10.1155/2010/297505>
- Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* 41(W1):W384–W388. <https://doi.org/10.1093/nar/gkt458>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Karosiene E, Lundegaard C, Lund O, Nielsen M (2012) NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 64(3):177–186. <https://doi.org/10.1007/s00251-011-0579-8>
- Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S (2017) The ClusPro web server for protein–protein

- docking. *Nat Protoc* 12(2):255–278. <https://doi.org/10.1038/nprot.2016.169>
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26(2):283–291. <https://doi.org/10.1107/S0021889892009944>
- Lee SJ, Shin SJ, Lee MH, Lee M-G, Kang TH, Park WS, Soh BY, Park JH, Shin YK, Kim HW, Yun C-H, Jung ID, Park Y-M (2014) A Potential Protein Adjuvant Derived from Mycobacterium tuberculosis Rv0652 Enhances Dendritic Cells-Based Tumor Immunotherapy. *PLoS ONE* 9(8):e104351. <https://doi.org/10.1371/journal.pone.0104351>
- Magnan CN, Randall A, Baldi P (2009) SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* 25(17):2200–2207. <https://doi.org/10.1093/bioinformatics/btp386>
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022) ColabFold—Making protein folding accessible to all. <https://arxiv.org/abs/2201.12149>
- Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 12(5):1007–1017. <https://doi.org/10.1110/ps.0239403>
- Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8(1):238. <https://doi.org/10.1186/1471-2105-8-238>
- Ong E, Cooke MF, Huffman A, Xiang Z, Wong MU, Wang H, Seetharaman M, Valdez N, He Y (2021) Vaxign2: The second generation of the first Web-based vaccine design program using reverse vaccinology and machine learning. *Nucleic Acids Res* 49(W1):W671–W678. <https://doi.org/10.1093/nar/gkab279>
- Peng J, Xu J (2011) RaptorX: Exploiting structure information for protein alignment by statistical inference. *Proteins Struct Funct Bioinform* 79(S10):161–171. <https://doi.org/10.1002/prot.23175>
- Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6(1):132. <https://doi.org/10.1186/1471-2105-6-132>
- Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct Funct Bioinform* 65(1):40–48. <https://doi.org/10.1002/prot.21078>
- SARS-CoV-2 Variants—Stanford Coronavirus Antiviral & Resistance Database (CoVDB) (2022) Retrieved December 16, from [https://covdb.stanford.edu/page/mutation-viewer/#sec\\_delta](https://covdb.stanford.edu/page/mutation-viewer/#sec_delta)
- Self WH, Tenforde MW, Rhoads JP, Gaglani M, Ginde AA, Douin DJ, Olson SM, Talbot HK, Casey JD, Mohr NM, Zepeski A, McNeal T, Ghamande S, Gibbs KW, Files DC, Hager DN, Shehu A, Preker ME, Erickson HL, Arter OG (2021) Comparative Effectiveness of Moderna, Pfizer-BioNTech, and Janssen (Johnson & Johnson) Vaccines in Preventing COVID-19 Hospitalizations Among Adults Without Immunocompromising Conditions—United States, March–August 2021. *MMWR Morb Mortal Wkly Rep* 70(38):1337–1343. <https://doi.org/10.15585/mmwr.mm7038e1>
- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F (2020) Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581(7807):221–224. <https://doi.org/10.1038/s41586-020-2179-y>
- Sharma N, Patiyl S, Dhall A, Pande A, Arora C, Raghava GPS (2021) AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform* 22(4):bbaa294. <https://doi.org/10.1093/bib/bbaa294>
- Singh H, Jakhar R, Sehrawat N (2020) Designing spike protein (S-Protein) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics. *Heliyon* 6(11):e05528. <https://doi.org/10.1016/j.heliyon.2020.e05528>
- Statement on Omicron sublineage BA.2 (2022) WHO <https://www.who.int/news/item/22-02-2022-statement-on-omicron-sublineage-ba.2>
- Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 17(6):555–561. <https://doi.org/10.1038/9858>
- Tahir ul Qamar M, Shahid F, Aslam S, Ashfaq UA, Aslam S, Fatima I, Fareed MM, Zohaib A, Chen L-L (2020) Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against SARS-CoV-2. *Infect Dis Poverty* 9(1):132. <https://doi.org/10.1186/s40249-020-00752-w>
- The PyMOL Molecular Graphics System, Version 2.0 (Version 2.0) (2022) [Computer software]. Schrödinger, LLC
- The UniProt Consortium (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res* 47(D1):D506–D515. <https://doi.org/10.1093/nar/gky1049>
- Tracking SARS-(2021) CoV-2 variants (2021) Retrieved December 16, from <https://www.who.int/emergencies/what-we-do/tracking-SARS-CoV-2-variants>
- Update on Omicron (2022) Retrieved December 16, from <https://www.who.int/news/item/28-11-2021-update-on-omicron>
- Vaccine Research & Development (2022) Johns Hopkins Coronavirus Resource Center. Retrieved December 16, from <https://coronavirus.jhu.edu/vaccines/timeline>
- Wang S, Li W, Liu S, Xu J (2016) RaptorX-Property: A web server for protein structure property prediction. *Nucleic Acids Res* 44(W1):W430–W435. <https://doi.org/10.1093/nar/gkw306>
- Wiederstein M, Sippl MJ (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35(Web Server):W407–W410. <https://doi.org/10.1093/nar/gkm290>
- Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Weekly epidemiological update on COVID-19—31 August 2021 (2021) Retrieved December 16, from <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---31-august-2021>
- Zhang H, Lund O, Nielsen M (2009) The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: Application to MHC-peptide binding. *Bioinformatics* 25(10):1293–1299. <https://doi.org/10.1093/bioinformatics/btp137>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.