



Immunoinformatics Analysis of Citrullinated Antigen as Potential Multi-peptide Lung Cancer Vaccine Candidates for Indonesian Population

Angelika Hermawan¹ · Fedric Intan Damai² · Leon Martin² · Matthew Chrisdianto² · Nadya Marcelina Julianto¹ · Ihsan Tria Pramanda² · Marsia Gustiananda¹

Accepted: 2 October 2022 / Published online: 10 November 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Non-small-cell lung cancer (NSCLC) is the most common lung cancer which has the highest mortality rate in Indonesia. One of the trends in treating cancer is by utilizing peptide vaccines, an immunotherapeutic approach that aims to stimulate the cell-mediated adaptive immune system to recognize cancer-associated peptides. Currently, no peptide vaccines are available in the market for NSCLC treatment. Therefore, this project aims to develop a multi-epitope peptide-based vaccine for NSCLC utilizing citrullinated peptides. Citrullination is a post-translational modification that occurs in cancer cells during autophagy that functions to induce immune responses towards modified self-epitopes such as tumor cells, through activation of PAD enzymes within the APC and target cells. It was found that introducing a common citrullinated neo-antigen peptide such as vimentin and enolase to the immune system could stimulate a higher specific CD4⁺ T cell response against NSCLC. Moreover, carcinoembryonic antigen (CEA), an antigen that is highly expressed in cancer cells, is also added to increase the vaccine's specificity and to mobilize both CD4⁺ and CD8⁺ T cells. These antigens bind strongly to the MHC Class II alleles such as HLA-DRB1*07:01 and HLA-DRB*11:01, which are predominant alleles in Indonesian populations. Through in silico approach, the peptides generated from CEA, citrullinated vimentin and enolase, were analyzed for their MHC binding strength, immunogenicity, ability to induce IFN γ response, and population coverage. It is expected that the immunodominant antigens presentation is able to induce a potent immune response in NSCLC patients in Indonesia, resulting in tumor eradication.

Keywords Immunoinformatics · Non-small cell lung cancer · Multipeptide vaccine · Citrullinated antigen

Introduction

Lung cancer is a malignant cancer that originates in the lung. It is the third highest case of cancer in Indonesia with roughly 8.8% of cases in 2020 (Global Cancer Observatory, 2020). There are two types of lung cancer: non-small cell lung cancer and small lung cancer; their difference is mainly in size and morphology of the cancer cell. Non-small-cell carcinoma

cells (NSCLC) are the most common lung cancer which has the highest mortality rate in Indonesia which took roughly 12.6% of Indonesia's mortality in 2018 (Suraya et al. 2021). The major cause of lung cancer is smoking which results in 80% deaths in lung cancer cases. Smoke contains carcinogens, co-carcinogens, and tumor promoters that can lead to lung cancer. Aside from smoking, high alcohol consumption that exceeds 30 gr/day can also increase the risk of lung cancer. In addition, high exposure to air pollution may lead to oxidative stress, inflammation on respiratory tract, and induction of procoagulatory. On the other hand, looking at the internal factor, lung cancer can be inherited in people with a family history of lung-related diseases; it is associated with the derived expression and function of regulating enzymes in tobacco carcinogens metabolism and DNA repair (Sholih et al. 2019).

To date, lung cancer is still remarkably challenging to overcome. Early stage (I and II) or locally advanced (stage

✉ Marsia Gustiananda
marsia.gustiananda@i3l.ac.id

Angelika Hermawan
hermawanangelika@gmail.com

¹ Biomedicine Department, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

² Biotechnology Department, Indonesia International Institute for Life Sciences, Jakarta, Indonesia

III) are commonly treated with complete resection. Still, recurrence is a major threat. Tumor resections are usually accompanied with neoadjuvant chemotherapy and study has shown that this method has improved overall survival (Song et al. 2010). Although the treatment can be considered as effective, chemotherapy leads to severe side effects; diarrhea, vomiting, constipation, mucositis, and anemia are among the most common observed adverse events (Singh et al. 2019).

Recent advances in immunotherapy have changed the landscape of treatment for NSCLC (Chen et al. 2020). Immunotherapy, a combination between computer science and basic immunology, applies the computational methods to better understand immunological information (Tomar and De 2014). Immunotherapy has been utilized to provide insight into the host–pathogen interactions thus justifying their use in vaccine development (Fadaka et al. 2021). The broad scope of immunotherapy include tumor targeting and immunomodulatory monoclonal antibody, immune checkpoint inhibitors, oncolytic virotherapy, tumor targeting monoclonal antibody, cell-based therapy, peptide vaccine, and many others (Galluzzi et al. 2014).

It is known that tumor lesions from NSCLC often express the same tumor associated antigens (TAA) and minimal peptide sequences in TAA are able to induce immunological response (Palata et al. 2020). The findings had drawn scientists to target these antigens by using a peptide vaccine. The peptide vaccine will be delivered by MHC-I and MHC-II molecules of the professional antigen presenting cell (APC) thus inducing the CD8+ and CD4+ T cell response.

Several NSCLC TAAs have been identified, such as MAGE-A3, PRAME1, MUC1, CEA, and many others. Carcinoembryonic antigen (CEA) is a glycoprotein that is overexpressed in a wide range of carcinoma including NSCLC (Criscitiello 2012). CEA is categorized as adhesion molecules, its overexpression in several cancers promotes adhesion and metastasis. Based on the fact that CEA expression in normal tissues is low, but significantly higher in cancer tissue, the antigen has been widely used as a predictive biomarker for NSCLC (Kataoka et al. 2018). Aside from that, CEA may also serve as a target for active-anticancer specific immunotherapy. Clinical trials of CEA-based vaccines have shown a strong efficacy while also having a good safety profile (Criscitiello 2012).

Although several clinical studies have shown activation of CD8+ and CD4+ T cells by peptide vaccine and thereby improved patient overall outcome, the same result cannot be confirmed in a larger patient cohort, suspected due to restriction to the peptide used for immunization. One NSCLC patient might express the particular antigen, while others do not. That is why peptide vaccines are considered to be more suitable as personalized vaccines instead for the general population. Therefore, the trends shifted

towards the use of multi-peptide vaccines that bring out a broad polyclonal tumor specific immune response which is able to target multiple antigenic epitopes at the same time (Palata et al. 2020).

Aside from the common tumor-associated antigens, the immune system has the ability to recognize modified self-antigens such as citrullination. Citrullination is a post-translational modification that occurs upon cell autophagy, mediated by peptidyl arginine deiminases (PAD). The process involves changes of the positively charged aldimine group (=NH) of arginine to the neutrally charged ketone (=O) group of citrulline. Autophagy is known to be induced by stress (e.g., hypoxia and nutrient starvation) and upregulated to promote tumor survival. The most common citrullinated proteins are vimentin and α -enolase. Vimentin is a cytoskeletal protein expressed by all mesenchymal cells. The protein is highly upregulated during epithelial to mesenchymal transition of metastasizing tumor cells. Moreover, the protein is also shown to be citrullinated in some human cancer cell lines including NSCLC (Brentville et al. 2015). Meanwhile, α -enolase is a glycolytic enzyme that plays an important role in glycolysis. Many tumors prefer to generate their energy via glycolysis compared to entering oxidative phosphorylation, the process is widely known as the Warburg effect. In conjunction, there is a significant increase in glycolytic-related enzyme expression, including α -enolase. Surprisingly, this particular protein is also known to be citrullinated in tumor cells (Cook et al. 2017). A study also showed that a combination of citrullinated vimentin and enolase induces a potent CD4+ anti-tumor response (Brentville et al. 2020). Activation of CD4+ anti-tumor response has beneficial effects because it is able to break self-tolerance. Moreover, it can overcome the immunosuppressive microenvironment because one of the subsets, Th1, is a pro-inflammatory CD4+, it secretes IFN- γ and TNF- α , provides chemokine gradients to promote infiltration of immune cells such as macrophages, CD8+, and antigen-presenting cells (APC) to the tumor site. In addition, CD4+ can also differentiate into potent killer T cells (Cook et al. 2017). This summarizes the reason why the study chose to include both citrullinated vimentin and α -enolase in the multi-peptide vaccine.

In this project, the protein sequences of peptides generated from carcinoembryonic antigen (CEA) from humans, citrullinated vimentin, and enolase were collected. Through in-silico approaches, the proteins were further analyzed for their MHC binding strength, immunogenicity, ability to induce IFN γ response, and population coverage in Indonesia. Epitopes with the highest score for these parameters were selected for the multi-epitope vaccine. EAAAK, AAY, and GPGPG linkers were used to join the different epitopes (Dong et al. 2020). To increase the immunogenicity, adjuvant such as β -defensin was added into the vaccine construct (Tahir ul Qamar et al. 2020). Moreover, the TAT sequence

was also added as an adjuvant to allow the intracellular delivery of the vaccine (Dong et al. 2020).

Materials and Methods

Protein Sequence Retrieval

The normal protein sequence for alpha-enolase (ID: P06733), vimentin (ID: P08670), and CEA fragment (ID: Q13982) were obtained from the UniProt database (<https://www.uniprot.org/>). The arginine (R) was replaced with “X” to represent the citrullinated amino acid to accommodate various in-silico testing servers, as various servers do not accept the three-letter code *cit* as the code for citrullinated peptide.

Identification of Indonesian Allele Frequencies

To identify the frequencies of HLA alleles of the Indonesian population, The Allele Frequency Net Database (<http://allelefrequencies.net/hla6006a.asp>) was utilized and several alleles were identified using the classical allele search with the parameter “Indonesia” for the country tab. The allele frequencies which were found in over 5% of Indonesian individuals were chosen for further processing with the peptides from the retrieved sequences.

Prediction of MHC Class I Epitopes

Prediction of class I epitopes was done by utilizing the NetCTLpan 1.1 server (<https://services.healthtech.dtu.dk/service.php?NetCTLpan-1.1>). The FASTA sequence for each amino acid was inputted, and the corresponding MHC class I allele (HLA-A02:01, HLA-A11:01, HLA-A24:02, HLA-A24:07, HLA-A33:03, HLA-A34:01, HLA-A34:01, HLA-A34:01, HLA-B15:02, HLA-B15:13, HLA-B15:21, HLA-B18:01, HLA-B35:05, HLA-B38:02, HLA-B44:03, HLA-B58:01) frequency for the Indonesian population was chosen. The result obtained was sorted and filtered to obtain the strong binding level epitopes indicated by the abbreviation E. The sorted results were then inputted into the IEDB immunogenicity tool (<http://tools.iedb.org/immunogenicity/>) to analyze the immunogenicity.

Prediction of MHC Class II Epitopes

Prediction of class II epitopes was done by utilizing the NetMHCIIpan-4.0 server (<https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-4.0>). The FASTA sequence for each amino acid was inputted, and the corresponding MHC class II allele (DRB1_07:01, DRB1_11:01, DRB1_12:02, DRB1_15:02) frequency for the Indonesian population was

chosen. The result obtained was sorted and filtered to obtain the strong binding level epitopes indicated by the abbreviation SB. The epitopes were inputted into the IFNepitope server (<http://crdd.osdd.net/raghava/ifnepitope/predict.php>) to predict their ability in inducing IFN γ . The methods utilized for prediction were the ‘Motif and SVM hybrid’ while the prediction model was set to ‘IFN γ versus non-IFN γ ’ and the job was submitted.

Prediction of Antibody Binding

For predicted binding to B cell antibody epitopes, the IEDB Antibody Epitope Prediction tool (<http://tools.iedb.org/bcell/>) was utilized. Each antigen sequence was inputted in the sequence tab and the Bepipred Linear Epitope 2.0 was chosen as the method.

Population Coverage Analysis of the Epitopes

The population coverage analysis was performed using the IEDB Population Coverage Analysis tool (<http://tools.iedb.org/population/>). The population area was set to “Indonesia”, all the calculation options were selected, while the epitopes from the previous analysis were inputted alongside its corresponding MHC alleles.

Multi-epitope Vaccine Construct Design

Various vaccine constructs were designed and assessed for their antigenicity and population coverage to obtain the best construct. Four types of multi-epitope vaccine construct combined with adjuvant and linkers were done and summarized as Table 1. The rank of each epitope was based on its antigenicity.

Three types of linkers were used to increase protein structure stability: (1) EAAAK was used to link between the adjuvant with the MHC class I epitopes, (2) AAY was used to link between MHC class I epitopes, and (3) GPGPG was used to link between the MHC class II epitopes. Two adjuvants were also added to the construct to increase the vaccine’s immunogenicity: β -defensin placed at the C-terminal of the construct and TAT peptide flanking the sequence at the N-terminal.

Reinspection of the Predicted Epitopes from the Vaccine Construct

The sequence of the vaccine construct was reinspected for predicted MHC class I and II peptide binding by utilizing the NetCTLpan 1.1 server and NetMHCIIpan-4.0 server accordingly. The steps in the “Prediction of MHC class I epitopes” and “Prediction of MHC class II epitopes” sections were repeated with the vaccine construct as the inputted FASTA

Table 1 Vaccine constructs and their epitopes

	Enolase epitope	Vimentin epitope	CEA epitope
Construct #1 (total: 6 epitopes)	Top 1 Epitope from each MHC Class I & II	Top 1 Epitope from each MHC Class I & II	Top 1 Epitope from each MHC Class I & II
Construct #2 (total: 12 epitopes)	Top 2 Epitope from each MHC Class I & II	Top 2 Epitope from each MHC Class I & II	Top 2 Epitope from each MHC Class I & II
Construct #3 (total: 18 epitopes)	Top 3 Epitope from each MHC Class I & II	Top 3 Epitope from each MHC Class I & II	Top 3 Epitope from each MHC Class I & II
Construct #4 (total: 7 epitopes)	Top 1 Epitope from each MHC Class I & II and 1 Epitope Capable of Binding to HLA-A _{24:02} & HLA-A _{24:07}	Top 1 Epitope from each MHC Class I & II	Top 1 Epitope from each MHC Class I & II

sequence. The result obtained was sorted and filtered to retrieve the strong binding level epitopes.

Blastp Analysis of the Vaccine Construct

The protein comparison analysis was done using the Blastp (Protein BLAST) from NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) in order to avoid autoimmunity. The overall vaccine construct sequence and the 9-mer of all epitopes used in the construct are inputted into the search box. Some settings in the “algorithm parameters” section are adjusted, such as the “Expect threshold” being 30,000, the “Word size” being 2, the “Matrix” changed to PAM30, the “Gap Costs” being Existence: 9 Existence: 1, and the “Compositional adjustments” set to no adjustments. The BLAST button was clicked to begin the analysis.

Allergenicity and Antigenicity Analysis of the Vaccine Construct

Allergenicity of the vaccine construct is analyzed with AllerTOP v2.0 server (<https://www.ddg-pharmfac.net/AllerTOP/>) and further verified with AllergenFP v.1.0 server (<https://ddg-pharmfac.net/AllergenFP/>). While the antigenicity prediction is predicted based on its sequence with the ANTI-GENpro tool provided in the SCRATCH Protein Predictor server (<http://scratch.proteomics.ics.uci.edu/>). The vaccine construct sequence was inputted and the prediction feature for both protein allergenicity and antigenicity was checked.

Physicochemical Properties Evaluation of the Vaccine Construct

The physicochemical properties of the vaccine construct were evaluated using the ProtParam tool (<https://web.expasy.org/protparam/>). The FASTA sequence of the vaccine construct was inputted. Various physicochemical properties were obtained, including the number of amino acids, molecular weight, theoretical pI (isoelectric point),

formula, extinction coefficients, estimated half-life, instability index, aliphatic index, and grand average of hydrophobicity (GRAVY).

Population Coverage Analysis of the Vaccine Construct

The population coverage analysis was performed once again using the IEDB Population Coverage Analysis tool (<http://tools.iedb.org/population/>). The population area was set to “Indonesia”, all the calculation options were selected, while the vaccine construct sequence was inputted alongside its corresponding MHC alleles.

Tertiary Structure Prediction of the Vaccine Construct

The tertiary structure of the vaccine construct was predicted utilizing the RaptorX web server (<http://raptorx.uchicago.edu/>). The sequence was inserted and the server generates the best prediction.

Refinement and Validation of the Vaccine Construct

The predicted tertiary structure of the vaccine construct was further refined through the Galaxy Refine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). The PDB file of the vaccine construct was inputted. The result obtained was filtered to obtain the most ideal vaccine construct according to the MolProbity and Rama favored score. Following that, the refined structure was assessed by using the Ramachandran plot analysis through the SwissProt Expasy server (<https://swissmodel.expasy.org/assess>). The PDB file of the refined vaccine construct was inserted. The Ramachandran plot and MolProbity score were obtained.

Molecular Docking of the Vaccine Construct with Human TLR (Toll-Like Receptor)

The 3D structure of the TLR4/MD2 (ID: 3FXI) was obtained from the PDB database (www.rcsb.org). The protein–protein docking between the validated vaccine construct and TLR4/MD2 structure was performed using ClusPro 2.0 online server (<https://cluspro.org/login.php>). The whole human TLR4/ MD2 was used as the receptor molecule while the validated vaccine construct acted as the ligand. Cluspro 2.0 docking programs sample the conformation of the receptor/ligand and rank these conformations using a scoring function. The 3D structure of the docking result with the highest score was then visualized using PyMOL software.

Codon Optimization and In Silico Cloning

Codon adaptation of the designed vaccine was carried out for cloning in order to examine the codon usage by the prokaryotic organism (*Escherichia coli*, *E. coli*). To optimize the codon, the Java Codon Adaptation tool (<http://www.jcat.de/>) was utilized. The *E. coli* K12 strain was chosen to improve the vaccine's protein expression efficiency. The ρ -independent transcription termination, prokaryotic ribosome binding site, and cleavage site of restriction enzymes were avoided during optimization. XhoI and BamHI restriction endonuclease sites were added to the vaccine's N and C terminals, respectively. The in silico cloning insertion was performed on the pET28a(+) vector between the XhoI and BamHI by SnapGene.

Results

HLA Allele Frequencies in Indonesia

The allele frequencies commonly found in Indonesia are obtained from the Allele Frequency Net Database. The obtained 20 hits of alleles for both MHC class I and II that can be found in Indonesia, which is listed in Table 2. The top 5 allele frequencies found for both MHC class I and II are HLA-A24:07 (22.2%), HLA-A11:01 (16.4%), HLA-A33:03 (16.2%), HLA-A24:02 (13.9%) and HLA-DRB107:01 (13.6%).

Predicted T Cell Epitopes

Both cytotoxic and helper T cell epitopes of CEA, vimentin, and enolase antigens that bind to HLA class I and class II are predicted using NetCTLpan 1.1 and NetMHCIIpan-4.0 correspondingly. These servers are also able to predict the epitopes restriction to both HLA class I and class II.

Table 2 The most common HLA allele in Indonesia and its frequency

HLA alleles	Population	Allele frequency
A*02:01	Indonesia Java-Western	0.0660
A*11:01	Indonesia Java-Western	0.1640
A*24:02	Indonesia Java-Western	0.1390
A*24:07	Indonesia Java-Western	0.2220
A*33:03	Indonesia Java-Western	0.1620
A*34:01	Indonesia Java pop 2	0.0830
A*34:01	Indonesia Java-Western	0.0730
A*34:01	Indonesia Sundanese and Javanese	0.0670
B*15:02	Indonesia Java Western	0.1220
B*15:13	Indonesia Java-Western	0.1150
B*15:21	Indonesia Java-Western	0.0730
B*18:01	Indonesia Java-Western	0.0640
B*35:05	Indonesia Java-Western	0.0860
B*38:02	Indonesia Java-Western	0.0540
B*44:03	Indonesia Java Western	0.0930
B*58:01	Indonesia Java Western	0.0590
DRB1*07:01	Indonesia Java Western	0.1360
DRB1*11:01	Indonesia Nusa Tenggara Island	0.0890
DRB1*12:02	Indonesia Java Western	0.3650
DRB1*15:02	Indonesia Java Western	0.2330

Furthermore, class I immunogenicity analysis was also performed for potential epitopes that bind to HLA class I using a tool from IEDB. Hence, various cytotoxic T cell epitopes with strong binding levels and high immunogenicity were discovered (Table 3).

Meanwhile, potential helper T cell epitopes that bind to HLA class II were also discovered. Epitopes that were chosen are those that have a strong binding affinity indicated by the %Rank EL score. For further analysis, the epitopes were also assigned to the IFNepitope server to predict whether the epitope was able to induce IFN- γ or not. As a result, several helper T cell epitopes with strong binding affinity and the ability to induce IFN- γ were obtained (Table 4).

Predicted Antibody Epitopes

The antibody epitope prediction is made to predict B-cell epitopes from the CEA antigen, citrullinated vimentin, and citrullinated enolase sequences (Figs. 1, 2, 3). The prediction was made using the IEDB analysis tool with Bepipred linear epitope prediction 2.0 method.

The graphs below show that the citrullinated vimentin has the best average score and has a lot of the yellow area in its graph. It also has the highest maximum score out of the three. Both CEA antigen and citrullinated enolase have similar average scores, and both have a lot of scores under the threshold indicated by the green color in their graphs.

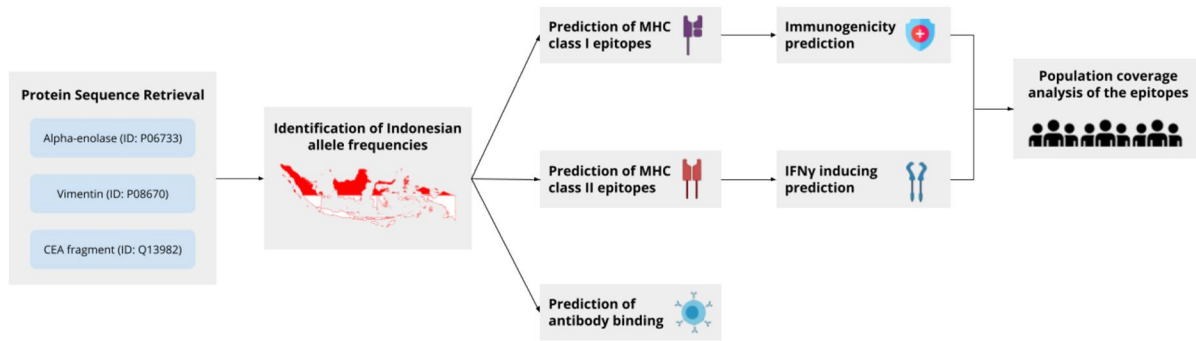
Table 3 The predicted peptide that binds to HLA class I

Antigen	Peptide	HLA	Class I immunogenicity
CEA	HLFGYSWYK	HLA-A*11:01, HLA-A*33:03, HLA-A*34:01	0.09755
	QQATPGAAY	HLA-B*15:02, HLA-B*15:21	0.12874
	EATGQFHVY	HLA-B*35:05	0.09552
Vimentin	SLQEEIAFL	HLA-A*02:01	0.38735
	LHEEEIQEL	HLA-B*38:02	0.31677
	LADAINTEF	HLA-B*15:13, HLA-B*35:05, HLA-B*58:01	0.26134
	XEMEENFAV	HLA-B*18:01, HLA-B*38:02, HLA-B*44:03	0.25682
	ELNDXFANY	HLA-A*34:01, HLA-B*15:02, HLA-B*15:21	0.21006
	EMAXHLXEY	HLA-A*34:01, HLA-B*15:02, HLA-B*15:21, HLA-B*35:05	0.18802
	NLAEDIMXL	HLA-A*02:01	0.14237
	ILLAELEQL	HLA-A*02:01	0.13965
	VETXDGQVI	HLA-B*44:03	0.04454
	IEIATYXKL	HLA-B*18:01, HLA-B*38:02, HLA-B*44:03	0.03457
	HLXEYQDLL	HLA-A*02:01	0.01715
Enolase	NYIDKVXFL	HLA-A*24:02	0.00646
	TEDTFIADL	HLA-B*38:02	0.33152
	LPVPAFNVI	HLA-B*35:05	0.1692
	ILPVPAFNV	HLA-A*02:01	0.15899
	ILPVGAAANF	HLA-A*24:02, HLA-A*24:07	0.13701
	FIADLVVGL	HLA-A*02:01, HLA-A*34:01	0.11772
	EVILPVPAF	HLA-A*34:01, HLA-B*15:02, HLA-B*35:05	0.0736
	EAMXIGAEV	HLA-A*34:01	0.2481
	DDWGAWQKF	HLA-B*18:01	0.12885
	ASEFFXSGK	HLA-A*11:01	0.1932

Table 4 The predicted peptide that binds to HLA class II

Antigen	Peptide (15-mer)	Core peptide (9-mer)	HLA	%Rank EL	IFNepitope
CEA	ESTPFNVAEGKEVLL	FNVAEGKEV	DRB1_0701	0.88	Positive
	STPFNVAEGKEVLLL	FNVAEGKEV	DRB1_0701	1.31	Positive
	FYTLQVIKSDLVNEE	LQVIKSDLV	DRB1_1502	1.46	Positive
Vimentin	SLGSALXPSTSXSPLY	LXPSTSXSPL	DRB1_0701	1.42	Positive
	LGSALXPSTSXSPLYA	LXPSTSXSPL	DRB1_0701	0.69	Positive
	GSALXPSTSXSPLYAS	LXPSTSXSPL	DRB1_0701	0.58	Positive
	MEENFAVEAANYQDT	FAVEAANYQ	DRB1_1101	1.95	Positive
Enolase	KAVEHINKTIAPALV	INKTIAPAL	DRB1_0701	1	Positive
	AGNSEVILPVPAFNV	ILPVPAFNV	DRB1_0701	1.63	Positive
	GNSEVILPVPAFNVI	ILPVPAFNV	DRB1_0701	0.57	Positive
	NSEVILPVPAFNVIN	ILPVPAFNV	DRB1_0701	0.32	Positive
	SEVILPVPAFNIVING	ILPVPAFNV	DRB1_0701	0.24	Positive
	EVILPVPAFNIVINGG	ILPVPAFNV	DRB1_0701	0.68	Positive
	DDLTVTNPKXIAKAV	LTVTNPKXI	DRB1_0701	1.72	Positive
	VSKAVEHINKTIAPA	VEHINKTIA	DRB1_1101	0.93	Positive
	SKAVEHINKTIAPAL	VEHINKTIA	DRB1_1101	1.24	Positive
	GNSEVILPVPAFNVI	ILPVPAFNV	DRB1_1502	0.54	Positive
	NSEVILPVPAFNVIN	ILPVPAFNV	DRB1_1502	0.17	Positive
	SEVILPVPAFNIVING	ILPVPAFNV	DRB1_1502	0.09	Positive
	EVILPVPAFNIVINGG	ILPVPAFNV	DRB1_1502	0.47	Positive

Preparation (Pre-Vaccine Construction):



Vaccine Construction:

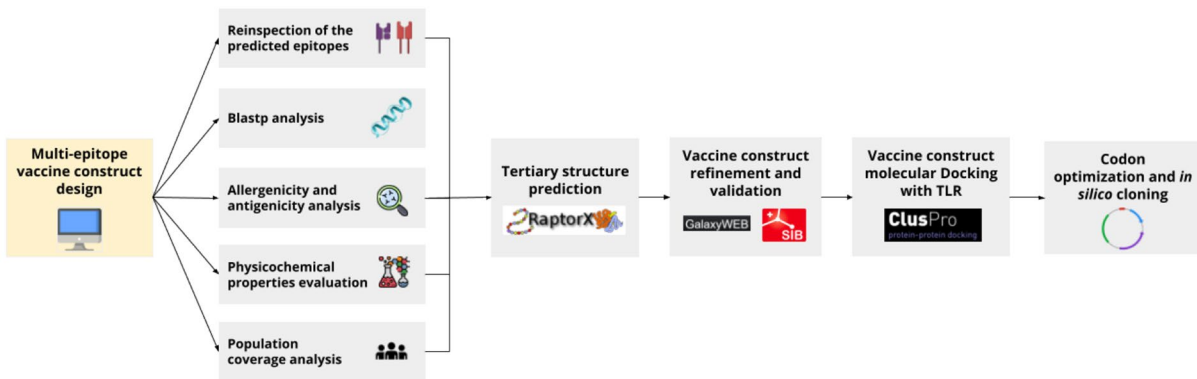
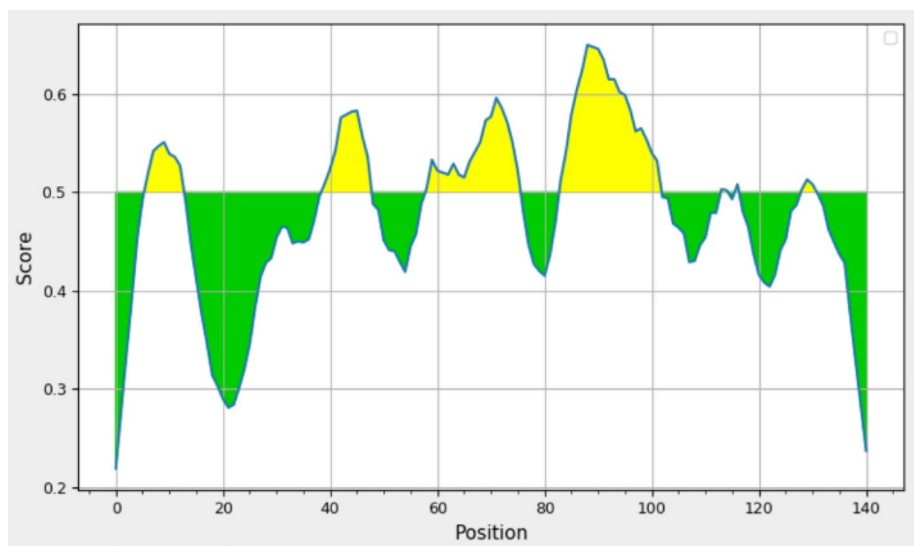


Fig. 1 Full methodology, which consisted of the preparation phase and vaccine construction phase

Fig. 2 The antibody epitope prediction result of CEA antigen using the IEDB tool and Bepipred linear epitope prediction 2.0 method. The average score is 0.478, the minimum score is 0.219, and the maximum score is 0.650

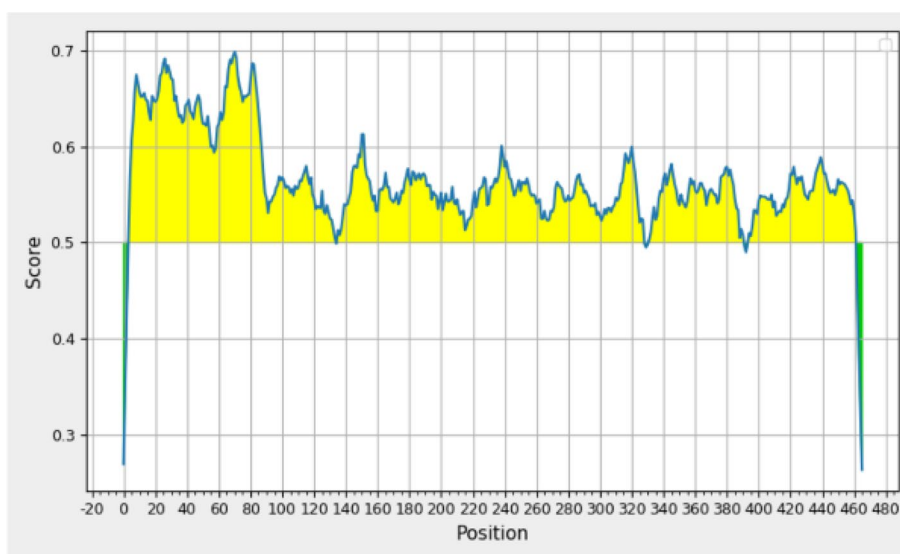


The threshold default value for residual antibody score is 0.5, indicated by the yellow area in the graphs. Those residues above the 0.5 value are probable to be part of the antibody epitope.

Population Coverage of the Epitopes

The population coverage was obtained from IEDB utilizing the epitope obtained from the binding prediction server for

Fig. 3 The antibody epitope prediction result of citrullinated vimentin using the IEDB tool and Bepipred linear epitope prediction 2.0 method. The average score is 0.565, the minimum score is 0.263, and the maximum score is 0.699



both MHC class I and II allele. The result shows that the coverage of the alleles is very high, reaching 96.7% coverage for both classes combined in enolase, 95.14% coverage in vimentin, and 91.39% coverage in CEA. However, if the alleles are observed individually, the coverage for MHC class II is quite low, only obtaining 59.56% coverage in enolase, 24.8% coverage in vimentin, and 56.39% coverage in CEA. The number is very low compared to the individual allele MHC class I coverage which reaches 91.84% in enolase, 93.54% coverage in vimentin, and 80.26% coverage in CEA. Overall, the result is quite satisfactory, where all of the epitopes could cover almost 100% coverage in class combined in all of the antigens (Table 5).

Construction of the Multi-epitope Vaccine

Out of all four constructs, construct #4 was chosen for the final vaccine construction (refer to Appendix 1). Construct #4, containing seven epitopes, was selected due to its high immunogenicity score as a whole vaccine and had the capability to bind unique Indonesian HLA (HLA-A_{24:02} and HLA-A_{24:07}). A total of 4 cytotoxic T cell epitopes and 3

helper T cell epitopes were fused with the aid of AAY and GPGPG linkers. The AAY linker was chosen to link the cytotoxic T cell epitopes, while the GPGPG linker fused the helper T cell epitopes. In addition, two adjuvant sequences (human β -defensin-3 and TAT sequence) were added to the vaccine construct to enhance immunogenicity further. The human β -defensin-3 sequence was merged to the N-terminal of the vaccine construct with the help of EAAAK linkers. On the other hand, the TAT sequence was added to the C-terminal of the vaccine construct by using GPGPG linkers (Fig. 4).

Predicted T-Cell Epitopes from the Vaccine Construct

The binding prediction between both cytotoxic and helper T cell epitopes of the vaccine construct and the HLA class I and class II are done utilizing NetCTLpan 1.1 and NetMHCIIpan-4.0 server, correspondingly. The purpose is to confirm the binding of chosen epitopes to HLA molecules and to discover other possible HLA binding peptides derived from the vaccine construct. The binding prediction of the vaccine construct epitopes to HLA class I is shown below

Table 5 Population coverage for CEA, vimentin, and enolase based peptide vaccine in Indonesia

Antigen	Class I			Class II			Class Combined		
	Coverage ^a (%)	Average hit ^b	pc90 ^c	Coverage ^a (%)	Average hit ^b	pc90 ^c	Coverage ^a (%)	Average hit ^b	pc90 ^c
CEA	80.26	1.51	0.51	56.39	0.82	0.23	91.39	2.33	1.06
Vimentin	93.54	3.49	1.36	24.80	0.66	0.13	95.14	4.15	1.67
Enolase	91.84	2	1.07	59.56	3.17	0.49	96.70	5.17	1.6

^aProjected population coverage

^bAverage number of epitope hits/ HLA combinations recognized by the population

^cminimum number of epitope hits/ HLA combinations recognized by 90% of the population

Fig. 4 The antibody epitope prediction result of citrullinated enolase using the IEDB tool and Bepipred linear epitope prediction 2.0 method. The average score is 0.457, the minimum score is 0.188, and the maximum score is 0.620

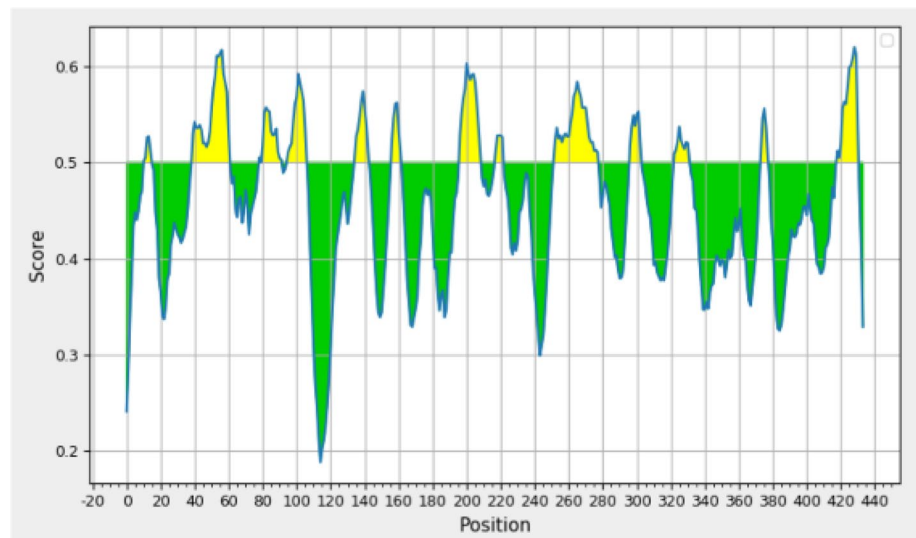


Table 6 Predicted vaccine construct epitopes that bind to HLA class I

Peptide	HLA
SLQEEIAFL	HLA-A*02:01
VTNPKXIAK	HLA-A*11:01
NTLQKYCYR	HLA-A*33:03
QQATPGAAY	HLA-B*15:02, HLA-B*15:21
LAAYTEDTF	HLA-B*15:13, HLA-B*35:05, HLA-B*58:01
KEAAKAAY	HLA-B*18:01, HLA-B*44:03
EEIAFLAAY	HLA-B*18:01, HLA-B*44:03
MEENFAVEA	HLA-B*18:01, HLA-B*44:03
TPGAAYAAAY	HLA-B*35:05
AKFVAAWTL	HLA-B*38:02
TEDTFIADL	HLA-B*38:02

in Table 6. The peptides and their corresponding HLA highlighted in bold are epitopes from our listed antigens, while the others originate from non-epitope.

Meanwhile, potential helper T cell epitopes that bind to HLA class II were also discovered. The binding prediction of the vaccine construct epitopes to HLA class II is shown in Table 7. The peptides and their corresponding HLA highlighted in bold are epitopes from our listed antigens, while the others originate from non-epitope.

Evaluation of the Vaccine Construct

The vaccine construct was further analyzed for its physical and chemical properties, including its allergenicity, antigenicity, and other physicochemical properties. Similarity with any other proteins in the human body is also checked

with the Blastp server. The result shows that the epitopes utilized in the vaccine do not show similarity with any protein found in humans, thus reducing the possibility of autoimmunity (Fig. 5).

Additionally, the allergenicity result shows that the vaccine construct is non-allergen in both AllerTOP v2.0 and AllergenFP v1.0 servers. Therefore, it signifies that the construct is non-allergenic and safe to be used as a vaccine. While for the antigenicity test, the prediction score obtained was 0.918289, where it is able to induce a 91.8% antigenic response.

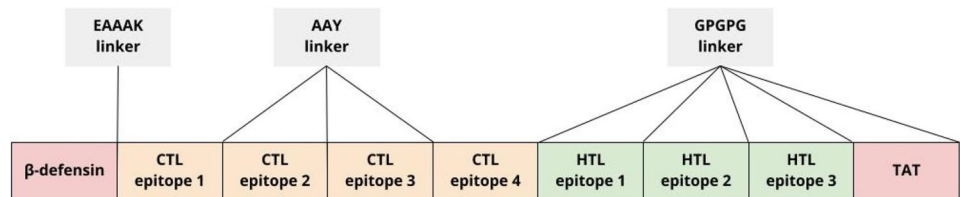
The physicochemical of the construct is also analyzed with ExPASy ProtParam. The properties are summarized in Table 8. Overall, the vaccine construct consists of 176 amino acids with the molecular formula of $C_{800}H_{1265}N_{223}O_{246}S_7$. The vaccine is relatively large, with 18,167.61 Da. In mammalian reticulocytes, the vaccine half-life is estimated to be 30 h, while in yeast (in vivo), it is predicted to be 20 h. According to the ExPASy ProtParam, the protein has an instability index of 35.62 which is considered stable. It has a GRAVY or grand average of hydropathicity value of -0.205, which means that the vaccine construct is hydrophilic.

Population Coverage of the Vaccine Construct

The IEDB population tool was utilized to obtain the population coverage of the whole vaccine construct (Table 9). According to the data presented, 92.33% of the Indonesian population could benefit from this whole vaccine construct. Meanwhile, when the allele coverage is observed individually, the coverage for MHC class I in Indonesia reached 81.04%. On the other hand, the coverage for MHC class II in Indonesia is lower, reaching 59.56%.

Table 7 Predicted vaccine construct epitopes that bind to HLA class II

Peptide (15-mer)	Core peptide (9-mer)	HLA	%Rank EL
AAKKAKFVAAWTLKA	AKFVAAWTL	DRB1_0701	1.89
AKKAKFVAAWTLKAA	AKFVAAWTL	DRB1_0701	1.1
KKAKFVAAWTLKAAA	FVAAWTLKA	DRB1_0701	1.21
KAKFVAAWTLKAAAE	FVAAWTLKA	DRB1_0701	1.15
GPGDDLTVTNPXIXIA	LTVTNPKXI	DRB1_0701	1.29
PGDDLTVTNPXIXIAK	LTVTNPKXI	DRB1_0701	0.53
GDDLTVTNPXIXIAKA	LTVTNPKXI	DRB1_0701	0.43
DDLTVTNPXIXIAKAV	LTVTNPKXI	DRB1_0701	1.72
AKKAKFVAAWTLKAA	FVAAWTLKA	DRB1_1101	1.7
KKAKFVAAWTLKAAA	FVAAWTLKA	DRB1_1101	0.95
KAKFVAAWTLKAAAE	FVAAWTLKA	DRB1_1101	0.67
KAAYQQATPGAAYAA	YQQATPGAA	DRB1_1101	1.85
PGPGFYTLQVIKSDL	FYTLQVIKS	DRB1_1101	1.96
MEENFAVEAANYQDT	FAVEAANYQ	DRB1_1101	1.95
EENFAVEAANYQDTG	FAVEAANYQ	DRB1_1101	1.34
KKAKFVAAWTLKAAA	FVAAWTLKA	DRB1_1502	1.82
KAKFVAAWTLKAAAE	FVAAWTLKA	DRB1_1502	1.35
FYTLQVIKSDLVNEE	LQVIKSDLV	DRB1_1502	1.46

Fig. 5 Schematic diagram of vaccine construct**Table 8** Physicochemical properties of the vaccine construct

Molecular formula	$C_{800}H_{1265}N_{223}O_{246}S_7$
Number of amino acids	176 aa
Molecular weight (MW)	18,167.61 Da
Piezoelectric point (pI)	7.71
Extinction coefficients (280 nm)	12,295
Half-time estimation	30 h (mammalian reticulocytes, in vitro) 20 h (yeast, in vivo) 10 h (<i>Escherichia coli</i> , in vivo)
Protein instability index (II)	35.62 (Stable)
Aliphatic index	73.41
Grand average of hydropathicity (GRAVY)	- 0.205

Table 9 Population coverage of the whole vaccine construct

Population/area	Class I			Class II			Class combined		
	Coverage ^a (%)	Average hit ^b	pc90 ^c	Coverage ^a (%)	Average hit ^b	pc90 ^c	Coverage ^a (%)	Average hit ^b	pc90 ^c
Indonesia	81.04	1.24	0.53	59.56	0.66	0.25	92.33	1.9	1.08

^aProjected population coverage^bAverage number of epitope hits/ HLA combinations recognized by the population^cMinimum number of epitope hits/ HLA combinations recognized by 90% of the population

Table 10 Vaccine construct tertiary structure Ramachandran plot and MolProbity validation results

Tertiary structure model	MolProbity score	Ramachandran favored (%)	Ramachandran outliers (%)
Model 1	3.53	82.18	6.32
Model 2	3.30	90.80	0.00
Model 3	3.40	91.95	1.72
Model 4	3.31	89.08	2.87
Model 5	3.48	84.48	4.02

Table 11 Refined vaccine construct tertiary structure Ramachandran plot and MolProbity validation results

Refined tertiary structure model	MolProbity score	Ramachandran favored (%)	Ramachandran outliers (%)
Refined model a	1.82	82.18	2.87
Refined model b	1.86	89.66	2.30
Refined model c	1.81	91.38	1.15
Refined model d	1.78	91.38	2.30
Refined model e	1.79	90.23	1.15

Structural Analysis of the Vaccine Construct

The tertiary structure prediction (RaptorX server) produces five tertiary structure models of the vaccine construct, where each was validated using Ramachandran plot and MolProbity (SwissProt Exspasy server). The tertiary structure validation results are shown below (Table 10). Model 2 is considered as the best model based on the lowest MolProbity score, relatively high Ramachandran favored score, and no Ramachandran outliers present in the structure.

The model 2 refinement (Galaxy Refine server) further produces five more models, which each are also validated using Ramachandran plot and MolProbity (SwissProt Exspasy server). The refined model validation results are shown below (Table 11). Refined model c is considered the best model as it is the only model with the highest Ramachandran favored score and lowest Ramachandran outliers, even though the MolProbity score is not particularly the highest from the rest.

Molecular Docking of the Vaccine Construct with TLR4

The final vaccine construct and human TLR4/MD2 were subjected to protein–protein docking using the Cluspro 2.0 server. The Cluspro 2.0 server ranked the model based on the cluster size, and the first model with a cluster size of 156 and lowest energy of -1023.7 was chosen for visualization using

the Pymol server (Fig. 6). As can be seen from the figure, Arg17, Val155, and Pro175 amino acids residue from the constructed vaccine interact with Ser141, Lys91, and Ser98 in human MD2 through hydrogen bonds.

In Silico Cloning of the Vaccine Construct

In silico cloning was performed to ensure expression of the constructed vaccine in *E. coli*. Initially, the codon was optimized according to the *E. coli* (Strain K12) by the JCAT server. The optimized vaccine construct contained 1515 nucleotides, an ideal range of GC content 56.38% (30–70%) and CAI value 0.93 (0.8–1.0), showing the high possibility of positive protein expression and reliability. After that, XhoI and BamHI restriction sites were inserted to both ends of the constructed vaccines optimized nucleotide sequence to aid the cloning/purification process. Finally, a refined constructed vaccine sequence was cloned between XhoI and BamHI restriction sites at the multiple cloning site of the pET28a (+) vector (Fig. 7). The total length of the clone was 5860 bp.

Discussion

Human leukocyte antigen (HLA), which is analogous to major histocompatibility complex (MHC) in other organisms, is a part of the immune system which targets specific parts of the pathogen (epitopes) and presents the antigen to T cells and B cells. HLA is encoded by genes located in chromosome 6 and is considered one of the most polymorphic regions in the human genome. The allele variations of HLA are characterized at the DNA sequence level through genomic evolution, which results in different allele frequencies in populations (Buhler and Sanchez-Mazas 2011). As Indonesia is diverse in ethnics, race, and cultural groups, more allele frequencies can be found in Indonesian populations (Pradana et al. 2019). However, the concern in high allele diversity is that not all allele frequencies will recognize the specific epitope presented in peptide-based vaccines (Nilsson et al. 2021). Since several alleles are more common in the population compared to others, it is required to list those alleles (Table 2) for potential vaccine peptides. The more peptides used from the common alleles listed, the vaccine population coverage will be better.

HLA presents the epitopes to the adaptive immune system, which consists of T cells and B cells. T cells play a critical role against foreign substances that enter the human body. It induces a receptor that will recognize various antigens while also maintaining memory and self-tolerance. Some forms of T cells present in the human body are naive

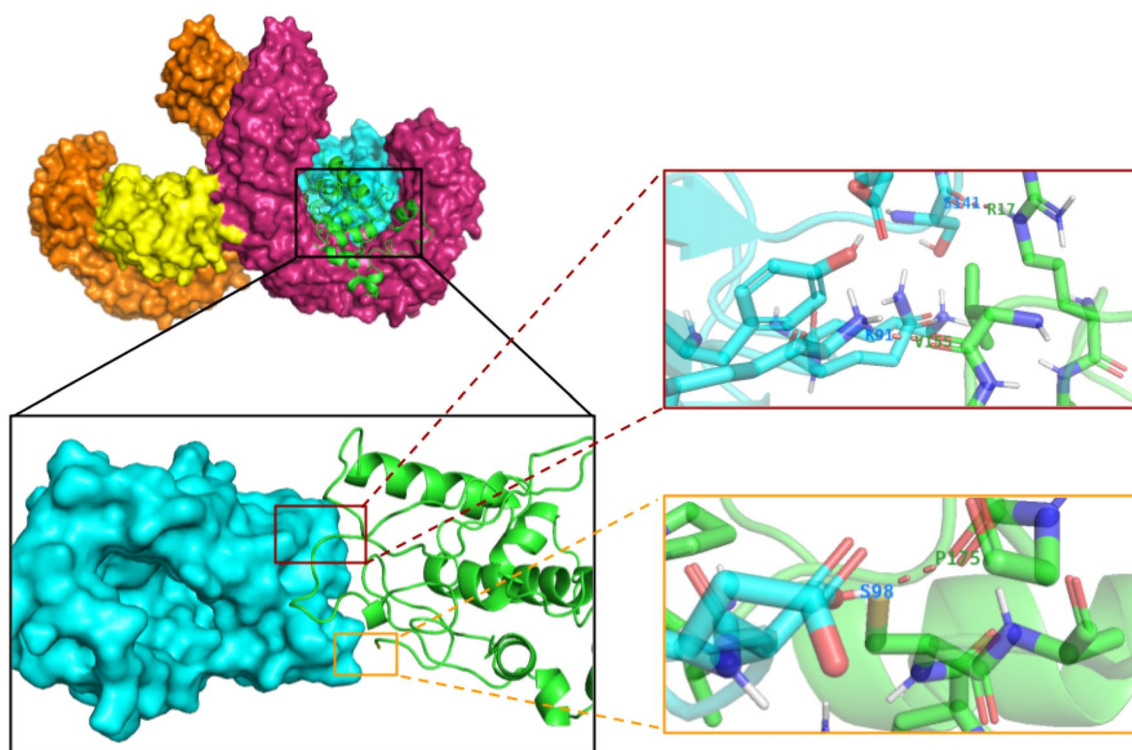


Fig. 6 TLR4/MD2-vaccine docking output. Molecular interaction between vaccine (green schematic ribbon) and MD2 (cyan wheat surface) by ClusPro 2.0. Zoomed regions show interactions of this

TLR4/MD2-vaccine complex. Sidechains are sticks. Hydrogen bonds are orange dotted lines. Amino acids are abbreviated with capitalized letters. The residue position is marked with the number

and memory T cells. Naive T cells are new T cells exposed to pathogenic antigens. In contrast, memory T cells are activated naive T cells with prolonged memory and specific to its antigen (Kumar et al. 2018). Moreover, T cells also have helper cells that secrete interleukins (IL), interferons (IFN), tumor necrosis factor (TNF) that helps induce inflammation, kill pathogens, and provide assistance to other immunogenic cells. Therefore, the T cell epitopes are essential for vaccines in inducing high-affinity antibodies and immune memory (Siegrist 2008).

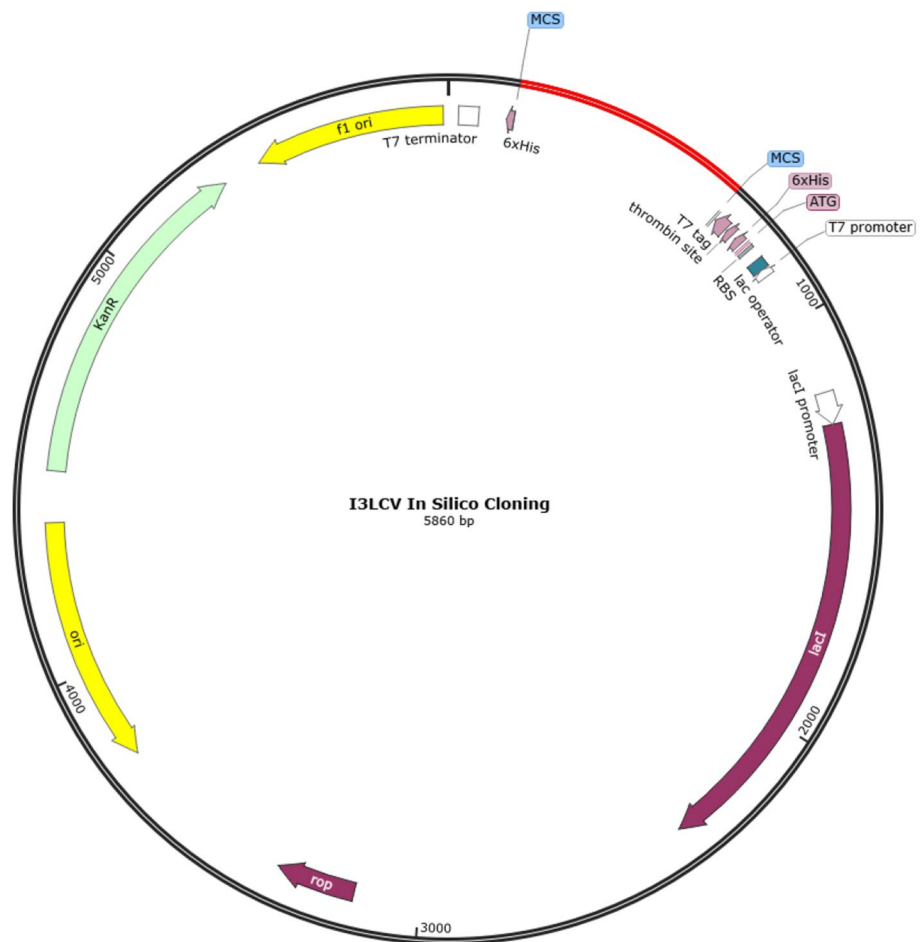
The predicted MHC class I epitopes to the designated antigens (Table 3) is obtained from NetCTLpan 1.1 server and the IEDB immunogenicity tool. The NetCTLpan 1.1 processes protein sequence input in 8-11mer peptides through an algorithm that calculates the peptides binding with MHC class I, the proteasomal C terminal cleavage, and TAP (transporter associated with antigen processing) transport. Those parameters are considered the differential points in MHC class I processing. Moreover, the threshold default values for the parameters (0.225 for cleavage and 0.025 for TAP) are considered the best weight based on the observation using the AUC fraction 0.1 as the benchmark value. The epitope identification threshold value is set to 1; therefore, the overall peptide score must meet the threshold

to be considered a good epitope candidate and marked with the “E” abbreviation (Stranzl et al. 2010).

The IEDB MHC I immunogenicity tool evaluates the ability of a peptide presented by the MHC complex to be recognized by the T cell and induce an immune response. The evaluation follows a propensity scale which is based on observations done on peptide residues that bind to T cells. Some residues are enriched with immunogenic peptides, while others are depleted. The tool validates 9-mers peptides specifically on residue 3–8, where usually those positions will bind with the T cell. The propensity scale will generate a score, where positive scores indicate possible peptide recognition by T cells (Fleri et al. 2017).

As for the predicted MHC class II epitopes (Table 4), NetMHCIIpan-4.0 and IFNepitope server were used. Similar to CTL prediction, the NetMHCIIpan-4.0 server utilizes an algorithm that measures both EL (eluted ligand) and BA (binding affinity) data, trained from various mass spectrometry (MS) experiments. Thus, the prediction does not only process the peptide-MHC class II binding but also includes the antigen presentation pathway process. In addition, the NetMHCIIpan-4.0 machine framework was upgraded from NNAlign into NNAlign_MA; therefore, it extends the server's ability to process multiple alleles (MA) data. The binding strength result is determined based on the %Rank EL

Fig. 7 In silico cloning of the vaccine in the vector, pET28a (+). Red areas represent the Lung Cancer vaccine, while the black areas represent the expression vector, pET28a (+)



score, which ranks the peptide prediction score based on the threshold for each binding strength category (strong and weak binding). The default setting for NetMHCIIpan-4.0 for binding strength threshold is set to %Rank score below the top 2% and 10% for strong and weak binding, respectively (Reynisson et al. 2020). The top 2% score will be marked with the "SB" abbreviation and considered a good epitope candidate.

Interferons (IFN) are major cytokines with antiviral, anti-tumor, and immunomodulatory features. IFN- γ , as the major product of Th1 CD4+ helper T-cell, has multiple roles in host defense, immune surveillance, and establishing innate and adaptive immunity. In addition, IFN- γ acts as a cytotoxic CD8+ T cell differentiation and proliferation signal, which is an important player in tumor cell eradication. Moreover, IFN- γ induces the "classical" polarization of macrophages towards those with a proinflammatory profile (Castro et al. 2018). These summarized why predicting the ability of peptide sequence to induce IFN- γ is crucial in vaccine construction. The prediction (Table 4) was generated using the IFNepitope server. The server adopts a special algorithm that collects data from the IEDB database for peptides that strongly bind to MHC class II. From there, the data was

divided into two, IFN- γ inducing and non-inducing peptides. The final prediction of whether the peptide sequence can induce IFN- γ production comes from the motif-based model generated by MERCI software, SVM-based model, and combination of both (Dhanda et al. 2013).

Apart from T cells, the adaptive immune system also has B cells capable of giving long-term protection against pathogenic molecules. B cells utilize its receptor named antibody or immunoglobulin (Ig) to recognize antigen via binding interactions between its receptor and a specific part of the antigen called an epitope. Therefore, B cell epitopes are used in vaccine development to ensure their prolonged persistence toward specific pathogens (Jespersen et al. 2019). In order to predict B cell epitopes, a sequence-based tool named Bepipred Linear Epitope 2.0 server was used. This server is trained on epitope data derived from crystal structures, resulting in an outstanding predictive power compared to other prediction tools. Hence, the B cell epitopes can be predicted from the protein sequence of the targeted antigen (Jespersen et al. 2017). The prediction of B cell epitopes showed that all antigens have epitopes that can bind to the B cell receptor indicated by the yellow area.

From all of the chosen CTL and HTL epitopes, a multi-epitope-based vaccine approach was performed for this study. This method has several advantages compared to a single-epitope vaccine approach. A multi-epitope vaccine is capable of eliciting strong humoral and adaptive immune responses due to the multiple epitopes contained within the vaccine (Zhang 2017). Not to mention, this type of vaccine also contains specific adjuvants that can further enhance the vaccine's immunogenicity, which in this case is the human β -defensin-3 and TAT sequence. The human β -defensin-3 is able to recruit naive T cells and induce innate immune responses by the interaction with the chemokine receptor-6 (CCR-6) (Dong et al. 2020). On the other hand, the TAT sequence was added to the construct to assist the vaccine's intracellular delivery. Lastly, all of the epitopes and adjuvants contained within the vaccine were fused with the help of EAAAK, AAY, and GPGPG linkers. In addition, these linkers also have crucial roles in maintaining the stability of the protein structure (Dong et al. 2020).

To evaluate the vaccine construct, BLASTp screening, antigenicity, immunogenicity, allergenicity, physicochemical test, and molecular docking was performed. The aim of this screening is to ensure that the vaccine construct does not elicit any autoimmunity potential. Any part of the construct that has a similarity higher than 35% is considered a homologous protein with the human proteome and thus unable to be used as the vaccine construct (Yang et al. 2021). As no similarity was found between the construct and the human proteome, the vaccine construct is considered safe for usage.

Antigenicity is the ability to bind specifically with a certain group of molecules that have adaptive immunity, while immunogenicity is the ability of a substance to induce a cellular and humoral immune response. Another term, allergenicity, is the antigen's ability to induce an unusual immune response known as allergy. It could be due to overreaction or distinct response by the immune response in comparison to the normal response, resulting in tissue damage or physiological disorders (Zhang and Tao 2015). These parameters are essential to ensure the capability of the construct to induce the host's immune response, as not all immunogenic substances are capable of inducing antigen response (Ilinskaya and Dobrovolskaia 2016). The antigenicity is based on the construct's sequence, where the primary sequence from multiple representations and five machine learning algorithms are utilized for the prediction. The result obtained is finalized with the support vector machine to determine whether the sequence of interest is antigenic or not (Heo et al. 2013). The common threshold for tumor antigenicity is set as 40% by the VaxiJen v2.0 prediction server (Doytchinova and Flower 2007), whereas the construct prediction score obtained was 91.8%. Therefore, the construct exhibits

very high potential as an antigenic and immunogenic protein with no risk of inducing allergy.

Protein's physicochemical properties are essential to understand its structural organization, stabilization, and how the protein of interest in the vaccine construct would interact with its environment (Yang et al. 2021). The understanding would enable identifications of cryptic elements for protein folding, residues that stabilize the protein's 3D structure, the behavior of amino acids, protein contact network, hydrophobicity, and others (Sengupta and Kundu 2012). From the negative GRAVY value of the vaccine construct, we could determine that it is hydrophilic, which could easily interact with water molecules (Yang et al. 2021). The half-life of the construct is approximately 30 h *in vitro* and 20 h *in vivo*, which is sufficient for a prolonged host immune system induction. Several methods which can prolong protein half-life are PEGylation and genetic fusion with other proteins such as human serum albumin (HSA) (AlQahtani et al. 2019). A high aliphatic index shows that the construct has high thermal stability based on the relative volume occupied by the aliphatic side chain. While the stability index below 40 indicates that the protein is stable in its solvent environment, based on the occurrence of certain dipeptide bonds (Pyasi et al. 2021). These results show that the construct has the potential to be the NSCLC vaccine candidate.

The Cluspro 2.0 server uses PIPER, a docking program that implements the Fast Fourier Transform (FFT) correlation method. FFT calculates energy function, thus enabling the sampling of billions of conformations of two interacting proteins. In Cluspro, the docked protein model was selected based on the cluster size. The bigger the cluster size, the higher probability of interactions. Furthermore, the server provides information regarding the lowest energy. It does not directly affect the binding affinity, but low-energy regions are more likely to produce more docked structures clusters (Kozakov et al. 2017; Natasya et al. 2021).

In this study, molecular docking is performed between TLR4/MD2 and our peptide vaccine with the intention of learning how the molecules interact with each other naturally. Toll-like receptor 4 (TLR4) is a leucine-rich repeat molecule, a surface protein expressed by innate immune sentinel cells such as dendritic cells and macrophages. Normally, TLR4 recognizes bacterial lipopolysaccharides and then induces the production of inflammatory signals to recruit other immune players (Hajjar et al. 2002). However, a study by Shimazu et al. (1999) mentioned that the activation of human TLR4 requires another additional molecule which is MD2. MD2 was found to be physically attached to TLR4 molecules on the cell surface and provide responsiveness towards bacterial LPS. Meanwhile, TLR ligands have long been used as adjuvants for peptide vaccine, and reports suggest that it plays an important role in enhancing antitumor immunity. Furthermore, it was known that lung cancer

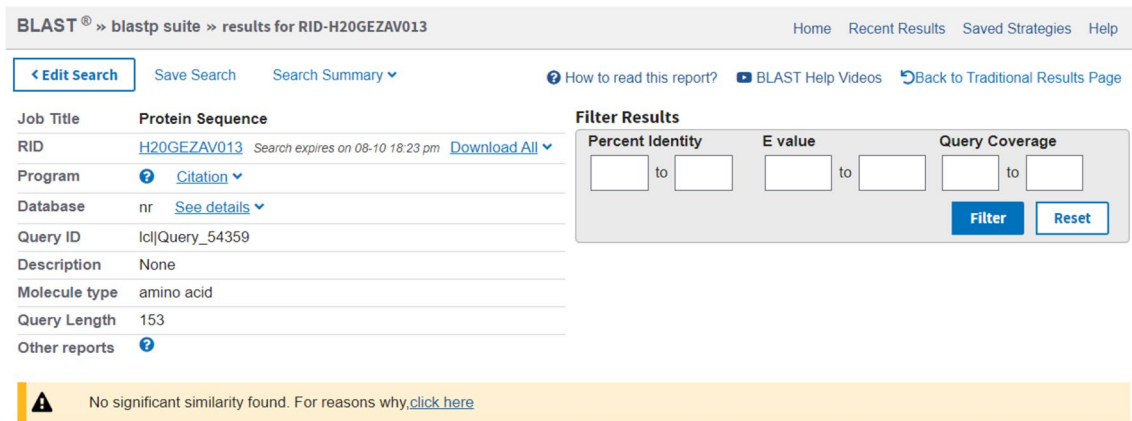


Fig. 8 Blastp result of the multi-epitope vaccine construct

patients express higher levels of TLR4 (Urban-Wojciuk et al. 2019), which further adds its role in eliminating lung cancer cells. As can be seen from Fig. 8, our constructed multi-peptide vaccine interacts with the TLR4, specifically the MD2 molecule. Therefore, the hydrogen bonds formed between the two molecules at different sites were expected to induce activation of human TLR4 and hence boost the antitumor immune response.

After the protein is translated into nucleotides, codon optimization on *Escherichia coli* is conducted to ensure and increase the protein expression (Mauro and Chappell 2014). The plasmid pET28a(+) is utilized as a vector; the optimized codon is designed with XhoI and BamHI restriction sites on the vaccine's N and C terminals, respectively. pET28a(+) contains the T7 promoter that initiates the translation, mediated by a Shine-Dalgarno (SD) sequence. The sequence is cloned with a poly-histidine purification tag (His₆) and a thrombin protease recognition site which will ease the protein purification. The plasmid also comes with the kanamycin resistance gene (KanR) that works as a selective marker. The constructed plasmid was created such that it can be realized in further experiments to turn this vaccine candidate into a vaccine.

Advances in bioinformatics have allowed the prediction of HLA binder epitopes, potential allergenicity, antigenicity and immunogenicity, physicochemical properties, 3D structure, and molecular docking. The bioinformatics approach greatly reduced the cost of the development of vaccines. In addition, it reduces the possibility of failure in the in vitro and in vivo experiments (Ranjbar et al. 2019). Therefore, our multi-peptide vaccine has great potential as a novel lung cancer vaccine. Nevertheless, further in vitro and in vivo is required to validate the efficacy of the multi-peptide vaccine against NSCLC.

Conclusion

In conclusion, the vaccine construct was developed using citrullinated antigens in order to induce a potent immune response and become an alternative treatment for NSCLC patients in Indonesia. The vaccine development is done in silico utilizing CEA, citrullinated vimentin, and citrullinated enolase. The resulting vaccine construct consists of seven T-cell epitopes obtained from the three antigens. The construct has been analyzed and evaluated for its structure, properties, and population coverage. The analysis indicates that the construct is potentially viable as a treatment against lung cancer, and it can cover most of the Indonesian population. Not to mention, the structural docking shows that the constructed vaccine is also capable of interacting with the TLR4, which was expected to induce its activation. However, further in vitro and in vivo experiments are suggested to verify the efficacy of the constructed vaccine towards NSCLC and could be expanded to world population.

Appendix

Appendix 1

The sequence of construct #1, where red indicates the sequence for β -defensin, green indicates the linkers, blue indicates the epitopes for MHC class I alleles, purple indicates the epitopes for MHC class II alleles, and yellow indicates the TAT peptide.

```
GIINTLQKYYCRVRRGRCVLSCLPKKEEQIGK-
STRGRKCCRRKKEAAAKAAYQQATPGAAYAA
YSLQEEIAFLAAYTETDFIADLPGPGFYTLQVIKS-
DLVNEEGPGPGMEENFAVEAANYQDTGPGPGDDLTVT-
NPKXIAKAVGPGPGTGALLAAGAAAGPGPG
```

Appendix 2

The sequence of construct #2, where red indicates the sequence for β -defensin, green indicates the linkers, blue indicates the epitopes for MHC class I alleles, purple indicates the epitopes for MHC class II alleles, and yellow indicates the TAT peptide.

```
GIINTLQKYYCRVRRGRCVLSCLPKKEEQIG-
KCSTRGRKCCRRKKEAAAKQQATPGAAYAA
YHLFGYSWYKAAAYSLQEEIAFLAAYLHEEEIQE-
LAAYTEDTFIADLAAYEAMXIGAEEVGGPGGFYTLQVIKS-
DLVNEEGPGPGSTPFNVAEGKEVLLGPGPGMEEN-
FAVEAANYQDTGPGPGSLGSALXPSTXSPLYGPGPGDD-
LTVTNPKXIAKAVGPGPGAGNSEVILPVPAFNVGPGPGT-
GALLAAGAAAGPGPG
```

Appendix 3

The sequence of construct #3, where red indicates the sequence for β -defensin, green indicates the linkers, blue indicates the epitopes for MHC class I alleles, purple indicates the epitopes for MHC class II alleles, and yellow indicates the TAT peptide.

```
GIINTLQKYYCRVRRGRCVLSCLPKKEEQIG-
KCSTRGRKCCRRKKEAAAKQQATPGAAYAA
YHLFGYSWYKAAAYEATGQFHVYAAAYSLQEEIAF-
LAAYLHEEEIQELAAAYNYIDKVVFLAAYTEDTFIAD-
LAAYEAMXIGAEEVAAYILPVGAANFGPGPGFYTLQVIKS-
DLVNEEGPGPGSTPFNVAEGKEVLLGPGPGESTPFNVAE-
GKEVLLGPGPGMEENFAVEAANYQDTGPGPGSLG-
SALXPSTXSPLYGPGPGLGSALXPSTXSPLYAGPGPGDDL-
VTNPKXIAKAVGPGPGAGNSEVILPVPAFNVGPGPGSKA-
VEHINKTIAPALGPGPGTGALLAAGAAAGPGPG
```

Appendix 4

The sequence of construct #4, where red indicates the sequence for β -defensin, green indicates the linkers, blue indicates the epitopes for MHC class I alleles, purple indicates the epitopes for MHC class II alleles, and yellow indicates the TAT peptide.

```
GIINTLQKYYCRVRRGRCVLSCLPKKEEQIG-
KCSTRGRKCCRRKKEAAAKQQATPGAAYAA
YHLFGYSWYKAAAYEATGQFHVYAAAYSLQEEIAF-
LAAYLHEEEIQELAAAYNYIDKVVFLAAYTEDTFIAD-
LAAYEAMXIGAEEVAAYILPVGAANFGPGPGFYTLQVIKS-
DLVNEEGPGPGSTPFNVAEGKEVLLGPGPGESTPFNVAE-
GKEVLLGPGPGMEENFAVEAANYQDTGPGPGSLG-
SALXPSTXSPLYGPGPGLGSALXPSTXSPLYAGPGPGDDL-
VTNPKXIAKAVGPGPGAGNSEVILPVPAFNVGPGPGSKA-
VEHINKTIAPALGPGPGTGALLAAGAAAGPGPG
```

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by AH, FID, LM, MC, and NMJ. The first draft of the manuscript was written by AH, FID, LM, MC, and NMJ and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other financial support were received during the preparation of this manuscript.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interest The authors have no relevant financial or non-financial interests to disclose.

References

- AlQahtani AD, O'Connor D, Domling A, Goda SK (2019) Strategies for the production of long-acting therapeutics and efficient drug delivery for cancer treatment. *Biomed Pharmacother* 113:108750. <https://doi.org/10.1016/j.biopha.2019.108750>
- Brentville VA, Metheringham RL, Gunn B, Symonds P, Daniels I, Gijon M, Cook K, Xue W, Durrant LG (2015) Citrullinated vimentin presented on MHC-II in tumor cells is a target for CD4+ T-cell-mediated antitumor immunity. *Cancer Res* 76(3):548–560. <https://doi.org/10.1158/0008-5472.can-15-1085>
- Brentville VA, Metheringham RL, Daniels I, Atabani S, Symonds P, Cook KW, Vankemmelbeke M, Choudhury R, Vaghela P, Gijon M, Meiners G, Krebber W-J, Melief CJM, Durrant LG (2020) Combination vaccine based on citrullinated vimentin and enolase peptides induces potent CD4-mediated anti-tumor responses. *J Immunother Cancer* 8(1):e000560. <https://doi.org/10.1136/jitc-2020-000560>
- Buhler S, Sanchez-Mazas A (2011) HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events. *PLoS ONE* 6(2):e14643. <https://doi.org/10.1371/journal.pone.0014643>
- Castro F, Cardoso AP, Gonçalves RM, Serre K, Oliveira MJ (2018) Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Front Immunol* 9:847. <https://doi.org/10.3389/fimmu.2018.00847>
- Chen R, Manochakian R, James L, Azzouqa A-G, Shi H, Zhang Y, Zhao Y, Zhou K, Lou Y (2020) Emerging therapeutic agents for advanced non-small cell lung cancer. *J Hematol Oncol* 13(1):1–23. <https://doi.org/10.1186/s13045-020-00881-7>
- Cook K, Daniels I, Symonds P, Pitt T, Gijon M, Xue W, Metheringham R, Durrant L, Brentville V (2017) Citrullinated α -enolase is an effective target for anti-cancer immunity. *Oncotarget* 7(2):e1390642. <https://doi.org/10.1080/2162402x.2017.1390642>
- Criscitello C (2012) Tumor-associated antigens in breast cancer. *Breast Care* 7(4):262–266. <https://doi.org/10.1159/000342164>
- Dhanda SK, Vir P, Raghava GP (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* 8(1):1–15. <https://doi.org/10.1186/1745-6150-8-30>
- Dong R, Chu Z, Yu F, Zha Y (2020) Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. *Front Immunol* 11:1784. <https://doi.org/10.3389/fimmu.2020.01784>

- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform* 8(1):1–7. <https://doi.org/10.1186/1471-2105-8-4>
- Fadaka A, Sibuyi N, Martin D, Goboza M, Klein A, Madiehe A, Meyer M (2021) Immunoinformatics design of a novel epitope-based vaccine candidate against dengue virus. *Sci Rep* 11(1):1–22. <https://doi.org/10.1038/s41598-021-99227-7>
- Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, Sette A (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol* 8:278. <https://doi.org/10.3389/fimmu.2017.00278>
- Galluzzi L, Vacchelli E, Pedro J-MB-S, Buqué A, Senovilla L, Baracco EE, Bloy N, Castoldi F, Abastado J-P, Agostinis P, Apte RN, Aranda F, Ayyoub M, Beckhove P, Blay J-Y, Bracci L, Caignard A, Castelli C, Cavallo F, Celis E (2014) Classification of current anticancer immunotherapies. *Oncotarget* 5(24):12472–12508. <https://doi.org/10.18632/oncotarget.2998>
- Global Cancer Observatory (2020) Population Fact Sheets, GLOBOCAN. <https://gco.iarc.fr/today/fact-sheets-populations>
- Hajjar AM, Ernst RK, Tsai JH, Wilson CB, Miller SI (2002) Human Toll-like receptor 4 recognizes host-specific LPS modifications. *Nat Immunol* 3(4):354–359. <https://doi.org/10.1038/ni777>
- 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>
- Iliinskaya AN, Dobrovolskaia MA (2016) Understanding the immunogenicity and antigenicity of nanomaterials: past, present and future. *Toxicol Appl Pharmacol* 299:70–77. <https://doi.org/10.1016/j.taap.2016.01.005>
- Jespersen MC, Peters B, Nielsen M, Marcatili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45(W1):W24–W29. <https://doi.org/10.1093/nar/gkx346>
- Jespersen MC, Mahajan S, Peters B, Nielsen M, Marcatili P (2019) Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes. *Front Immunol* 10:298. <https://doi.org/10.3389/fimmu.2019.00298>
- Kataoka Y, Hirano K, Narabayashi T, Hara S, Fujimoto D, Tanaka T, Ebi N, Tomii K, Yoshioka H (2018) Carcinoembryonic antigen as a predictive biomarker of response to Nivolumab in non-small cell lung cancer. *Anticancer Res* 38(1):12259. <https://doi.org/10.21873/anticancer.12259>
- Kozakov D, Hall DR, Xia B, Porter KA, Padhony 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>
- Kumar BV, Connors TJ, Farber DL (2018) Human T cell development, localization, and function throughout life. *Immunity* 48(2):202–213. <https://doi.org/10.1016/j.immuni.2018.01.007>
- Mauro VP, Chappell SA (2014) A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* 20(11):604–613. <https://doi.org/10.1016/j.molmed.2014.09.003>
- Natasya N, Tera T, Savero S, Swastika N, Parikesit AA (2021) In silico computation of coagulation factor II: a potential water treatment agent against gram negative bacteria. *Notulae Scientia Biologicae* 13(1):11021. <https://doi.org/10.15835/nsb13311021>
- Nilsson JB, Grifoni A, Tarke A, Sette A, Nielsen M (2021) PopCover-2.0. Improved selection of peptide sets with optimal HLA and pathogen diversity coverage. *Front Immunol* 12:728936. <https://doi.org/10.3389/fimmu.2021.728936>
- Palata O, Podzimekova Hradilova N, Mysiková D, Kutna B, Mrazkova H, Lischke R, Spisek R, Adkins I (2020) Detection of tumor antigens and tumor-antigen specific T cells in NSCLC patients: correlation of the quality of T cell responses with NSCLC subtype. *Immunol Lett* 219:46–53. <https://doi.org/10.1016/j.imlet.2020.01.001>
- Pradana KA, Widjaya MA, Wahjudi M (2019) Indonesians Human Leukocyte Antigen (HLA) distributions and correlations with global diseases. *Immunol Invest*. <https://doi.org/10.1080/08820139.2019.1673771>
- Piyasi S, Sharma V, Dipti K, Jonniya NA, Nayak D (2021) Immunoinformatics approach to design multi-epitope-subunit vaccine against bovine ephemeral fever disease. *Vaccines* 9(8):925. <https://doi.org/10.3390/vaccines9080925>
- Ranjbar MM, Ebrahimi MM, Shahsavandi S, Farhadi T, Mirjalili A, Tebianian M, Motedayen MH (2019) Novel applications of immunobioinformatics in vaccine and bio-product developments at research institutes. *Arch Razi Inst* 74(3):219–233. <https://doi.org/10.22092/ari.2018.122523.1224>
- Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 48(W1):W449–W454. <https://doi.org/10.1093/nar/gkaa379>
- Sengupta D, Kundu S (2012) Role of physico-chemical properties of amino acids in protein's structural organization: a network perspective. In: International conference on information processing in cells and tissues. Springer, Berlin, pp. 74–81. https://doi.org/10.1007/978-3-642-28792-3_11
- Shimazu R, Akashi S, Ogata H, Nagai Y, Fukudome K, Miyake K, Kimoto M (1999) MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J Exp Med* 189(11):1777–1782. <https://doi.org/10.1084/jem.189.11.1777>
- Sholih MG, Perwitasari DA, Hendriani R, Sukandar H, Barliana MI, Suwantika A, Diantini A (2019) Risk factors of lung cancer in Indonesia: a qualitative study. *J Adv Pharm Educ Res* 9(2):41–45
- Siegrist CA (2008) Vaccine immunology. *Vaccines* 5(1):17–36. <https://doi.org/10.1016/b978-1-4557-0090-5.00004-5>
- Singh N, Muthu V, Myllemngap B, Prasad K, Behera D (2019) Adverse effects observed in lung cancer patients undergoing first-line chemotherapy and effectiveness of supportive care drugs in a resource-limited setting. *Lung India* 36(1):32. https://doi.org/10.4103/lungindia.lungindia_321_17
- Song W-A, Zhou N-K, Wang W, Chu X-Y, Liang C-Y, Tian X-D, Guo J-T, Liu X, Liu Y, Dai W-M (2010) Survival benefit of neoadjuvant chemotherapy in non-small cell lung cancer: an updated meta-analysis of 13 randomized control trials. *Thorac Oncol* 5(4):510–516. <https://doi.org/10.1097/jto.0b013e318cd3345>
- Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 62(6):357–368. <https://doi.org/10.1007/s00251-010-0441-4>
- Suraya A, Nowak D, Sulistomo AW, Icksan AG, Berger U, Syahrudin E, Bose-O'Reilly S (2021) Excess risk of lung cancer among agriculture and construction workers in Indonesia. *Ann Glob Health* 87(1):8. <https://doi.org/10.5334/aogh.3155>
- Tahir ul Qamar M, Rehman A, Tusleem K, Ashfaq UA, Qasim M, Zhu X, Fatima I, Shahid F, Chen L-L (2020) Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: immunoinformatics and in silico approaches. *PLoS ONE* 15(12):e0244176. <https://doi.org/10.1371/journal.pone.0244176>
- Tomar N, De R (2014) Immunoinformatics: a brief review. *Methods Mol Biol*. https://doi.org/10.1007/978-1-4939-1115-8_3
- Urban-Wojciuk Z, Khan MM, Oyler BL, Fähraeus R, Marek-Trzonkowska N, Nita-Lazar A, Hupp TR, Goodlett DR (2019) The role of TLRs in anti-cancer immunity and tumor rejection. *Front Immunol* 10:2388. <https://doi.org/10.3389/fimmu.2019.02388>
- Yang Z, Bogdan P, Nazarian S (2021) An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Sci Rep* 11(1):1–21. <https://doi.org/10.1038/s41598-021-81749-9>

- Zhang L (2017) Multi-epitope vaccines: a promising strategy against tumors and viral infections. *Cell Mol Immunol* 15(2):182–184. <https://doi.org/10.1038/cmi.2017.92>
- Zhang J, Tao A (2015) Antigenicity, immunogenicity, allergenicity. *Allergy Bioinform*. https://doi.org/10.1007/978-94-017-7444-4_11

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

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