The epitopes analysis and construction of recombinant plasmid of fused ESAT-6/Tb10.4 for tuberculosis vaccine development

Sabar Pambudi[1,2*], *Anwar Shafrial* Irsyad[2], *Doddy Irawan Setyo* Utomo[1], *Astutiati* Nurhasanah[1], *Titta* Novianti[2]

[1]Research Center for Vaccine and Drugs, National Research and Innovation Agency, Building 611, LAPTIAB, KST BJ Habibie, Tangerang Selatan, Banten, Indonesia, 15314
[2]Department of Biotechnology, Universitas Esa Unggul, Jl, Arjuna Utara no.9, Jakarta, Indonesia, 11510

Abstract. *Mycobacterium tuberculosis* (Mtb) is a pathogenic bacteria responsible for tuberculosis (TB), an infectious disease that poses a significant threat in Indonesia. Despite extensive and thorough research efforts throughout the years, Bacille Calmette-Guerin (BCG) remains the sole authorized vaccination with varying levels of effectiveness. It offers immunity against tuberculosis in children but is not efficacious in treating tuberculosis in adults. Epidemiological modelling indicates that, despite advancements in pharmacological treatments for tuberculosis, the World Health Organization's efforts to contain the spread of the illness necessitate the development of a novel vaccine with the ability to prevent tuberculosis. The B-cell epitope prediction algorithms have significant medical and economic value because of their practical use in vaccine development. In this study, we employed immunoinformatic prediction tools such as Alphafold, Ellipro, VaxiJen, and IFNepitop to analyze the epitopes of fuse antigen ESAT-6/Tb10.4. We then performed molecular cloning of fuse gene ESAT-6/Tb10.4 into bacteria expression vector pET21d(+). The predicted template modelling (pTM) from Alphafold 3 of our fused protein is 0.57. Which means it might be similar to the true structure. The B cell epitope from Ellipro analysis showed 5 linear and 5 discontinuous epitopes. Our analysis using IFNepitop predicted 126 candidates induced interferon gamma-inducing epitopes out of 198 peptides. Moreover, we successfully fused the ESAT-6 gene and Tb10.4 gene into expression vector pET21d(+) and confirmed by restriction enzyme digestion.

1 Introduction

In 2018, tuberculosis (TB) caused 1.5 million people deaths, making it the leading cause of death due to a single infectious agent. The need for innovative methods to handle this disease is of utmost importance. The BCG vaccination, developed by Calmette and Guerin in the

* Corresponding author: saba002@brin.go.id; sabar.pambudi@esaunggul.ac.id

early 1900s and first administered to individuals 99 years ago, is an essential element of tuberculosis control. The BCG vaccine was created by repeatedly deriving the bovine TB bacillus, Mycobacterium bovis, resulting in genetically and phenotypically distinct progeny strains. When applied intradermally (i.d.) at birth, BCG effectively shields newborns from severe and widely distributed tuberculosis. BCG, however, has varying but limited effectiveness in preventing pulmonary tuberculosis in adults and adolescents [1].

BCG immunization during infancy does not consistently provide effective protection against pulmonary tuberculosis in adults and adolescents. BCG, being a live attenuated vaccine, relies on replication and/or persistence to provide protection. Nevertheless, the presence of nontuberculous mycobacteria (NTM) has been discovered to impede the reproduction of BCG and diminish its efficacy in producing immune responses that can react to other pathogens [2, 3].

Nonreplicating subunit vaccines differ from BCG and other live injections in that they employ synthetic immunostimulatory adjuvants and are not influenced by prior mycobacterial sensitisation. Due to the encouraging progress in the clinical development of TB vaccines, it has been suggested that subunit vaccines could be given in addition to BCG (re)vaccination to create a more potent vaccine regimen [4].

The ESX systems of Mycobacterium TB are responsible for generating highly immunogenic proteins that are essential for the bacterium's survival and rapid growth. The two prototypic proteins, ESAT-6 (EsxA from ESX-1) and TB10.4 (EsxH from ESX-3), share many characteristics in terms of their genomic structure, size, antigenic properties, and potential for use in vaccines. Nevertheless, these two compounds have different purposes in bacterial physiology [5]. Both ESAT-6 and TB10.4 are secreted by a Type VII secretion transport mechanism. ESAT-6 is secreted by ESX-1, whereas TB10.4 is secreted by ESX-3. The pathogenicity of the bacteria is dependent on the region of difference 1 (RD1), which largely encodes ESX-1. The primary cause for the debilitation of the bacteria throughout the development of BCG is the absence of this specific area. The secretion of ESX-1 and the subsequent export of ESAT-6 have been associated with several actions of mycobacteria that contribute to the development of disease, including as the suppression of IFN-γ production by human T cells, the hindrance of fusion between phagosomes and lysosomes, and the induction of death in host cells [6-8].

The process of designing a vaccine is both crucial and intricate since the goal is to create a candidate that can effectively stimulate both the humoral and cell-mediated immune responses against a specific disease. Over the past few decades, the field of vaccine design known as "Immunoinformatics" has achieved significant advancements and has contributed to the creation of vaccines [9]. Immunoinformatics is the utilisation of computational techniques and resources to study immunological processes. Immunoinformatics employs statistical, computational, mathematical, and biological expertise and methods to accurately store and analyse data pertaining to the immune system and its functions [10]. Immunity against infection can be achieved through two branches of the immune response: B-cell mediated and T-cell mediated. The humoral immune response, mediated by B-cells, rapidly neutralises infections, whereas the T-cell-mediated reaction creates immunological memory in the host. Memory is established by include prominent epitopes in the vaccination. The development of a multi-epitope vaccination utilising immunoinformatics entails a series of procedures [9].

In this study, we evaluated the fused form of ESAT-6/Tb10.4 recombinant protein by using immunoinformatic tools. Various immunoinformatic techniques, including Alphafold, Ellipro, VaxiJen, and IFNepitop, were developed to forecast the structure, antigenicity, and immune response triggered by the fused ESAT-6/Tb10.4 recombinant protein. Our analysis showed the fused protein could induce a type of cytokine such as interferon-gamma, which plays crucial roles in adaptive and innate immunity.

2 Methods

2.1 Sequences retrieval

The ESAT-6 and Tb10.4 sequences were retrieved from the NCBI database. A consensus sequence of ESAT-6 and Tb10.4 was generated from Indonesian *Mycobacterium tuberculosis* clinical samples. The ESAT-6 gene fused to the Tb10.4 gene by linking both genes with the GGSGG linker.

2.2 Immunoinformatic analysis

The fused ESAT-6/Tb10.4 gene sequence was uploaded into the AlphaFold (Abramson et al., 2024) Server (https://alphafoldserver.com) to predict the model structure of the fused gene. The model of the fused gene was saved in PDB file extension. Furthermore, the best model from AlphaFold 3 was analyzed using Ellipro (Ponomarenko et al., 2008) from IEDB to obtain the predicted linear and discontinuous B-cell epitopes of the fused ESAT-6/Tb10.4. The threshold value for linear and discontinuous epitopes was set at 0.5 to increase the validity of the prediction. The antigenicity and interferon-gamma release of the fused recombinant protein were evaluated by using VaxiJen and the IFNepitope server. The protein sequence of the fused protein was uploaded into the VaxiJen server. To increase the validity of the prediction, the threshold prediction was set at 0.5. Interferon-gamma release was evaluated using the IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/). The linear epitopes predicted from IEDB were re-analyzed using IFNepitope to evaluate their potency inducing the interferon-gamma release.

2.3 Cloning of fused ESAT-6/Tb10.4 gene

The ESAT-6 and Tb10.4 genes were amplified from pGEX-6P-1-SpyCatcher003-ESAT6 and pGEX-6P-1-SpyCatcher003-Tb10.4 (unpublish data), respectively. The ESAT-6 gene was amplified using ESAT Fw: TAATggatccATGACCGAACAGCAATGGAATTTTGC and ESAT Rv: cccaccggaaccgccAGCAAACATACCGGTCACATTGC. The Tb10.4 gene was amplified using Tb Fw: ggcggttccggtgggATGAGCCAGATTATGTATAATTATCCGGCG and Tb Rv: TAATctcgagACCGCCCCATTTCGCG. These primers were designed to introduce a *Bam*HI site of forward primer of the ESAT-6 gene and a *Xho*I site of reverse primer for the Tb10.4 genes. Overlapping PCR of ESAT-6 and Tb10.4 was performed using two-step PCR. The first step was to amplify each gene using its own primer set at 98ºC for 10 sec, 57ºC for 15 sec, and 72ºC for 30 sec, followed by a final extension step at 72ºC for 2 min. The PCR product from the first step was used as the template for the second step PCR and amplified using ESAT Fw: TAATggatccATGACCGAACAGCAATGGAATTTTGC and Tb Rv: TAATctcgagACCGCCCCATTTCGCG. The PCR fuse products were purified through MEGA quick-spin Plus Total Fragment DNA Purification Kit (Intron) as per the manufacturer's instructions. The pET-21d(+)and the PCR product were double digested with *Bam*HI and *Xho*I. The linearized plasmid DNA and insert fuse DNA were purified using the MEGA quick-spin Plus Total Fragment DNA Purification Kit followed by ligation using T4 DNA ligase and transformed into competent E. *coli* TOP10 cells. The design of the recombinant plasmid is shown in Figure 1.

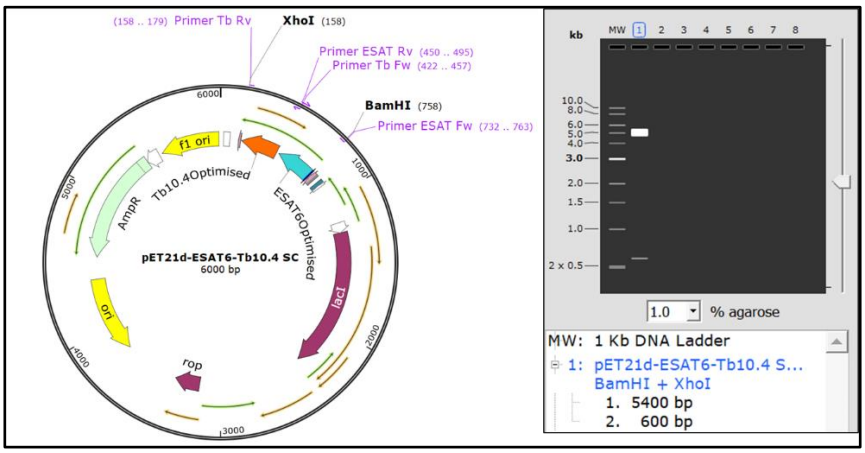

Fig. 1. Design of recombinant plasmid pET-21d(+)-ESAT6/Tb10.4. The position of all primers used in the study is shown in the plasmid construct. The predicted double digestion result of the recombinant plasmid is shown in the right picture.

3 Result and Discussion

To develop a new subunit vaccine for TB as a booster for the BCG vaccine, we carried out in silico study followed by cloning the gene target. The ESAT-6 and Tb10.4 gene sequences were retrieved from NCBI. A consensus sequence of each gene was generated from the retrieved sequences and the codon optimized for expression in the E. *coli* expression system. To generate a stable fusion recombinant protein, ESAT-6, and Tb10.4 were fused by the addition of the GGSGG linker sequence. Previous reports have shown additional linkers could increase the construction of stability and bioactive fusion proteins [11]. The sequence of fusion ESAT-6 and Tb10.4 (Figure 2A) was constructed under pET21d(+) (Figure 1).

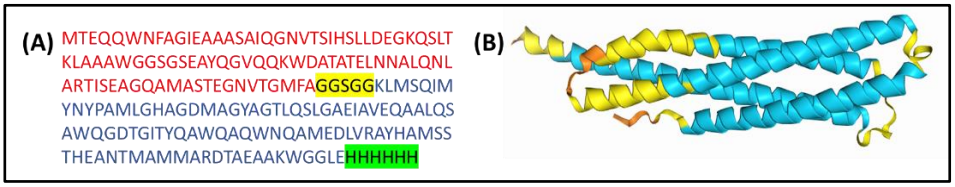

Fig. 2. Sequence and structure model of fuse ESAT-6/Tb10.4 recombinant protein. A. The fused ESAT-6/Tb10.4 amino acids sequence. Red font indicates the ESAT-6 gene, yellow highlight indicates the linker, blue font indicates the Tb10.4 gene and green highlight indicates the fusion tag histidine. B. The best model of fusion recombinant protein ESAT6/Tb10.4 generated by Alphafold 3.

The model of the structure protein of fused ESAT6/Tb10.4 was generated by Alphafold 3. The protein's validity structure is determined by calculating its pTM score. A pTM score greater than 0.5 indicates that the predicted fold for the complex is likely to be comparable to the actual structure [12]. The best model of our fused protein (Figure 2B) has a pTM score of 0.57 which is above the threshold score of Alphafold 3. Based on the prediction model and the pTM score, our fusion recombinant protein could be generated and expressed in the expression system.

Immunoinformatic evaluation of the vaccine design based on the fusion recombinant protein of ESAT6/Tb10.4 was analyzed using several *in silico* tools. The first evaluation related to the ability of the recombinant protein as the protective antigen was evaluated using the VaxiJen server. The VaxiJen model for bacterial protective antigen had 82% accuracy, 91% sensitivity, and 72% specificity. Even though the external validation of the server showed a lower result, however overall the performance of the prediction is satisfactory. [13]. The threshold for the prediction was set at 0.4 following the default setting from the server. Analysis from VaxiJen showed our construct has a protective score of 0.534 which indicates the fusion recombinant protein has potency as a vaccine.

Reliable prediction of antibodies especially B-cell epitopes remains challenging yet highly desirable for the design of vaccines. In this study, we exploited the ElliPro to determine the discontinues and linear B-cell epitopes of the fused ESAT6/Tb10.4. The ElliPro is based on the geometrical properties of protein structure [14] and the PDB file generated from the Alphafold 3 model was used as the template for predicting the B-cell of fused recombinant protein. Five discontinuous epitopes (Table 1) and linear epitopes (Table 2) were predicted as the best B-cell epitopes from the fused recombinant protein. One discontinuous epitope intersects with two linear epitopes (epitope 2: TKLAAAWGGSGSEAYQGVQQKWDA: 37--60aa and epitope 4: VEQAALQSAWQGDTGITYQAWQAQWNQ: 136--162aa). Furthermore, both epitopes were evaluated with an IFNepitope server to predict their ability to induce Interferon Gamma (IFN-γ). Interestingly, both epitopes showed positive results with IFN-γ scores were 4 and 3.7 for epitope 2 and epitope 4, respectively. Based on the immunoinformatic analysis, the fusion recombinant protein ESAT6/Tb10.4 showed a potent candidate for TB vaccine.

Table 1. Discontinues epitopes of Fuse ESAT-6/Tb10.4

No	Residues	Number of Residue	Score
1	A:M1, A:T2, A:E3, A:Q5	4	0.85
2	A:A187, A:T190, A:A191, A:E192, A:A193, A:A194, A:K195, A:W196, A:G197, A:G198, A:L199, A:E200, A:H201, A:H202	14	0.77
3	A:E31, A:Q34, A:T37, A:K38, A:L39, A:A40, A:A41, A:A42, A:W43, A:G44, A:G45, A:S46, A:G47, A:S48, A:E49, A:A50, A:Y51, A:Q52, A:G53, A:V54, A:Q55, A:Q56, A:K57, A:D59, A:A60, A:V136, A:E137, A:A139, A:A140, A:Q142, A:S143, A:A144, A:W145, A:Q146, A:G147, A:D148, A:T149, A:G150, A:I151, A:T152, A:Y153, A:Q154, A:A155, A:Q157, A:A158, A:Q159, A:N161	47	0.74
4	A:S85, A:G88, A:N89, A:T91, A:G92, A:M93, A:A95, A:G96, A:G97, A:S98, A:G99, A:G100, A:K101, A:M103, A:S104, A:Q105, A:M107	17	0.72
5	A:N7, A:A9, A:E12, A:A13, A:S16	5	0.57

Table 2. Linear epitopes of Fuse ESAT-6/Tb10.4

No	Chain	Start	End	Peptide	Number of residues	Score
1	A	190	206	TAEAAKWGGLEHHHHHH	17	0.83
2	**A**	**37**	**60**	**TKLAAAWGGSGSEAYQGVQQKWDA**	**24**	**0.75**
3	A	88	108	GNVTGMFAGGSGGKLMSQIMY	21	0.68
4	**A**	**136**	**162**	**VEQAALQSAWQGDTGITYQAWQAQWNQ**	**27**	**0.66**
5	A	1	9	MTEQQWNFA	9	0.63

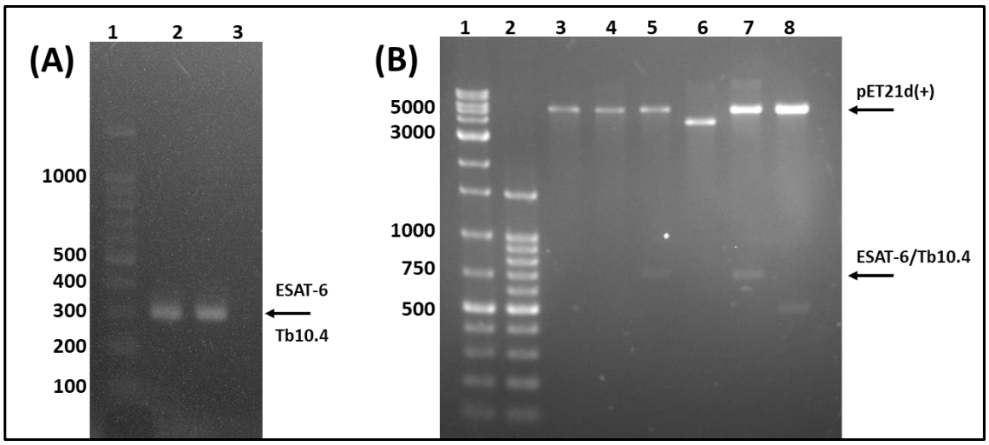

Fig. 3. Cloning recombinant plasmid pET21d(+)-ESAT6/Tb10.4. A. PCR product of ESAT-6 and Tb10.4. B. Verification of recombinant plasmid by double digestion. Lane 1: 1kb DNA ladder. Lane 2: 100 bp DNA ladder. Lane 3: digested pET21d(+). Lane 4--8: digested clones of recombinant plasmid.

The cloning of the fusion gene was initiated by amplifying each gene using the standard PCR protocol (Figure 3A). The overlapping PCR was carried out to fuse the gene (data not shown) followed by molecular cloning of the fragment into the pET21d(+). Two out of five clones showed positive recombinant plasmid verified by double digestion with *Bam*H1 and *Xho*1. Positive clones are indicated by the appearance of 2 bands, the higher band corresponds to plasmid pET21d(+) with a size of around 5600 bp, and the lower band corresponds to a fusion gene fragment with a size around 600 bp. Taken together, we succeeded in designing, evaluating, and constructing the fusion recombinant plasmid which will encode the fusion recombinant protein ESAT6/Tb10.4 in the E. *coli* expression system. Further study needs to be carried out to confirm the expression of the fusion genes and also characterize the recombinant protein in inducing the protective immune response against TB infection.

Determining B cell epitopes within a recombinant protein can greatly improve the efficacy of vaccines. By selectively targeting particular B cell epitopes, the vaccination can stimulate a robust and precise antibody response, so enhancing the overall immune defence against the disease [15]. Moreover, epitope-based vaccinations have the ability to reduce the likelihood of negative responses by eliminating non-essential or potentially dangerous elements of the pathogen. This comprehension facilitates the logical formulation of vaccines, enhancing their efficacy and simplifying their production process [16]. In addition, the combination of numerous epitopes can augment the vaccine's capacity to stimulate both humoral (antibody-mediated) and cellular (T cell-mediated) immunity, resulting in more potent and safer vaccinations. This has the potential to revolutionise the field of infectious disease prevention [17].

References

1. Scriba, T.J. and V. Mizrahi, *Renewing the Fight Against TB with an Old Vaccine.* Cell, 2020. **180**(5): p. 829-831.
2. Poyntz, H.C., et al., *Non-tuberculous mycobacteria have diverse effects on BCG efficacy against Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2014. **94**(3): p. 226-37.

3. Brandt, L., et al., *Failure of the Mycobacterium bovis BCG Vaccine: Some Species of Environmental Mycobacteria Block Multiplication of BCG and Induction of Protective Immunity to Tuberculosis.* Infection and Immunity, 2002. **70**(2): p. 672-678.
4. Woodworth, J.S., et al., *A Mycobacterium tuberculosis-specific subunit vaccine that provides synergistic immunity upon co-administration with Bacillus Calmette-Guérin.* Nat Commun, 2021. **12**(1): p. 6658.
5. Hoang, T., et al., *ESAT-6 (EsxA) and TB10.4 (EsxH) based vaccines for pre- and post-exposure tuberculosis vaccination.* PLoS One, 2013. **8**(12): p. e80579.
6. Lewis, K.N., et al., *Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guérin attenuation.* J Infect Dis, 2003. **187**(1): p. 117-23.
7. Ligon, L.S., J.D. Hayden, and M. Braunstein, *The ins and outs of Mycobacterium tuberculosis protein export.* Tuberculosis, 2012. **92**(2): p. 121-132.
8. Pym, A.S., et al., *Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti.* Molecular Microbiology, 2002. **46**(3): p. 709-717.
9. Rawat, S.S., et al., *Immunoinformatics Approaches for Vaccine Design: A Fast and Secure Strategy for Successful Vaccine Development.* Vaccines (Basel), 2023. **11**(2).
10. Oli, A.N., et al., *Immunoinformatics and Vaccine Development: An Overview.* Immunotargets Ther, 2020. **9**: p. 13-30.
11. Chen, X., J.L. Zaro, and W.C. Shen, *Fusion protein linkers: property, design and functionality.* Adv Drug Deliv Rev, 2013. **65**(10): p. 1357-69.
12. Abramson, J., et al., *Accurate structure prediction of biomolecular interactions with AlphaFold 3.* Nature, 2024. **630**(8016): p. 493-500.
13. Doytchinova, I.A. and D.R. Flower, *VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines.* BMC Bioinformatics, 2007. **8**: p. 4.
14. Ponomarenko, J., et al., *ElliPro: a new structure-based tool for the prediction of antibody epitopes.* BMC Bioinformatics, 2008. **9**: p. 514.
15. Biavasco, R. and M. De Giovanni, *The Relative Positioning of B and T Cell Epitopes Drives Immunodominance.* Vaccines, 2022. **10**.
16. Shawan, M., et al., *Advances in Computational and Bioinformatics Tools and Databases for Designing and Developing a Multi-Epitope-Based Peptide Vaccine.* Int J Pept Res Ther, 2023. **29**(4): p. 60.
17. Mortazavi, B., A. Molaei, and N.A. Fard, *Multi-epitopevaccines, from design to expression; an in silico approach.* Hum Immunol, 2024. **85**(3): p. 110804.