# Optimization of Soluble Expression of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD Fusion Proteins as Candidates for COVID-19 Intranasal Vaccines

Simson Tarigan[1], Sumarningsih[1], Gita Sekarmila[1], Apas[1], Ronald Tarigan[2], Ryandini Putri[1], Damai Ria Setyawati[3]

[1]Research Center for Veterinary Sciences, Research Organization for Health, National Research and Innovation (BRIN), Cibinong, Indonesia,
[2]School of Veterinary Medicine and Biomedical Sciences, IPB University, Bogor, Indonesia
[3]Research Center for Vaccine and Drugs, Research Organization for Health, National Research and Innovation (BRIN), Cibinong, Indonesia

*Corresponding author
Simson Tarigan
Email: sitariganta@gmail.com

## Abstract

Developing intranasal vaccines against pandemics and devastating airborne infectious diseases is imperative. The superiority of intranasal vaccines over injectable systemic vaccines is evident, but the challenge in developing effective intranasal vaccines is more substantial. Fusing a protein antigen with the catalytic domain of cholera toxin (CTA1) and the two-domain D of staphylococcal protein A (DD) has significant potential for intranasal vaccines. In the present study, we constructed two fusion proteins containing CTA1, tandem repeat linear epitopes of the SARS-CoV-2 spike protein (S14P5 or S21P2), and DD. The *in silico* characteristics and solubility of the fusion proteins CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were analyzed when overexpressed in *Escherichia coli*. Structural predictions indicated that each component of the fusion proteins was compatible with its origin. Both fusion proteins were predicted by

computational tools to be soluble when overexpressed in *E. coli*. Contrary to these predictions, the constructs exhibited limited solubility. The solubility did not improve even after lowering the cultivation temperature from 37°C to 18°C. Induction with IPTG at the early log phase, instead of the usual mid-log phase growth, significantly increased soluble CTA1-(S21P2)4-DD but not CTA1-(S14P5)4-DD. The solubility of overexpressed fusion proteins significantly increased when a non-denaturing detergent (Nonidet P40, Triton X100, or Tween 20) was added to the extraction buffer. In a scale-up purification experiment, the yields were low, only 1-2 mg/L of culture, due to substantial losses during the purification stages, indicating the need for further optimization of the purification process.

## Introduction

The respiratory tract is constantly exposed to the external environment and serves as a primary route for the entry of airborne pathogens, including those causing devastating diseases such as COVID-19, SARS, MERS, influenza, and tuberculosis. It is advantageous for these pathogens to be eliminated by the immune system in the nasal cavity before they enter the lungs or the rest of the body. Eliciting mucosal immune responses in the nasal cavity through intranasal vaccination is a promising approach, as the nasal cavity is equipped with an advanced local lymphoid system, specifically the nasopharynx-associated lymphoid tissue (NALT) [1]. Intranasal vaccines offer several advantages over traditional injectable vaccines, including the ability to induce systemic and mucosal immunity, which is particularly effective for pathogens entering through mucosal surfaces. Additionally, intranasal administration is needle-free, enhancing patient compliance and reducing the risk of needle-associated infections and injuries [2].

The most significant progress in developing intranasal vaccines has occurred for influenza. Several intranasal vaccines, such as FluMist or Fluenz (AstraZeneca) and Nasovac (Serum Institute of India), have been approved for human use. These intranasal vaccines have demonstrated superiority over systemic vaccines, eliciting robust immune responses, including mucosal-neutralizing antibodies and systemic protection against homologous and heterologous viruses, without additional adjuvants [2].

Possibly inspired by the success of the intranasal influenza vaccine, there is significant interest in developing intranasal vaccines for COVID-19. Various types of intranasal vaccines are being developed, including recombinant Sendai virus expressing the receptor-binding domain of SARS-CoV-2 [3], VLP-based vaccines paired with adjuvants [4], and DelNS1-nCoV-RBD live attenuated influenza vaccine [5]. These vaccines have shown efficacy in preclinical and clinical

studies, with some at the stage of clinical trials demonstrating safety, immunogenicity, and potential for broader protection against emerging variants like Omicron. Intranasal vaccination presents a promising approach to combat COVID-19.

Most of the aforementioned intranasal vaccines are attenuated or adenovirus-vectored vaccines. While these vaccines induce strong immunity, they require cold chains and often have unacceptable side effects [6]. Subunit vaccines, especially those based on epitopes, are expected to be safer. However, subunit vaccines without potent adjuvants appear ineffective in eliciting protective mucosal immune responses. Inadequate stimulation by a protein antigen may even lead to immune tolerance, which further complicating the issue. Cholera toxin is one of the most potent adjuvants for mucosal vaccination, but its high toxicity prevents its use as an adjuvant. Agren and colleagues have pioneered the elimination of the toxic nature of the protein by fusing the catalytic component of the toxin (CTA1) with a two-domain D of staphylococcal protein A (DD). The latter component enables the fusion protein to bind to B cells of all isotypes and then transform into plasma cells, producing immunoglobulins. The resulting fusion protein, CTA1-DD, retains the full adjuvanticity of cholera toxin but is entirely nontoxic [7-9]. The CTA1-DD demonstrates superior promotion of long-term immune responses compared to aluminum salts (Alum) and Ribi adjuvants [10]. An intranasal CTA1-DD-adjuvanted H3N2 split influenza vaccine elicited high titers of specific IgA in bronchoalveolar and vaginal lavages, as well as IgM and IgG in the sera of experimental animals [11]. The CTA1-DD is not only potent in inducing an IgA response but also effective in preventing the development of immune tolerance [12].

In addition to mixing antigens with CTA1-DD, protein antigens or peptides may be conjugated with CTA1-DD into a single fusion protein to ensure consistent adjuvanticity towards the protein antigen. An intranasal vaccine made by incorporating the ectodomain matrix-2 protein (M2e) of an influenza virus between CTA1 and DD was reported to be highly effective in mice, promoting high specific serum IgG and mucosal IgA and providing strong protection against a potentially lethal challenge infection with influenza virus [13, 14]. An intranasal vaccine for human respiratory syncytial virus (hRSV) made by conjugating the prefusion F protein (RBF) of the virus to the C-terminal end of CTA1-DD was reported to be effective based on vaccination and challenge trials in mice. The CTA1-DD-RBF vaccine stimulated the production of hRSV F-specific neutralizing antibodies (IgG1, IgG2a, sIgA) and T cell immunity in mice, effectively protecting the vaccinated animals from hRSV challenge [15].

In the present study, we aim to construct fusion proteins consisting of tandem repeat epitopes S14P5 or S21P2 with CTA1 and DD, resulting in CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD fusion proteins, which are prokaryotically expressible, functional, and soluble. Epitopes S14P5 and S21P4 are linear epitopes of the spike protein of SARS-CoV-2 that have been shown to induce neutralizing antibodies [16]. Previously, S14P5 and S21P4 have been formulated in tandem repeats and, when injected into rabbits, induced robust antibody responses that recognized the SARS-CoV-2 virus [17]. For these reasons, CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD are expected to elicit strong neutralizing antibody responses and, consequently, are suitable candidates for intranasal vaccines against COVID-19. The vaccines are expected to be thermostable because they are based on linear epitopes. In this study, we seek to identify potential challenges in protein expression and purification processes that may hinder the development of these fusion proteins as vaccine candidates. By elucidating the underlying mechanisms contributing to limited solubility and low yield, our study aims to provide valuable insights for optimizing the production of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, thereby enhancing their potential as viable components in future vaccine formulations. The approach used in this study should also be readily applicable to other infectious diseases.

## Materials and methods

### Construction and *in silico* analysis of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD

The CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD constructs were generated by conjugating the catalytic subunit of cholera toxin (CTA1) at the N-terminal end of the tandem repeat SARS-CoV-2 epitopes S14P5 or S21P2 and the two-domain D of the staphylococcal protein A at the C-terminal end (Fig 1A). The sequence for CTA1 was obtained from a previous study [18] with a modification where residue phenylalanine at residu132 was replaced with serine to increase protein solubility [19]. The amino acid sequence of domain D staphylococcal protein A was from an earlier publication [20], and the tandem repeats S14P5 or S21P2 were from previous studies [16, 17]. A peptide linker, GGGS, was placed between CTA1 and the peptide, between the peptide and DD, and between the peptide repeats.

The characteristics of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were predicted based on their respective amino acid sequences. The general attributes of the recombinant proteins were assessed using the ExPASy application (https://web.expasy.org/protparam/). Solubility

predictions were conducted employing NetSolP0.1 (https://services.healthtech.dtu.dk/services/NetSolP-1.0/) and SoluProt (https://loschmidt.chemi.muni.cz/soluprot/), while propensity to aggregate was evaluated using Aggrescan (http://bioinf.uab.es/aggrescan/). Tertiary structure prediction of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD was performed utilizing the open-source software Alphafold2 and Chimera X [21, 22].

**Expression of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD**

Synthetic genes for CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were generated by reverse translation of the amino acid sequences and codon optimization for prokaryotic expression. The synthetic genes were inserted into the pET38a(+) expression plasmids (Genscript Inc. USA). The plasmids were transformed into a competent BL21 strain of *E. coli* using a protocol described in a previous study, with some modifications [17]. Incubation temperatures of 37°C and 18°C were tested for 2 hours and 6 hours to obtain the highest amount of recombinant protein in soluble form. Induction was carried out at $OD_{600}$ values of 0.1 and 0.4. The concentration of IPTG for induction was 0.3 m*M* in all experiments. Different solutions were tested to extract proteins from bacterial cells. Various solutions, including native buffer (0.5 M NaCl, 0.1 M sodium phosphate. pH 9), Nonidet P40, Triton X-100, and Tween-20, each at concentrations of 0.01%, 0.05%, and 0.1% in the native buffer, were tested for protein extraction. Chemicals were sourced from Sigma Aldrich.

**Analysis of Total and Soluble Fusion Proteins**

For the analysis of total protein, 0.5 mL of bacterial culture was centrifuged at 10,000 x g for 5 minutes to pellet the cells. The resulting pellet was solubilized in 50 µL of SDS-PAGE sample buffer, and 10 µL of this suspension was loaded into a well of a 15-well polyacrylamide gel of the Mini-PROTEAN II electrophoresis apparatus (Bio-Rad). The percentage of polyacrylamide was 10% for the separating gel and 4% for the stacking gel.

To analyze soluble protein, 1 mL of bacterial culture was centrifuged at 10,000 x g for 5 minutes to pellet the cells. The pellet was resuspended in 0.5 mL of PBS containing 10 mM EDTA and sonicated for three intervals of 30 seconds each. Following sonication, the sample was centrifuged again at 10,000 x g for 5 minutes, and the supernatant was carefully collected. The

proteins in the supernatant were precipitated with trichloroacetic acid according to the standard protocol [23]. The precipitated proteins were then solubilized in 50 µL of SDS-PAGE sample buffer, and 10 µL of this solution was loaded into an SDS-PAGE gel as previously described.

For the immunoblot assay, proteins from the polyacrylamide gel were transferred onto a nitrocellulose membrane, and CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were identified respectively with rabbit-anti-(S14P5)4 and rabbit-anti-(S21P2)4 antibodies produced in our previous studies [17]. Protein concentrations were assessed by densitometry of protein bands on SDS-PAGE gel or immunoblot, using ImageJ open-source software (NIH, USA).

### Statistical analysis

Descriptive statistics were used to summarize the data. Differences in protein concentration were analyzed using the Wilcoxon rank-sum test, with *p*-values less than 0.05 considered statistically significant.

## RESULT

### In silico analysis of CTA1-(S14P5)4-DD and CTA1-(S21P2)4

The general characteristics of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, including the number of amino acids, molecular weight, aliphatic index, Grand Average of Hydropathicity (GRAVY), and instability index, are presented in Table 1. To facilitate a better understanding of the proteins, we included the well-known protein bovine serum albumin (BSA), whose amino acid sequence was obtained from GenBank (accession numberAAA51411.1), for comparison. Both CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD have molecular weights of approximately 47 kDa, are acidic (pI<7), and more hydrophilic than BSA, as indicated by their lower GRAVY scores. According to the instability index, both proteins were less stable than BSA. The solubility of CTA1-(S14P5)4-DD and CTA1-(S21P4)4-DD were predicted to be higher than the highly soluble BSA. This higher solubility was in agreement with the higher hydrophilicity. Moreover, the CTA1-(S14P5)4-DD and (S21P2)4 were predicted to have a lower propensity for aggregation than BSA, as indicated by the higher number of hot spot area (nHS) and normalized nHS for 100 residues (NnHS).

Table 1. General or Predicted Characteristics of CTA1-(S14P5)4-DD, CTA1-(S21P2)4-DD and Bovine Serum Albumin (BSA).

| Characteristics | CTA1-(S14P5)4-DD | CTA1-(S21P2)4-DD | BSA |
|---|---|---|---|
| No. amino acids[1] | 419 | 423 | 590 |
| Molecular weight[1] | 46.5 kDa | 46.9 kDa | 67.3 kDa |
| Isoelectric point (pI) [1] | 5.97 | 6.35 | 5.76 |
| Aliphatic index[1] | 55.73 | 60.69 | 75.73 |
| Grand average of hydropathicity (GRAVY)[1] | -0.910 | -0.866 | -0.482 |
| Instability index[1] | 53.6 | 59.3 | 40.91 |
| Solubility prediction (DeepSoluE[2]) | 0.9513 | 0.9647 | 0.8553 |
| Number of Hot Spots (nHS) (Aggrescan[3]) | 6 | 10 | 17 |
| Normalized nHS for 100 residues (NnHS) (Aggrescan[3]) | 1.432 | 2.364 | 2.881 |
| **Solubility prediction of protein over-expressed in *E. coli*** | | | |
| SoluPro[4] | 0.858 | 0.887 | not relevant |
| NetSolP[5] | 0.5975 | 0.5946 | not relevant |

Note:
[1]https://web.expasy.org/protparam/
[2] http://39.100.246.211:10505/DeepSoluE/
[3]http://bioinf.uab.es/aggrescan/
[4]https://loschmidt.chemi.muni.cz/soluprot/
[5]https://services.healthtech.dtu.dk/services/NetSolP-1.0/

Upon overexpression in *E. coli*, CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were predicted to exhibit solubility using online computational tools. The SoluPro application generated high prediction values, precisely 0.858 for CTA1-(S14P5)4-DD and 0.887 for CTA1-(S21P2)4-DD. In contrast, the NetSolP application yielded considerably lower predictions of only 0.6 for both proteins.

The schematic and tertiary structures of CTA1-(S14P5)4-DD and (S21P2)4 are shown in Fig 1. The schematic representation provides an overview of the domain organization and sequences of these construct (Fig 1A). The CTA1 domain consists of two groups of three antiparallel beta sheets, several short alpha helices, and connecting loops. The DD component comprised a pair of three long, parallel alpha helices connected by a long coil. The (S14P5)4 appeared as a long coil, whereas the (S21P2)4 appeared as four alpha helices with coils at both ends (Fig 1B,D). The structures of S14P5 and S21P2 in the predicted structure of CTA1-(S14P5)4-DD and (S21P2)4 were similar to those in the cryo-electron microscopy structure of SARS-CoV2 spike protein previously submitted to the Protein Data Bank ((PDB # 6VXX) (Fig 1C, E).

Fig 1.
**(A)** Schematic structures of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD.
**(B, D)** Tertiary structure of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD predicted by AlphaFold2 ChimeraX with CTA1 coloured in blue, (S21P2)4 or (S21P2)4 in red and DD in violet.
**(C, E)** The cryo-electron microscopy structure of the SARS-CoV2 spike protein (PDB # 6VXX) highlights the secondary structure of linear epitopes S14P5 and S21P2, colored in red.

## Expression of CTA1-(S14P5)4-DD and CTA1-(S14P5)4-DD

The *E. coli* cells transformed with the CTA1-(S14P5)4-DD or CTA1-(S21P2)4-DD constructs exhibited robust protein expression upon induction with IPTG (Fig 2A). No protein expression was observed before IPTG induction, indicating correct regulation of genetic constructs and a responsive expression system. Substantial expression levels were evidenced by the intense protein bands observed in the gels. Densitometry analysis determined the expression levels of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD to be 30.5% and 45.8% of total cell protein, respectively. Immunoblot analysis using monospecific anti-(S14P5)4 and anti-(S21P2)4 antibodies confirmed the identity of the expressed proteins (Fig 2B, C), validating their expected characteristics. The molecular weights of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were determined to be 61.2 kDa and 61.4 kDa, respectively, significantly higher than the calculated weights based on their amino acid compositions, which were 46.5 kDa and 46.9 kDa, respectively (Table 1).

Fig 2. Total proteins of *E. coli* transformed with expression plasmid carrying either *CTA1-(S14P5)4-DD* or *CTA1-(S21P2)4-DD* genes. A. Three colonies grown in kanamycin-supplemented LB broth harvested before (A, B, and C) and after induction with IPTG (A', B' and C'). B and C: The *CTA1-(S14P5)4-DD*- and *CTA1-(S21P2)4-DD*-transformed *E. coli* probed with relevant specific antibodies.

## Effect of growth stage ($OD_{600}$) at induction, temperature and duration of incubation

Contrary to the in silico prediction, both CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD were poorly soluble when overexpressed in E. coli. The soluble fractions constituted only a negligible portion of the total proteins, and exact quantification was often unreliable due to smearing observed in many protein bands (S1 Fig, S1 Table). Lowering the culture temperatures from 37°C to 18°C did not increase protein solubility ($p>0.05$). In fact, lowering the temperature

decreased the amount of soluble CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD. Prolonging incubation from 3 to 6 hours increased the amount of soluble proteins ($p<0.05$). Induction at an early log phase ($OD_{600}$ of 0.1) resulted in higher production of soluble CTA1-(S21P2)4-DD ($p<0.05$), but not for CTA1-(S14P5)4-DD (S2 and S3 Tables).

**Effect of detergent in the extraction and purification buffer**

Protein purification using the native buffer (0.5 *M* NaCl, 0.1 *M* sodium phosphate, pH 9) with the NiNTA purification system yielded low amounts of CTA1-(S14P5)4-DD and even lower amounts of CTA1-(S21P2)4-DD (Fig 3). Adding detergents to the native buffer significantly increased the yields, particularly for CTA1-(S21P2)4-DD. Supplementing the buffer with 0.01% Nonidet P40 or Triton X-100 resulted in a fivefold increase in CTA1-(S14P5)4-DD yield and a seven- to eightfold increase for CTA1-(S21P2)4-DD, as measured by densitometry. However, 0.01% Tween 20 increased yields only twofold for both proteins. Further yield improvements were observed with higher concentrations of detergents, including Tween 20 at 0.05% and 0.1%. The optimal detergent concentration appears to be around 0.1%, as increasing the concentration from 0.05% to 0.1% did not result in significant additional increases in soluble protein amounts.

Fig 3. Effect of Detergents on the Yield of Proteins Purified Using NiNTA Column

**Effectivity of purification of CTA1-(S14P5)4-DD and CTA1-(S14P5)4-DD**

The efficacy of purifying CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD was tested under the optimized conditions determined in this study: 37°C culture temperature and 0.1% Nonidet-P40 extraction buffer, scaled up to a larger culture volume. The results of these experiments are detailed in Tables 2 and 3 and S2 Fig. After a 2-hour induction period, the bacterial cells contained 24-42 mg/L of CTA1-(S14P5)4-DD and 32-40 mg/L of CTA1-(S21P2)4-DD, based on densitometry using BSA as the standard. Despite the substantial initial protein amounts, the yields were low at the end of the purification process. Following elution, dialysis, and concentration in buffers without detergent, the yields were less than 5% of the total protein present in the bacterial cells for both proteins. Adding Nonidet P-40 to the elution and dialysis buffers only marginally increased the yields to 2.4 mg/L of culture for CTA1-(S14P5)4-DD and 1.15 mg/L for CTA1-(S21P2)4-DD. These yields in the scale-up volume experiment contrasted with previous experiments, where the addition of Nonidet P-40 to the extraction buffer led to a sevenfold increase in the yield of CTA1-(S14P5)4-DD and a sixteen-fold increase in the yield of CTA1-(S21P2)4-DD (Fig 3). The most significant protein losses occurred during the NiNTA

column separation stage, with over 90% of the target protein being lost during this process. During the lysate preparation stage, there was only a slight reduction, likely due to overestimated protein quantification caused by target band smearing, possibly indicating contamination with bacterial nucleic acids resulting from sonication (S2 Fig, Lane 2). Further substantial losses occurred during concentration and desalting in PBS, with more than 50% of the target proteins present in the eluate being lost. Adding Nonidet-P40 to the elution increased recovery, although significant losses could not be prevented. The discrepancy in yields between the small and large-volume experiments necessitates further fine-tuning of the experimental conditions.

**Table 2. Purification of CTA1-(S14P5)4-DD at different stages of the purification process, comparing elution without Nonidet-P40 and with Nonidet-P40.**

| Purification stage | Eluant without Nonidet-P40 | | | Eluant containing Nonidet P-40 | | |
|---|---|---|---|---|---|---|
| | Protein (µg) in the band | Protein µg/mL culture | Recovery | Protein (µg) in the band | Protein µg/mL culture | Recovery |
| Original cells (~0.05 mL culture) | 1.2 | 24 | 100% | 2.1 | 42 | 100% |
| Cells lysate (~0.125 mL culture) | 1.9 | 15.2 | 63.3% | 5.5 | 44 | 105% |
| Eluate (~0.5 mL culture) | 0.6 | 1.7 | 7.1% | 1.0 | 2 | 4.8% |
| Dialysed and concentrated eluate (2 mL) | 1.1 | 0.55 | 4.2% | 4.8 | 2.4 | 5.7 % |

**Table 3. Purification of CTA1-(S21P2)4-DD at different stages of the purification process, comparing elution without Nonidet-P40 and with Nonidet-P40.**

| Purification stage | Eluant without Nonidet-P40 | Eluant containing Nonidet P-40 |
|---|---|---|

| | Protein (µg) in the band | Protein µg/mL culture | Recovery | Protein (µg) in the band | Protein µg/mL culture | Recovery |
|---|---|---|---|---|---|---|
| Original cells (~0.05 mL culture) | 2 | 40 | 100% | 1.6 | 32 | 100% |
| Cells lysate (~0.125 mL culture) | 2.5 | 20 | 50 % | 3.5 | 28 | 88.5 % |
| Eluate (0.5 mL culture) | 1.3 | 2.6 | 6.5% | 1.2 | 2.4 | 7.5 % |
| Dialysed and concentrated eluate (~2 mL culture) | 1.9 | 0.95 | 2.4 % | 2.3 | 1.15 | 3.6 % |

## DISCUSSION

In the present study, we construct two fusion proteins, CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, as candidates for intranasal vaccines against SARS CoV-2. Each protein contains a linear epitope S14P5 or S21P2 of SARS CoV-2 spike protein, which is efficacious in inducing neutralizing antibodies [16]. The linear epitopes were made in the form of four tandem repeats. Our previous study confirmed that the S14P5 and S21P2 in this form enhanced the immunogenicity of the epitopes, and the induced antibodies effectively recognized the SARS-CoV-2 virus[17]. To further enhance the immune responses and to optimize them as intranasal vaccines, each of the tandem repeat epitopes has been conjugated to the catalytic subunit of cholera toxin (CTA1), and the domain D of staphylococcal protein A in dimer form (DD). The choice of domain D, instead of other domains, was beneficial due to its ability to bind not only IgG but also IgA and IgM [20].

*In silico* analysis of protein structure and function has become indispensable in modern biomedical research, particularly with the rapid advancement of bioinformatics tools. These computational methods can predict protein behavior, provide insights into protein interactions, stability, and potential binding sites, and guide the design of experiments, thus saving time and resources. To assess the functionality of the components within the fusion proteins, CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, we performed *in silico* analyses to predict their compatibility with their native structures. The three-dimensional structures of the fusion proteins

were predicted using AlphaFold-2, a computational method that predicts 3D protein structures from their respective amino acid sequences with near-experimental accuracy [22].

The structure of the CTA1-(S14P5)4-DD and CTA1-(S14P5)4-DD components, as predicted by AlphaFold-2, were compatible with their original structure, supporting the proper function of the proteins. The CTA1, which is the catalytic, ADP-ribosyltransferase and a NAD-glycohydrolase domain of cholera toxin, can be delineated into three discrete regions: $CTA1_1$ (residues 1-132), $CTA1_2$ (residues 133 to 161), and $CTA1_3$ (residues 162-193). The $CTA1_1$ forms a compact globular unit characterized by a combination of alpha-helices and beta-strands, harboring a catalytic cleft presumed to be the site of NAD and substrate binding. The $CTA1_2$ acts as a flexible linker bridging $CTA1_1$ and $CTA1_3$. The $CTA1_3$ is marked by a dense arrangement of hydrophobic residues [18]. Our predicted molecular model of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD aligns well with the established crystallographic structure. Specifically, residues 1-132 exhibit a similar secondary structure composition of alpha-helices and beta-strands, suggesting the presence of the $CTA1_1$ subunit. Residues 133-161 manifest as an elongated coil, consistent with the characteristics of $CTA1_2$, while residues 162-193 adopt a conformation resembling a tangled loop interspersed with short alpha helices, indicative of $CTA1_3$.

The DD component of the proteins appeared as a pair of three long, parallel alpha helices. This structure is compatible with the crystal structure domain D of staphylococcal protein A previously reported [20]. The secondary structure of epitopes S14P5, which appeared as a long loop, and S21P2, which appeared as an alpha helix and loops, were compatible with the relevant segment in the crystal structure of spike protein SARS CoV2 deposited in the Protein Data Bank [24].

Proper design constitutes the initial step in obtaining CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD that function as intended. The subsequent crucial step involves the expression and purification of the constructed fusion proteins, ensuring the preservation of its structure and function. Prokaryotic expression remains the preferred choice for protein expression due to its simplicity and cost-effectiveness. However, a significant challenge in prokaryotic expression is the formation of proteins in insoluble aggregates [25, 26]. Expressing proteins in soluble forms within *E. coli* offers notable advantages. Soluble expression guarantees the maintenance of proteins' native conformation, thereby preserving their functionality. In contrast, expression in inclusion bodies often requires denaturation and subsequent renaturation steps, posing risks of

protein function loss. Moreover, isolating proteins from insoluble inclusion bodies entails higher costs and longer processing times [27].

Predicting protein solubility before initiating expression confers significant advantages, particularly for engineered constructs like CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD. Therefore, it was imperative to evaluate their solubility, primarily upon overexpression in *E. coli*. However, current bioinformatic tools still exhibit limited reliability in predicting the solubility of overexpressed proteins, as evident from this study. Notably, based on their amino acid sequences, both CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD demonstrated high solubility. In fact, these proteins displayed a notably higher probability of solubility compared to bovine serum albumin (BSA), a protein renowned for its solubility. Although highly soluble proteins are not necessarily soluble when over-expressed in *E. coli*, this is true for both CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, which were predicted to be soluble [28].

The solubility of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD on overexpression was predicted by two computational tools: (1) SoluPro, which predicts solubility based on amino acid composition, dipeptide composition, and physicochemical properties, employing a machine learning approach, and (2) NetSolP, which predicts solubility based on amino acid composition, secondary structure, and solvent accessibility employing a neural network-based approach. These tools have undergone significant improvements, achieving accuracies of 74% for SOLUPro and 70% for NetSolP, surpassing many previous methods [29, 30]. However, despite these advancements, there remains a risk of misprediction, particularly for engineered proteins like CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, as these computational tools may not have been trained on similar proteins.

In experiments aimed at optimizing the solubility of expressed protein in *E. coli*, quantifying expressed protein is challenging. In the present study, we used densitometry techniques to quantify the proteins in SDS PAGE gels or immunoblots with specific antibodies. This technique can measure both soluble and insoluble proteins. Additionally, it can be used to measure proteins in mixtures with other unrelated proteins present in the samples. Despite these advantages, the technique requires high-resolution SDS PAGE gels or immunoblots that allow for the complete separation of proteins, which is often unattainable.

Formation of insoluble expressed proteins as inclusion bodies emerged as one of the most significant obstacles in prokaryotic expression systems. Inclusion bodies are formed due to a rapid rate of protein expression, a key objective in recombinant protein production. When

protein expression exceeds the host cells' capacity for post-translational modifications and folding, misfolding occurs, leading to aggregation into inclusion bodies as hydrophobic residues become exposed. Slowing down the expression rate by modifying culture conditions, such as lowering culture temperature and the concentration of induction agents, comprise the most practical approach to increasing the solubility of expressed proteins. Lowering temperature also diminishes the hydrophobic interaction between the expressed proteins, thus reducing aggregation [26].

Attempts to convert expressed proteins in *E. coli* from insoluble inclusion bodies to soluble forms by lowering the cultivation temperature have proven successful in many proteins. For instance, a three-fold increase in the soluble fraction of Green Fluorescent Protein (GFP) was achieved by lowering the temperature from 37°C to 16°C [31]. Similarly, human interferon-α2 and γ, which formed insoluble aggregates when expressed in *E. coli* at 37°C, demonstrated increased solubility by 30-90% when cells were cultivated at 23°C [32]. The increased solubility of heterologous proteins expressed in *E. coli* at lower temperatures has also been observed in various other proteins, including β-lactamase, human epidermal growth factor, human hemoglobin, and β-galactosidase [33-35]. However, the effectiveness of lowering the culture temperature in increasing protein solubility varies among different proteins [26]. As demonstrated in this study, lowering the cultivation temperature to 18°C instead of the typical 37°C did not yield the expected increase in solubility.

Moreover, in a previous study, induction of *P5βR*-transformed *E. coli* with IPTG during the early-log phase substantially increased the solubility of the expressed P5βR (progesterone 5β-reductase [36]. The increase in the solubility of expressed proteins by IPTG induction at the early log phase is likely dependent on the protein. In the present study, expression of soluble CTA1-(S21P2)4-DD significantly increased, but not CTA1-(S14P5)4-DD. It is unknown why the solubility of some expressed proteins increases and not others.

The production of fusion proteins, CTA1-DD or CTA1-peptide-DD, utilizing *E. coli* expression systems has been extensively explored in various studies. Despite numerous attempts, the expression of CTA1-DD or CTA1-peptide-DD consistently yields insoluble inclusion bodies, with no reported instances of soluble expression. This phenomenon has been observed across different *E. coli* expression systems, including the pUC19 expression vector with TG-1 *E. coli* in 2YT medium [8, 19, 37-39], DH5 *E. coli* in SYPPG medium [13], as well as the use of pET28a with BL-21 *E. coli* strains [15]. Despite forming insoluble inclusion bodies, the extracted

418 proteins have been successfully purified using guanidine HCl and chromatography techniques,
419 followed by subsequent renaturation steps. Remarkably, the purified proteins have exhibited
420 functionality despite these initial hurdles, as evidenced by various assays. Specifically, the ADP-
421 ribosyltransferase activity of CTA1 has been validated through NAD: agmatine assay [7, 19, 37,
422 39], while the DD component has been demonstrated to bind specifically to immunoglobulin
423 receptors on B cells [7].

424 In previous studies, purification of the fusion protein CTA1-DD expressed in *E. coli* as inclusion
425 bodies involved solubilization with guanidine-HCl, isolation using a IgG-Sepharose-affinity
426 column, and renaturation, resulting in yields of 8-60 mg/L culture [19, 40]. Although it is
427 inappropriate to directly compare different proteins in this aspect of purification, the yields of
428 CTA1-(S14P5)-DD and CTA1-(S21P2)-DD using native techniques were significantly lower.
429 Despite efforts to enhance solubility in the present study, purification using native techniques
430 yielded only 1-2 mg of pure proteins/L of culture.

431 The low yield of native CTA1-(S14P5)-DD and CTA1-(S21P2)-DD was due to substantial
432 losses during the purification stages. Higher yields are likely achievable by optimizing the
433 purification process, as the expressed fusion proteins constitute 30-45% of the total bacterial
434 protein after just 3 hours of induction. Additionally, as found in this study, the inclusion of non-
435 denaturing detergents such as Nonidet P40, Triton X100, or Tween 20 in the extraction buffer
436 effectively solubilizes the expressed CTA1-(S14P5)-DD and CTA1-(S21P2)-DD.

437 The approach used in this study for intranasal vaccine development can be easily adapted to
438 other infectious diseases. For diseases that have been existed for some time, lists of linear
439 epitopes that evoke neutralizing antibodies are likely to have been identified. Fusion proteins
440 consisting of these tandems repeat epitopes, and CTA1-DD can be constructed for mucosal
441 vaccines. For new emerging infectious diseases, such as those that may arise in the future, the
442 complete sequence of the causal agents, similar to what occurred with SARS-CoV-2, will be
443 available immediately. Linear epitopes can be identified from protein sequences using
444 computational tools, which are becoming increasingly accurate in their predictions. As a result,
445 vaccines could be available immediately before the disease spreads to more expansive areas.

446 In summary, two fusion proteins, CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, have been
447 constructed as candidates for intranasal vaccines against COVID-19. The constructs were
448 expressed abundantly in *E. coli* as insoluble inclusion bodies. The solubility of the expressed

proteins did not increase by lowering the cultivation temperature. Expression of soluble CTA1-(S21P2)4-DD, but not CTA1-(S14P5)4-DD, increased when the culture was induced with IPTG at the early log-phase growth. The solubility of both CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD was significantly enhanced by adding non-denaturing detergents to the extraction buffer.

## Acknowledgments

The authors would like to express their sincere gratitude to Dr. Fx. Sudirman, Director of Biotis Pharmaceuticals Indonesia, for his invaluable support in the publication of this study.

## References

1. Date Y, Ebisawa M, Fukuda S, Shima H, Obata Y, Takahashi D, et al. NALT M cells are important for immune induction for the common mucosal immune system. Int Immunol. 2017;29(10):471-8. Epub 2017/12/01. doi: 10.1093/intimm/dxx064. PubMed PMID: 29186424.

2. Esposito S, Montinaro V, Groppali E, Tenconi R, Semino M, Principi N. Live attenuated intranasal influenza vaccine. Hum Vaccin Immunother. 2012;8(1):76-80. Epub 2012/01/19. doi: 10.4161/hv.8.1.18809. PubMed PMID: 22251995.

3. Morimoto S, Saeki K, Takeshita M, Hirano K, Shirakawa M, Yamada Y, et al. Intranasal Sendai virus-based SARS-CoV-2 vaccine using a mouse model. Genes Cells. 2023;28(1):29-41. Epub 2022/11/20. doi: 10.1111/gtc.12992. PubMed PMID: 36401755.

4. Lee KS, Rader NA, Miller-Stump OA, Cooper M, Wong TY, Shahrier Amin M, et al. Intranasal VLP-RBD vaccine adjuvanted with BECC470 confers immunity against Delta SARS-CoV-2 challenge in K18-hACE2-mice. Vaccine. 2023;41(34):5003-17. Epub 2023/07/06. doi: 10.1016/j.vaccine.2023.06.080. PubMed PMID: 37407405; PubMed Central PMCID: PMCPMC10300285.

5. Deng S, Liu Y, Tam RC, Chen P, Zhang AJ, Mok BW, et al. An intranasal influenza virus-vectored vaccine prevents SARS-CoV-2 replication in respiratory tissues of mice and hamsters. Nat Commun. 2023;14(1):2081. Epub 2023/04/13. doi: 10.1038/s41467-023-37697-1. PubMed PMID: 37045873; PubMed Central PMCID: PMCPMC10092940 related to the generation and application of DelNS1 live attenuated influenza vaccines and the associated platform, with H.C., P.W., and K-Y.Y. included as co-inventors. There is no restriction on the publication of data. The other authors declare that they have no competing interests.

6. Zhang C, Berg A, Joe CCD, Dalby PA, Douglas AD. Lyophilization to enable distribution of ChAdOx1 and ChAdOx2 adenovirus-vectored vaccines without refrigeration. NPJ Vaccines. 2023;8(1):85. Epub 2023/06/06. doi: 10.1038/s41541-023-00674-2. PubMed PMID: 37277337; PubMed Central PMCID: PMCPMC10240132 of simian adenovirus-vectored vaccines. ADD has received research grant funding and payment for consultancy from AstraZeneca, unconnected to the current work.

7. Agren L, Sverremark E, Ekman L, Schon K, Lowenadler B, Fernandez C, et al. The ADP-ribosylating CTA1-DD adjuvant enhances T cell-dependent and independent responses by direct action on B cells involving anti-apoptotic Bcl-2- and germinal center-promoting effects. J Immunol. 2000;164(12):6276-86. Epub 2000/06/08. doi: 10.4049/jimmunol.164.12.6276. PubMed PMID: 10843681.

8. Agren LC, Ekman L, Lowenadler B, Lycke NY. Genetically engineered nontoxic vaccine adjuvant that combines B cell targeting with immunomodulation by cholera toxin A1 subunit. J Immunol. 1997;158(8):3936-46. Epub 1997/04/15. PubMed PMID: 9103464.

9. Agren L, Lowenadler B, Lycke N. A novel concept in mucosal adjuvanticity: the CTA1-DD adjuvant is a B cell-targeted fusion protein that incorporates the enzymatically active cholera toxin A1 subunit. Immunol Cell Biol. 1998;76(3):280-7. Epub 1998/07/31. doi: 10.1046/j.1440-1711.1998.00750.x. PubMed PMID: 9682972.

10. Bemark M, Bergqvist P, Stensson A, Holmberg A, Mattsson J, Lycke NY. A unique role of the cholera toxin A1-DD adjuvant for long-term plasma and memory B cell development. J Immunol. 2011;186(3):1399-410. Epub 2011/01/05. doi: 10.4049/jimmunol.1002881. PubMed PMID: 21199899.

11. Fan XT, Wang YL, Su QD, Qiu F, Yi Y, Jia ZY, et al. Intranasal Immunization Using CTA1-DD as a Mucosal Adjuvant for an Inactivated Influenza Vaccine. Biomed Environ Sci. 2019;32(7):531-40. Epub 2019/07/25. doi: 10.3967/bes2019.070. PubMed PMID: 31331437.

12. Lycke N. From toxin to adjuvant: basic mechanisms for the control of mucosal IgA immunity and tolerance. Immunol Lett. 2005;97(2):193-8. Epub 2005/03/09. doi: 10.1016/j.imlet.2004.12.008. PubMed PMID: 15752558.

13. Eliasson DG, El Bakkouri K, Schon K, Ramne A, Festjens E, Lowenadler B, et al. CTA1-M2e-DD: a novel mucosal adjuvant targeted influenza vaccine. Vaccine. 2008;26(9):1243-52. Epub 2008/02/05. doi: 10.1016/j.vaccine.2007.12.027. PubMed PMID: 18243429.

14. Schussek S, Bernasconi V, Mattsson J, Wenzel UA, Stromberg A, Gribonika I, et al. The CTA1-DD adjuvant strongly potentiates follicular dendritic cell function and germinal center formation, which results in improved neonatal immunization. Mucosal Immunol. 2020;13(3):545-57. Epub 2020/01/22. doi: 10.1038/s41385-020-0253-2. PubMed PMID: 31959882; PubMed Central PMCID: PMCPMC7223721.

15. Li H, Ren H, Zhang Y, Cao L, Xu W. Intranasal vaccination with a recombinant protein CTA1-DD-RBF protects mice against hRSV infection. Sci Rep. 2021;11(1):18641. Epub 2021/09/22. doi: 10.1038/s41598-021-97535-6. PubMed PMID: 34545126; PubMed Central PMCID: PMCPMC8452643.

16. Poh CM, Carissimo G, Wang B, Amrun SN, Lee CY, Chee RS, et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat Commun. 2020;11(1):2806. Epub 2020/06/03. doi: 10.1038/s41467-020-16638-2. PubMed PMID: 32483236.

17. Tarigan S, Dharmayanti N, Sugiartanti D, Putri R, Andriani, Nuradji H, et al. Characterization of two linear epitopes SARS CoV-2 spike protein formulated in tandem repeat. PLoS One. 2023;18(1):e0280627. Epub 2023/01/21. doi: 10.1371/journal.pone.0280627. PubMed PMID: 36662754; PubMed Central PMCID: PMCPMC9857970.

18. Zhang RG, Scott DL, Westbrook ML, Nance S, Spangler BD, Shipley GG, et al. The three-dimensional crystal structure of cholera toxin. J Mol Biol. 1995;251(4):563-73. Epub 1995/08/25. doi: 10.1006/jmbi.1995.0456. PubMed PMID: 7658473.

19. Agren L, Norin M, Lycke N, Lowenadler B. Hydrophobicity engineering of cholera toxin A1 subunit in the strong adjuvant fusion protein CTA1-DD. Protein Eng. 1999;12(2):173-8. Epub 1999/04/09. doi: 10.1093/protein/12.2.173. PubMed PMID: 10195289.

20. Graille M, Stura EA, Corper AL, Sutton BJ, Taussig MJ, Charbonnier JB, et al. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A. 2000;97(10):5399-404. Epub 2000/05/11. doi: 10.1073/pnas.97.10.5399. PubMed PMID: 10805799; PubMed Central PMCID: PMCPMC25840.

21. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 2018;27(1):14-25. Epub 2017/07/16. doi: 10.1002/pro.3235. PubMed PMID: 28710774; PubMed Central PMCID: PMCPMC5734306.

22. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9. Epub 2021/07/16. doi: 10.1038/s41586-021-03819-2. PubMed PMID: 34265844; PubMed Central PMCID: PMCPMC8371605 have filed non-provisional patent applications 16/701,070 and PCT/EP2020/084238, and provisional patent applications 63/107,362, 63/118,917, 63/118,918, 63/118,921 and 63/118,919, each in the name of DeepMind Technologies Limited, each pending, relating to machine learning for predicting protein structures. The other authors declare no competing interests.

23. Rosenberg IM. Protein Analysis and Purification. Benchtop Technique. . Birkhauser, Boston p79. 1996. Epub 2nd.

24. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181(2):281-92 e6. Epub 2020/03/11. doi: 10.1016/j.cell.2020.02.058. PubMed PMID: 32155444; PubMed Central PMCID: PMCPMC7102599.

25. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172. Epub 2014/05/27. doi: 10.3389/fmicb.2014.00172. PubMed PMID: 24860555; PubMed Central PMCID: PMCPMC4029002.

26. Bhatwa A, Wang W, Hassan YI, Abraham N, Li XZ, Zhou T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. Front Bioeng Biotechnol. 2021;9:630551. Epub 2021/03/02. doi: 10.3389/fbioe.2021.630551. PubMed PMID: 33644021; PubMed Central PMCID: PMCPMC7902521.

27. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. Trends Biotechnol. 2006;24(4):179-85. Epub 2006/03/01. doi: 10.1016/j.tibtech.2006.02.007. PubMed PMID: 16503059.

28. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnol. 2004;4:32. Epub 2004/12/16. doi: 10.1186/1472-6750-4-32. PubMed PMID: 15598350; PubMed Central PMCID: PMCPMC544853.

29. Hon J, Marusiak M, Martinek T, Kunka A, Zendulka J, Bednar D, et al. SoluProt: prediction of soluble protein expression in Escherichia coli. Bioinformatics. 2021;37(1):23-8. Epub 2021/01/09. doi: 10.1093/bioinformatics/btaa1102. PubMed PMID: 33416864; PubMed Central PMCID: PMCPMC8034534.

30. Thumuluri V, Martiny HM, Almagro Armenteros JJ, Salomon J, Nielsen H, Johansen AR. NetSolP: predicting protein solubility in Escherichia coli using language models. Bioinformatics. 2022;38(4):941-6. Epub 2022/01/29. doi: 10.1093/bioinformatics/btab801. PubMed PMID: 35088833.

31. Vera A, González-Montalbán N, Arís A, Villaverde A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng. 2007;96(6):1101-6. Epub 2006/10/03. doi: 10.1002/bit.21218. PubMed PMID: 17013944.

32. Schein CH, Noteborn MHM. Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. Bio/Technology. 1988;6(3):291-4. doi: 10.1038/nbt0388-291.

33. Chalmers JJ, Kim E, Telford JN, Wong EY, Tacon WC, Shuler ML, et al. Effects of temperature on Escherichia coli overproducing beta-lactamase or human epidermal growth factor. Applied and Environmental Microbiology. 1990;56(1):104-11. doi: doi:10.1128/aem.56.1.104-111.1990.

34. Vasina JA, Baneyx F. Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the Escherichia coli cspA and tac promoter systems. Protein Expr Purif. 1997;9(2):211-8. Epub 1997/03/01. doi: 10.1006/prep.1996.0678. PubMed PMID: 9056486.

35. Weickert MJ, Pagratis M, Curry SR, Blackmore R. Stabilization of apoglobin by low temperature increases yield of soluble recombinant hemoglobin in Escherichia coli. Appl Environ Microbiol. 1997;63(11):4313-20. Epub 1997/11/15. doi: 10.1128/aem.63.11.4313-4320.1997. PubMed PMID: 9361418; PubMed Central PMCID: PMCPMC168751.

36. San-Miguel T, Pérez-Bermúdez, P. and Gavidia, I. Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. SpringerPlus. 2013;2:89-93.

37. Eriksson AM, Schon KM, Lycke NY. The cholera toxin-derived CTA1-DD vaccine adjuvant administered intranasally does not cause inflammation or accumulate in the nervous tissues. J Immunol. 2004;173(5):3310-9. Epub 2004/08/24. doi: 10.4049/jimmunol.173.5.3310. PubMed PMID: 15322194.

38. Nedrud JG, Bagheri N, Schon K, Xin W, Bergroth H, Eliasson DG, et al. Subcomponent vaccine based on CTA1-DD adjuvant with incorporated UreB class II peptides stimulates protective Helicobacter pylori immunity. PLoS One. 2013;8(12):e83321. Epub 2014/01/07. doi: 10.1371/journal.pone.0083321. PubMed PMID: 24391754; PubMed Central PMCID: PMCPMC3877028.

39. Mowat AM, Donachie AM, Jagewall S, Schon K, Lowenadler B, Dalsgaard K, et al. CTA1-DD-immune stimulating complexes: a novel, rationally designed combined mucosal vaccine adjuvant effective with nanogram doses of antigen. J Immunol. 2001;167(6):3398-405. Epub 2001/09/07. doi: 10.4049/jimmunol.167.6.3398. PubMed PMID: 11544331.

40. Agren LC, Ekman L, Lowenadler B, Nedrud JG, Lycke NY. Adjuvanticity of the cholera toxin A1-based gene fusion protein, CTA1-DD, is critically dependent on the ADP-ribosyltransferase and Ig-binding activity. J Immunol. 1999;162(4):2432-40. Epub 1999/02/11. PubMed PMID: 9973526.

# Supporting information

**S1 Fig.** Visualization of CTA1-(S14P5)4-DD (A) and CTA1(S21P2)4-DD (B) proteins using specific antibodies in immunoblot assays. Proteins were analyzed as either soluble fractions or total proteins within bacterial cells. Cultures were grown at 37°C or 18°C, incubated for 3 or 6 hours, and induced with IPTG at early log-phase ($OD_{600}$ of 0.1) or mid-log-phase ($OD_{600}$ of 0.4). Note that the samples for the soluble fractions were derived from four times more bacterial cells than those for total protein.

**S2 Fig**. Purification of CTA1-(S14P5)4-DD and CTA1-(S21P2)4-DD, each from 150 mL cultures. Bacterial cell lysates were prepared in PBS containing 0.1% Nonidet P40 and subjected to NiNTA column chromatography. Proteins adsorbed by the column were eluted with either 0.5 M imidazole (A, C) or 0.5 M imidazole containing 0.1% Nonidet P40.
Lane 1: Proteins from 50 µL culture after induction
Lane 2: Proteins from the supernatant of cell lysate from 125 µL culture
Lane 3: Eluate from the NiNTA column derived from 500 µL culture
Lane 4: Dialyzed and concentrated eluate from 2 mL culture.

S1 Table. Percentage of Soluble Protein Recombinants at Different Incubation Temperatures, Durations of Incubation, and Growth Stages of IPTG Induction.

S2 Table. Production of Soluble CTA1-(S14P5)4-DD, as indicated by at Different Cultivation Temperatures, Incubation Times, and Growth Stages of Induction

S3 Table. Production of Soluble CTA1-(S21P2)4-DD, as indicated by at Different Cultivation Temperatures, Incubation Times, and Growth Stages of Induction

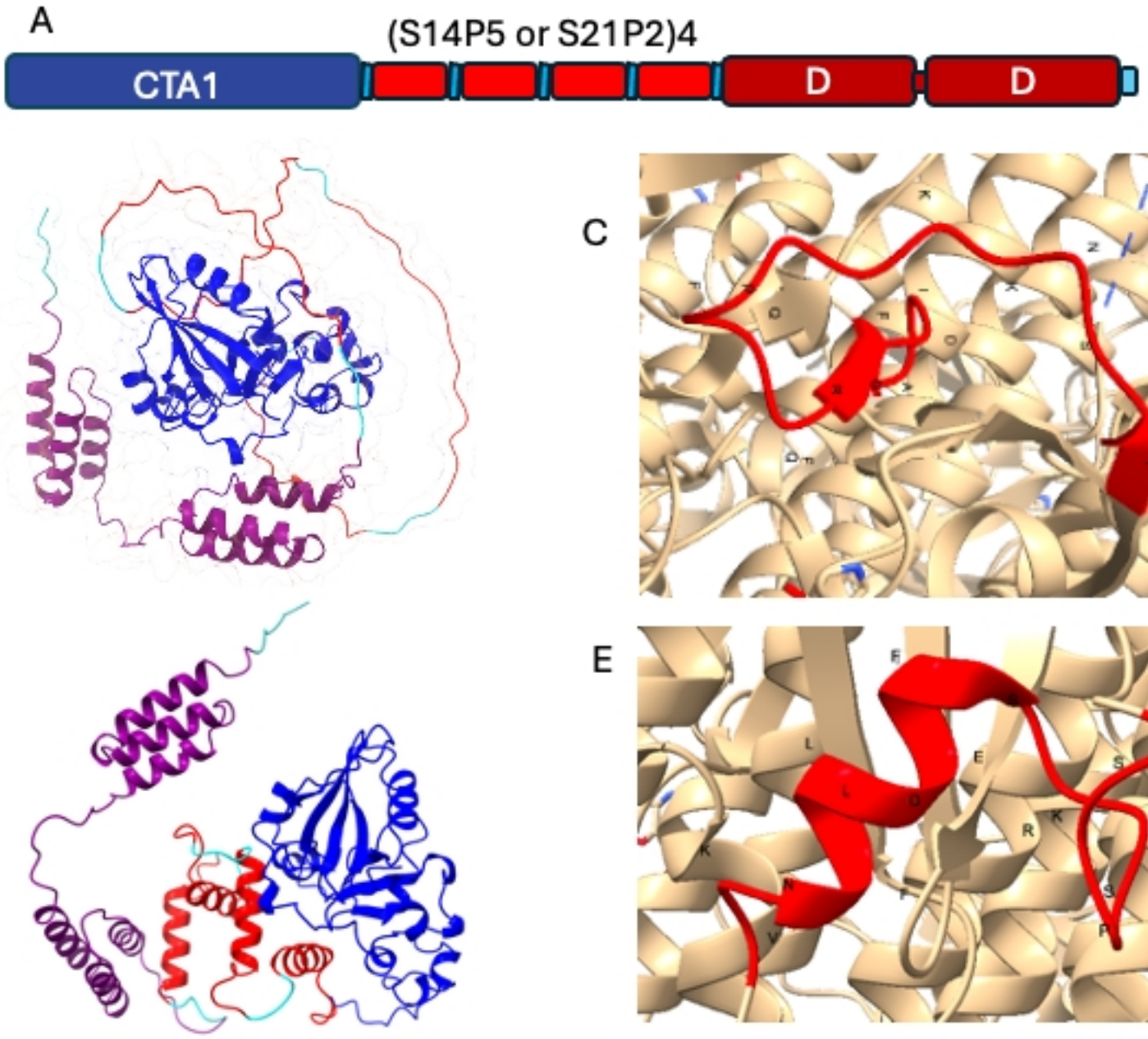


Figure 1

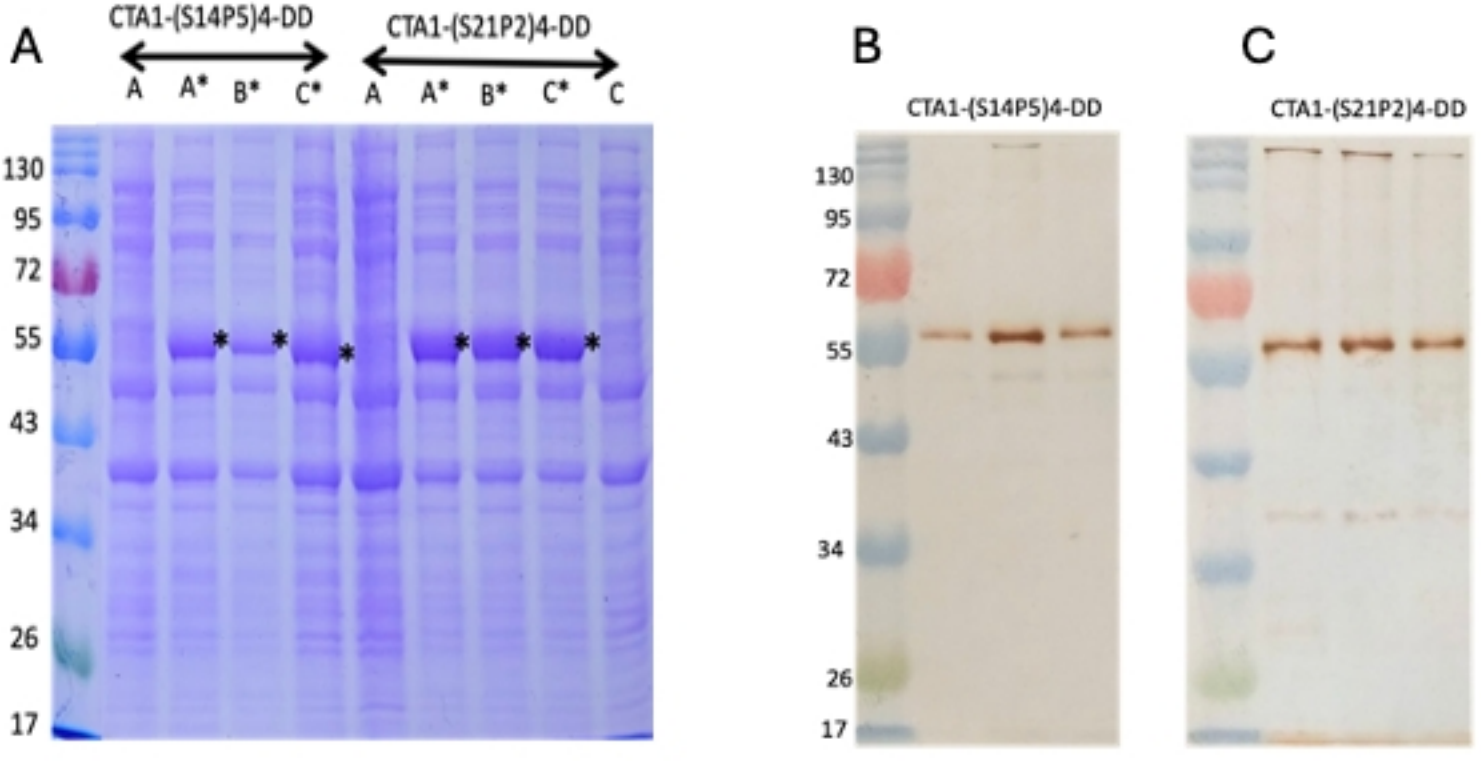
A
CTA1-(S14P5)4-DD
CTA1-(S21P2)4-DD
A
A*
B*
C*
A
A*
B*
C*
C
130
95
72
55
43
34
26
17
B
CTA1-(S14P5)4-DD
C
CTA1-(S21P2)4-DD

Figure 2

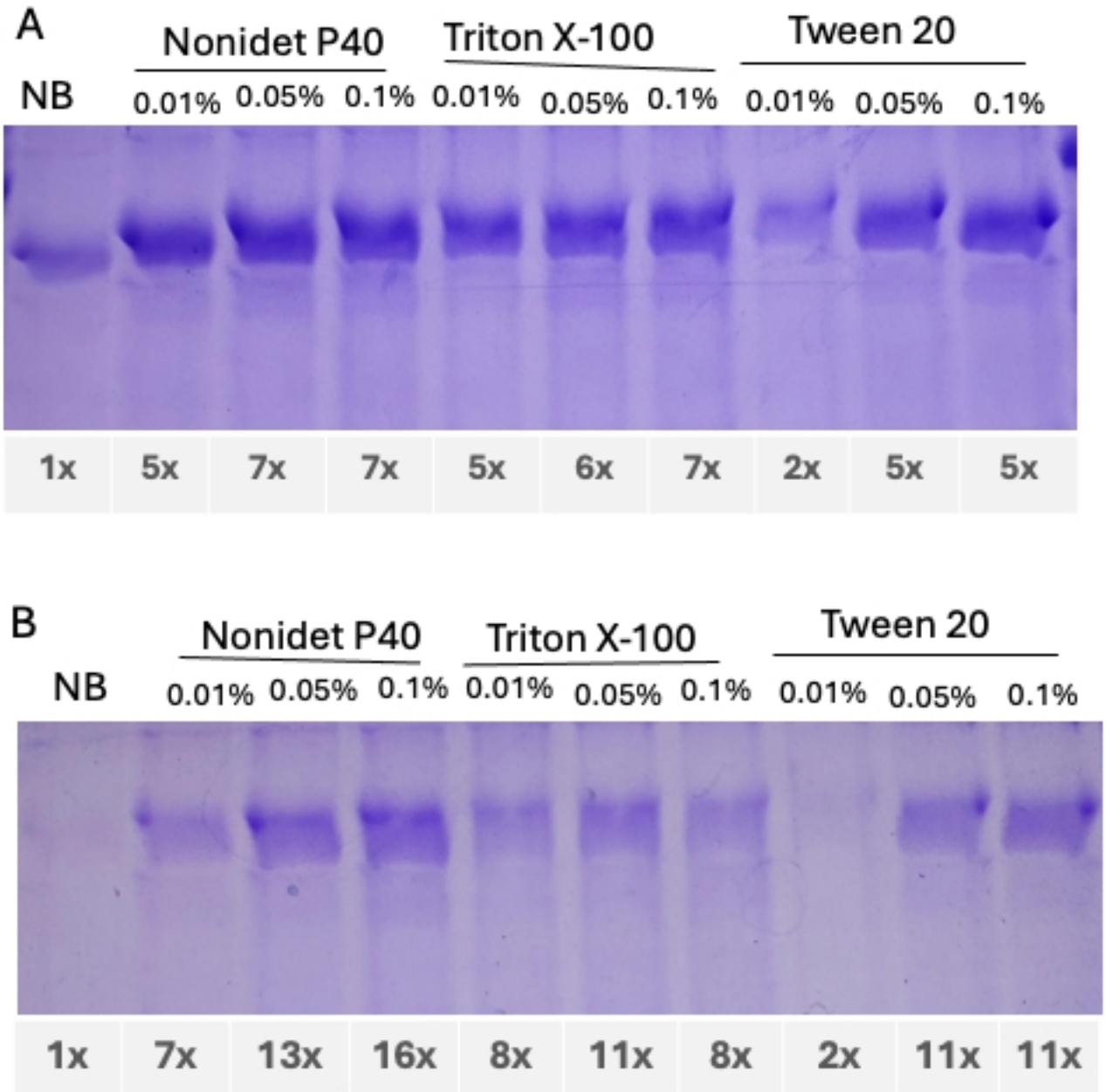
A
Nonidet P40
Triton X-100
Tween 20
NB
0.01% 0.05% 0.1% 0.01% 0.05% 0.1% 0.01% 0.05% 0.1%
1x 5x 7x 7x 5x 6x 7x 2x 5x 5x
B
Nonidet P40
Triton X-100
Tween 20
NB
0.01% 0.05% 0.1% 0.01% 0.05% 0.1% 0.01% 0.05% 0.1%
1x 7x 13x 16x 8x 11x 8x 2x 11x 11x

Figure 3