

## Expression and *in silico* Analysis of CIDR $\alpha$ 1 Recombinant Protein from *Plasmodium Falciparum* as a Malaria Subunit Vaccine Candidate

Erma Sulistyaningsih\*, Nurul Istinaroh, Rosita Dewi and Hairrudin

Department of Parasitology, Faculty of Medicine, University of Jember, Jember 68121, Indonesia

(\*Corresponding author's e-mail: [sulistyaningsih.fk@unej.ac.id](mailto:sulistyaningsih.fk@unej.ac.id))

Received: 30 December 2021, Revised: 12 February 2022, Accepted: 28 February 2022, Published: 10 November 2022

### Abstract

Malaria vaccination is an essential approach to combat malaria. One major protein studied for vaccine development is *Plasmodium falciparum* erythrocyte membrane protein-1 (PfEMP1). It contains several important domains for malaria pathogenesis. The binding of Cysteine-rich interdomain region  $\alpha$ 1 (CIDR $\alpha$ 1) of PfEMP1 to endothelial protein C receptor (EPCR) is associated with cerebral malaria, while CIDR $\alpha$ 1 binding to CD36 has been correlated with uncomplicated malaria. The vital function of CIDR $\alpha$ 1 of PfEMP1 makes it a potential vaccine candidate to prevent clinical features of malaria. A long journey of vaccine development can be shortened by the advancement of bioinformatics and biotechnology techniques. This study aimed to express the recombinant CIDR $\alpha$ 1 of PfEMP1 and investigate its potency as a malaria subunit vaccine candidate by *in silico* analysis. Constructed CIDR $\alpha$ 1-PfEMP1 was expressed in *E. coli* BL21(DE3) after induction with Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and purified using Ni-NTA column. *In silico* analysis on CIDR $\alpha$ 1 of PfEMP1 sequence was conducted using ProtParam Tool for its physicochemical properties, Iterative Threading ASSEMBLY Refinement (I-TASSER) server and JPred4 program to predict secondary structure, 3D modelling, and ligand-binding site, BepiPred 2.0 and Kolaskar-Tangaonkar to predict B-cell epitope, NetCTL server to determine T-cell epitope, and Vaxijen v2.0 server to predict its antigenicity. The chimeric CIDR $\alpha$ 1 of PfEMP1 protein had a 27 kDa molecular weight and was classified as a stable protein. The secondary structure consisted of 6 helices connected with loops. It revealed similarity to CD36-binding protein, EPCR-binding domain, and protein involved in rosetting. The 3D structure modelling demonstrated conserved ligand-binding sites and accessible surface area, which are vital for receptor binding. It had B-cell and T-cell epitopes and was non-allergenic. The properties of the chimeric CIDR $\alpha$  of PfEMP1 indicated its potential as a malaria subunit vaccine candidate.

**Keywords:** CIDR $\alpha$ , *In silico*, Malaria vaccine, *Plasmodium falciparum*, Recombinant protein

### Introduction

Malaria is one global health problem, with an estimated 229 million cases and 409,000 deaths in 2019 in 87 malaria-endemic countries. There was slow progress in decreasing global malaria cases from 2015 to 2019 by less than 2 % and steadily reducing malaria deaths. Several approaches have been conducted to control malaria, including prevention programs by insecticide-treated mosquito nets (ITNs), indoor residual spraying (IRS), malaria chemoprevention, early diagnosis by rapid diagnostic test (RDT), and prompt treatment using artemisinin-based combination therapy (ACT) [1]. Malaria vaccine could be an additional approach to a malaria control program. However, the development of malaria vaccines faces obstacles due to the complexity of malaria infectious agents.

Malaria due to *Plasmodium falciparum* potentially results in severe clinical manifestations. It is related to cytoadherence that occurred through specific interactions between host receptors and PfEMP1, an antigenic protein expressed on the surface of infected erythrocytes (IEs). This process causes host microvasculature obstruction leading to complications of severe malaria. In addition, PfEMP1 mediates rosettes formation shown by the attachment of IEs to other uninfected erythrocytes, further obstructing the host microvasculature [2,3].

PfEMP1 is a key protein in severe malaria pathogenesis. The protein is encoded by a *var* gene family comprising of 60 variable genes. The diversity of PfEMP1 members was determined by the number and size of the extracellular domain, Duffy-binding-like (DBL) and CIDR. Based on sequence similarity, DBL domains are divided into 6 main groups, i.e.,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\xi$ , while CIDR domains are divided into 4 main groups, i.e.,  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ [4]. PfEMP1 universally has a semi-conserved head structure adjacent to the N-

terminus consisting of a tandem DBL $\alpha$ 1-CIDR $\alpha$ 1 domain. It can be followed by the second DBL $\delta$ -CIDR tandem domain or other types of DBL domains [3,5].

The N-terminal part of CIDR $\alpha$ 1-PfEMP1 has the capability to bind EPCR on endothelial cells; cluster of differentiation 36 (CD36) receptors on host cells, including endothelial cells, mononuclear phagocytes, and platelets; and other receptors with unknown binding capabilities, probably related to rosetting [2,3,5,6]. EPCR-binding CIDR $\alpha$ 1 domains are associated with cerebral malaria in Indian adults and African children, contributing to the increase of malaria-related death [3,7,8]. Binding to EPCR inhibits activity for its ligand-activated protein C, resulting in vascular inflammation, increased endothelial permeability, and coagulation, leading to cerebral malaria [3,6,9]. CIDR $\alpha$ 1 from a child with cerebral malaria inhibits the barrier protective function of EPCR in brain endothelial *in vitro* related to brain swelling pathogenesis [10]. In comparison, CD36 binding CIDR $\alpha$ 1 domains has been related to uncomplicated malaria [3]. Meanwhile, almost all *P. falciparum* are isolated directly from patients bound to the human CD36 receptor, which belongs to the CIDR $\alpha$ 1 domains group [5].

PfEMP-1 is a primary target of naturally acquired immunity to malaria [6]. The specific antibodies against PfEMP1 potentially have a clinical protection effect. These antibodies can disrupt the adhesion of IEs to host receptors and mediate IEs clearance through opsonization and phagocytosis or antibody-mediated cytotoxicity [3,5]. Antibodies against EPCR binding CIDR $\alpha$ 1 are elicited more rapidly than those against other CIDR domains in high malaria transmission areas and boosted by severe malaria [11]. These antibodies are acquired in children in intense and seasonal malaria transmission areas and could protect children against severe malaria [3,6]. Meanwhile, the antibody against CD36 binding CIDR $\alpha$ 1 domains probably decreases the risk of severe malaria similar to the level of antibodies against EPCR binding CIDR domains [5].

Studies found that all CIDR $\alpha$ 1 domains adopt a similar fold in order to bind EPCR and the binding mechanisms mimic that of activated protein C to EPCR. However, immune selection pressure has forced sequence diversity across surface-exposed amino acids on the CIDR $\alpha$ 1 domain. As a result, CIDR $\alpha$ 1 domains have varied in sequence to escape immune recognition but have retained their overall structure for high-affinity binding to EPCR [7]. The present study focused on the CIDR $\alpha$ 1-PfEMP1 domain, an adhesion molecule involved in pathogenic mechanism and a potential vaccine candidate. This study expressed the recombinant CIDR $\alpha$ 1-PfEMP1 from a clinically severe malaria patient and utilized a bioinformatics tool to evaluate the capacity of candidate protein by predicting the physicochemical characteristics, secondary structure, B-cell and T-cell epitopes, and ligand-binding site. The computational approaches could speed up the research and simplify the evaluation process to a great extent.

## Materials and methods

### Ethical statement

The study sampling protocol was approved by the Ethical Committee for Research, Faculty of Medicine, University of Jember, Jember, Indonesia, with reference number 1.114/H25.1.11/KE/2017. All procedures complied with the ethical standards of the institutional and national committees on human studies and with the Declaration of Helsinki. Respondent received clear informed consent and signed the consent form.

### Malaria sample

A blood sample was collected from a malaria patient in Jember district, East Java province, Indonesia. It was an imported case, since the patient came from Papua province, a high endemic malaria area in the eastern part of Indonesia. The patient had severe anemia as a complication of severe malaria.

### Expression, purification of CIDR $\alpha$ 1-PfEMP1 recombinant and sequence alignment

The CIDR $\alpha$ 1-PfEMP1 recombinant construct was obtained by cloning PCR product of DNA from the blood of malaria patient, which was amplified using specific primers CIDR $\alpha$ 1\_Fw 5'-CGGGATCCAAATGGAAATGTTATTATG-3' and CIDR $\alpha$ 1\_Rev 5'-CCCTCGAGTTGTAGTAATTTATCAATT-3. The recombinant plasmid was extracted, sequenced, analyzed using Sequence Scanner Software 2, checked for similarity and pairwise comparison using Basic Local Alignment Search Tool (BLAST) and translated into amino acid using ExPASy Translate Tool [12].

The CIDR $\alpha$ 1-PfEMP1 recombinant was expressed in *E. coli* BL21 (DE3). The *E. coli* BL21 (DE3) transformant was grown in 250 mL LB broth containing 50  $\mu$ g/ml kanamycin in an incubator shaker at 150 rpm at 37 °C until OD 600 reached 0.6 - 0.8. The culture was induced with 0.3 mM IPTG concentration for 8 h, followed by centrifugation at 4 °C of 6.000 rpm for 15 min. The pellet was solubilized using extraction

buffer, then lysed with 1 mg/mL lysozyme in ice for 15 min before sonication, and centrifuged at 12,000 rpm at 4 °C for 20 min. The supernatant containing soluble protein and marker was heated for 5 min at 95 °C. Then sample and marker were run in 15 % Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel and run on the 80 V.

The recombinant protein was purified using the Ni-NTA column due to the presence of histidine sequences in the upstream and downstream protein. The supernatant containing soluble protein from harvested *E. coli* BL21 (DE3) recombinant was applied to the Ni-NTA column and washed using a buffer containing 20 mM and 50 mM imidazole. The soluble protein was separated from the column in response to the buffer containing 80 mM imidazole, placed in a suspension buffer, and dialyzed in response to PBS buffer. The purified protein was run in 15 % SDS-PAGE.

### **General physicochemical properties**

The various physical and chemical parameters of protein, including the molecular weight, theoretical pI (isoelectric point), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were computed using ExPASy ProtParam Tool [12].

### **Secondary structure prediction and protein solubility prediction**

The secondary protein structure and its structure based-function were predicted using I-TASSER [13-15] and JPred4 [16]. The confident prediction is presented as Confident score (Conf. Score, range of 0 - 9). A higher score means a more confident secondary structure prediction. Normalized B-factor displays the residue stability in the protein structure, the negative value (below 0) indicates the stable residue in the structure. The prediction of the mean residue accessible surface area (ASA) of the CIDR $\alpha$ 1-PfEMP1 recombinant protein was performed using the NetSurfP-2.0 server. The algorithm used a threshold of 25 % [17].

### **Protein modelling and model evaluation**

The 3D protein modelling was performed using an online I-TASSER server [13-15,18]. It used the SPICKER program to cluster structural conformations based on the pairwise structure similarity, and the protein model was quantitatively calculated by C-score, which was measured based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. The C-score ranges from -5 to 2, with a cut-off of -1.5. A C-score more than -1.5 indicates a model of correct global topology.

### **Ligand-binding site prediction**

The ligand-binding site was predicted by COFACTOR and COACH based on the I-TASSER structure prediction. COFACTOR deduces ligand-binding sites using structure comparison and protein-protein networks [13-15,19] and COACH combines multiple function annotation results on ligand-binding sites from the COFACTOR, TM-SITE and S-SITE programs [13-15,19].

### **B-cell epitope prediction**

B-cell epitope prediction of CIDR $\alpha$ 1-PfEMP1 recombinant protein was conducted using Immune Epitope Database and Analysis Resource by 2 algorithms, i.e., BepiPred 2.0 [20] and Kolaskar-Tangaonkar [21]. BepiPred 2.0 predicts the location of linear B-cell epitopes using a combination of the hidden Markov model and propensity scale method. The score ranges from 0 to 1, with a cut-off of 0.5. A score of more than 0.5 was considered an epitope sequence. Kolaskar-Tangaonkar predicts antigenic determinants based on physicochemical properties of amino acid residues.

### **T-cell epitope prediction**

T-cell epitope prediction of CIDR $\alpha$ 1-PfEMP1 recombinant protein was performed using the NetCTL online server [22], which predicts epitopes of human cytotoxic T lymphocyte (CTL) by combining the Major histocompatibility complex (MHC) Class I affinity, transporter associated with antigen processing (TAP) transport efficiency and proteasomal cleavage for 12 MHC Class I supertypes. The threshold for C terminal cleavage, TAP transport efficiency, and epitope prediction were 0.15, 0.05 and 0.75, respectively. The NetCTL database predicts a maximum score of 1. A score of greater than 0.5 was considered as an epitope [23].

### **Antigenicity and allergenicity prediction**

The antigenicity prediction of recombinant protein to induce immune response was done using Vaxijen v2.0 server [24], which is based on the physical and chemical of protein. The threshold value of

0.5 was used to differentiate between antigenic and non-antigenic proteins. Allergenicity analysis of recombinant protein was conducted using the online tool AllerTOP v2.0.

## Results and discussion

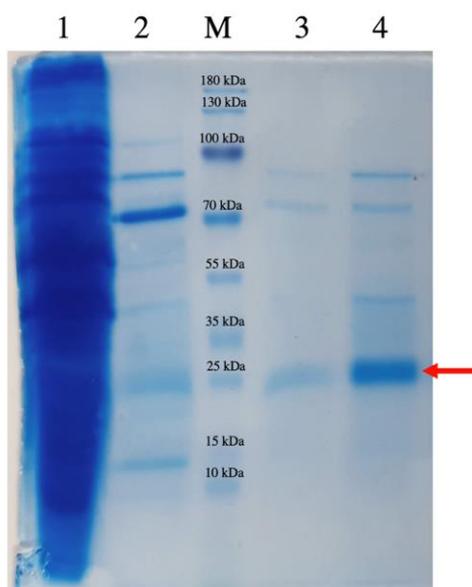
### Expression, purification and sequence alignment of CIDR $\alpha$ 1-PfEMP1 recombinant

In this study, the CIDR $\alpha$ 1 of PfEMP1 from Indonesian isolate was constructed and expressed in *E. coli* BL21(DE3) strain. The constructed CIDR $\alpha$ 1 of PfEMP1 recombinant consisted of 699 nucleotides and was translated into 233 amino acids. The expressed recombinant protein in *E. coli* BL21(DE3) was 27 kDa, as calculated by the ExPASy ProtParam tool (**Table 1**), which measured the protein from the start codon ATG (Met) until Histidine Tag as posed by the vector. The chimeric protein was confirmed by a purification process conducted based on Histidine Tag and classified as a stable protein, as shown in **Figure 1**. BLAST analysis confirmed the sequence as PfEMP1 protein. Pairwise comparison with a reference sequence that has a CD36 binding function, i.e., MC179 showed 58.67 % amino acid identity (**Figure 2**).

**Table 1** The physicochemical properties of the CIDR $\alpha$ 1-PfEMP1 recombinant protein.

Number of amino acids	233 amino acids		
Molecular weight	27066.45 Dalton		
Theoretical pI	6.37		
Amino acid composition	Ala (A)	9	3.9 %
	Arg (R)	4	1.7 %
	Asn (N)	11	4.7 %
	Asp (D)	24	10.3 %
	Cys (C)	7	3.0 %
	Gln (Q)	11	4.7 %
	Glu (E)	15	6.4 %
	Gly (G)	16	6.9 %
	His (H)	17	7.3 %
	Ile (I)	12	5.2 %
	Leu (L)	15	6.4 %
	Lys (K)	30	12.9 %
	Met (M)	8	3.4 %
	Phe (F)	7	3.0 %
	Pro (P)	3	1.3 %
	Ser (S)	11	4.7 %
	Thr (T)	12	5.2 %
	Trp (W)	7	3.0 %
	Tyr (Y)	5	2.1 %
	Val (V)	9	3.9 %
Pyl (O)	0	0.0 %	
Sec (U)	0	0.0 %	
Total number of negatively charged residues (Asp + Glu)	39 amino acids		
Total number of positively charged residues (Arg + Lys)	34 amino acids		
Atomic composition	Carbon	C	1,184
	Hydrogen	H	1,824
	Nitrogen	N	338
	Oxygen	O	362
	Sulfur	S	15
Formula	C <sub>1184</sub> H <sub>1824</sub> N <sub>338</sub> O <sub>362</sub> S <sub>15</sub>		
Total number of atoms	3723		
Extinction coefficients			
Extinction coefficients are in units of M <sup>-1</sup> cm <sup>-1</sup> , at 280 nm measured in water.			
Ext. coefficient	46325		
Abs 0.1 % (= 1 g/L)	1.712, assuming all pairs of Cys residues form cystines		
Ext. coefficient	45950		

Abs 0.1 % (= 1 g/L)	1.698, assuming all Cys residues are reduced
Estimated half-life	M (Met)
The N-terminal of the sequence	30 h (mammalian reticulocytes, <i>in vitro</i> ).
The estimated half-life	>20 h (yeast, <i>in vivo</i> ).
	>10 h ( <i>Escherichia coli</i> , <i>in vivo</i> )
Instability index (II)	28.61
	This classifies the protein as stable
Aliphatic index/ Grand average of hydropathicity (GRAVY)	60.26/-0.973



**Figure 1** Expression and purification of the CIDR $\alpha$ 1-PfEMP1 recombinant protein in 15 % SDS-PAGE. The target protein was 27 kDa (red arrow). M: Protein marker; 1: Soluble recombinant protein in washed buffer containing 20 mM imidazole; 2: Soluble recombinant protein in washed buffer containing 50 mM imidazole; 3: Purified recombinant protein in the 1<sup>st</sup> elution buffer containing 80 mM imidazole; 4: Purified recombinant protein in the 2<sup>nd</sup> elution buffer containing 80 mM imidazole.

Query	36	YNKFFWDVVDMLIDSIQWRDEHGKCI	NKDN	NTCISGCKKKDCFLKWVEKKTEWDKI	95
Sbjct	1	YN FFW VVDMLIDSII+WRDEHG+CINKD	G TCI	GC KKC CF KWVE+KKTEW KI	60
Query	96	KEHFKQKDIGQKGLGSMVTPDFVLQ	QVLEK	NLLQIIQDAYGDAKETEHIRKMLDE--	153
Sbjct	61	K+HF+KQKDI + T D LQ +L K+LLL+IIQD	YGDA	E + I +L++ KDHFRKQKDIPK-----WTHDDFLQTL	114
Query	154	----EETAVAGV----LGGENKTTIDKLLQ			175
Sbjct	115	+ A+AG+ E TTIDKLLQ			144

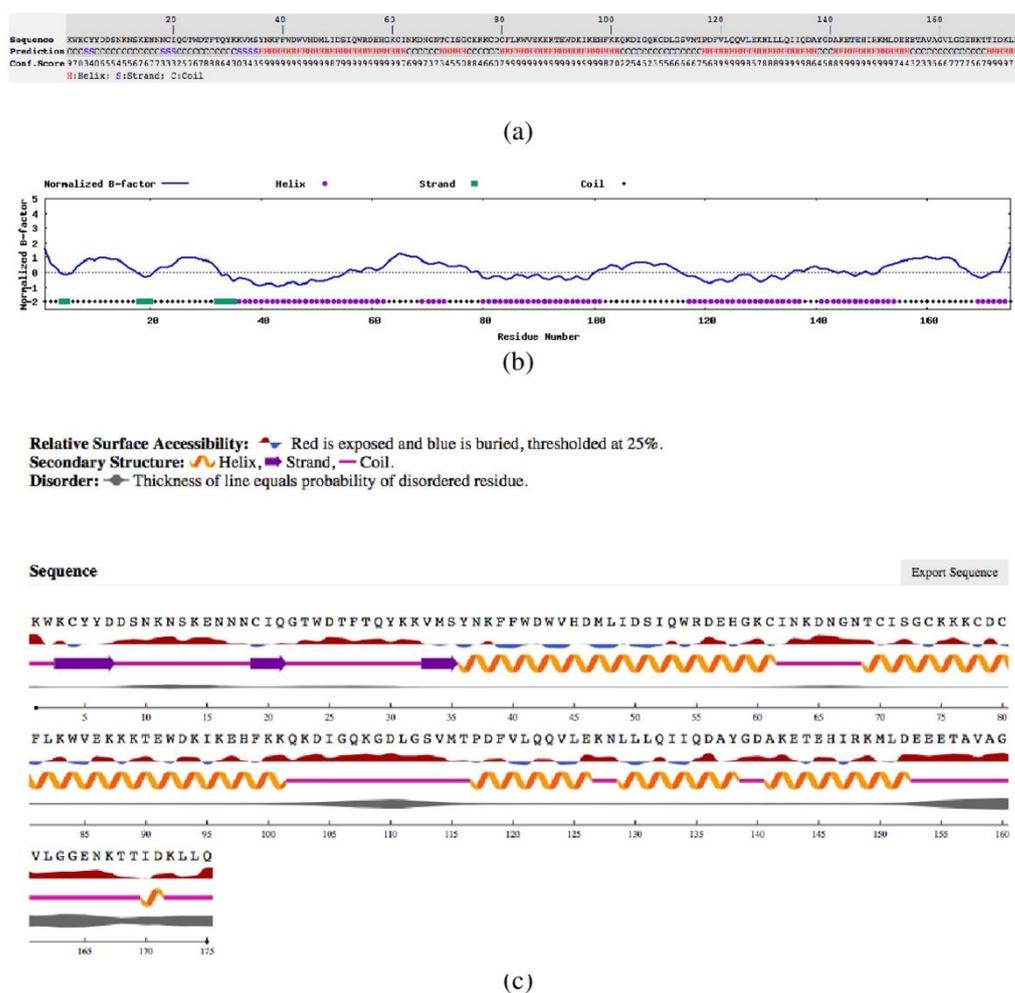
**Figure 2** Pairwise comparison of the CIDR $\alpha$ 1-PfEMP1 with the MC179.

### Physicochemical sequence characteristics

The physicochemical properties of the CIDR $\alpha$ 1 of PfEMP1 recombinant protein were predicted. Parameters such as molecular weight, theoretical pI, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) that indicate the proteins' solubility were summarized in **Table 1**.

BLAST analysis has validated the chimeric protein as a part of PfEMP1 protein. The secondary structure prediction by I-TASSER demonstrated that the CIDR $\alpha$ -PfEMP1 had 6 helices connected with loops. The CIDR $\alpha$ 1-PfEMP1 secondary structure was similar to CIDR1 of MC179 from Malayan Camp variant 2 (MCvar2 PfEMP1) and MC var1 PfEMP1 that bound to CD36, as confirmed by the JPre4 server. These proteins are composed of a bundle of 3 helices connected by loops and 3 additional helices in the order of H1-H2-a-b-c-H3 in the V-shape [5]. The CIDR $\alpha$ 1-PfEMP1 amino acid sequences also showed 58.67 % identity with the CIDR1 of MC179 from Malayan Camp variant 2 (MCvar2 PfEMP1), which is needed for CD36 binding. It is reported that CIDR1 of MC179 corresponds to the M2 subdomain. It is a minimal domain needed for CD36 binding. Meanwhile, the CIDR $\alpha$ 1-PfEMP1 corresponds to M1 and M2 subdomains that have 7 cysteine residues [5,25].

Further calculation using B-factor in I-TASSER server that indicates the extent of inherent thermal mobility of residues/atoms in protein showed the residue is relatively stable in the protein structure, especially at the helix sites (**Figure 3(b)**). The exposed and surface accessibility of the residues made the benefit of protein for ligand-binding.



**Figure 3** (a) Prediction of the chimeric CIDR $\alpha$ 1-PfEMP1 secondary structure by I-TASSER, it consisted of 6 helices connected with loops; (b) Prediction of residues stability in protein by I-TASSER, presented as B-factor, negative values indicated relative residues stability in the protein structure; (c) Prediction of surface accessibility of the secondary structure by NetSurfP-2.0 server.

#### Protein predicted secondary structure and protein solubility

The CIDR $\alpha$ 1-PfEMP1 secondary structure was predicted by I-TASSER and the NetSurfP-2.0 server. The protein consisted of 6 helices connected with loops (**Figure 3(a)**) and had relatively stable residues in protein structure (**Figure 3(b)**). The prediction of the mean residue accessible surface area (ASA) of the

CIDR $\alpha$ 1-PfEMP1 recombinant protein using the NetSurfP-2.0 server showed that the chimeric protein is exposed and accessible (**Figure 3(c)**).

The secondary structure similarity indicated the similar function of the chimeric CIDR $\alpha$ 1 of PfEMP1 protein as the CD36 binding protein (Protein Database (PDB) identity: 3c64, 5lgd) (**Table 2**). Further secondary structure analysis also showed chimeric CIDR $\alpha$ 1 of PfEMP1 similar with the CIDR $\alpha$  from IT4var07 PfEMP1 and CIDR $\alpha$  from HB3var03 PfEMP1 that bound to EPCR. The binding of CIDR $\alpha$  to EPCR is reported to have implications in severe malaria in child patients. Another interesting finding is that the chimeric CIDR $\alpha$ 1 of PfEMP1 is similar to the CIDR $\gamma$  of PfEMP1-varO strain (PDB identity: 2yk0), which is essential for erythrocyte binding or rosetting, lead to severe malaria outcomes. Thus, it emphasized the potential role of chimeric CIDR $\alpha$ 1 of PfEMP1 as an adhesion molecule that affect severe malaria clinical symptom. However, studies showed that CIDR $\alpha$ 1-PfEMP1 domains have evolved to diversify due to the immune pressure and have retained the capacity to bind with specific endothelial ligands. The binding sites of CIDR $\alpha$ 1 to EPCR and CD36 are very different in shape, but similar chemical properties mediate both interactions, residues involved in stabilizing the fold of binding are conserved and sequence variation in residues that interact directly with the ligands is conservative in chemistry, maintaining the capacity to bind. This increases the possibility of generating antibodies that recognize these conserved chemical features and show extensive inhibitory potential [26].

Analysis using JPred4 server showed that CIDR $\alpha$ 1 of Indonesian *Plasmodium falciparum* is similar to those of PfEMP1, CIDRA, erythrocyte membrane protein 1-Chain A that recognized in PDB. The results are in accordance with I-TASSER analysis on the best threading templates, which works based on the templates identified by LOMETS from the PDB library. All similar proteins had a Z-score > 1, which indicated good alignment (**Table 2**).

**Table 2** Secondary structure similarity of CIDR $\alpha$ 1 of Indonesian *Plasmodium falciparum* using JPred4 and I-TASSER server.

JPRred4 server							
PDB identity	Protein description			Chain type	Blast E-value		
3c64	PfEMP1 variant 2 of strain MC			A	2e-45		
5lgd	PfEMP1 variant 1 of strain MC			B	3e-44		
4v3e	IT4VAR07 CIDRA			A	5e-17		
4v3d	HB3VAR03 CIDRA DOMAIN			C	9e-17		
4v3d	HB3VAR03 CIDRA DOMAIN			A	9e-17		
2yk0	Erythrocyte membrane protein 1			A	7e-06		
I-TASSER server							
PDB identity	Protein identity			Identity 1	Identity 2	Cov*	Normalized Z-score**
2yk0A	Membrane protein	<i>Plasmodium falciparum</i> varO strain (CIDR $\gamma$ )		0.21	0.26	0.9	1.97
4v3dA	Signaling protein	<i>Plasmodium falciparum</i> HB3-CIDR $\alpha$		0.32	0.35	0.88	2.09
5lgdB	Cell adhesion	<i>Plasmodium falciparum</i> CIDR $\alpha$ domain from MCvar1		0.65	0.52	0.75	3.35

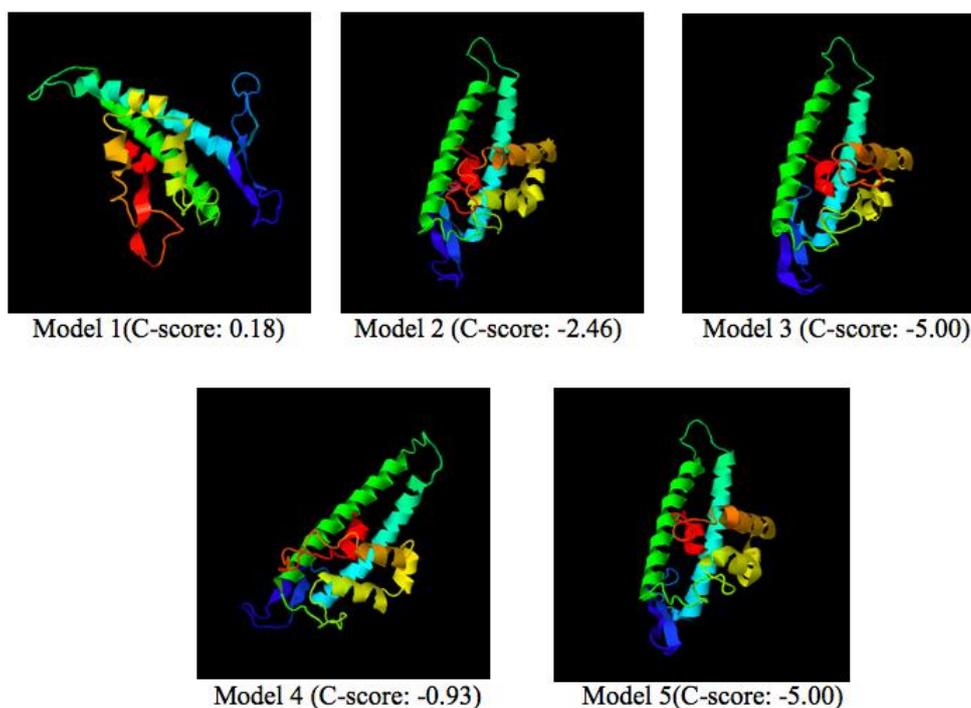
\*Cov; Coverage, it represents the coverage of the threading alignment, it is equal to number of aligned residues divided by the length of query protein.

\*\* Normalized Z-score: the highest significant alignments from the protein database (PDB) library. The Z-score > 1 indicated a good alignment, and the higher Z-score, the better alignment.

### Protein modelling methods and model evaluation

The 3D structure modelling and the ligand-binding site regions enhanced the prediction confidence of the chimeric CIDR $\alpha$ 1-PfEMP1 potency. **Figure 4** presents 5 protein modellings predicted by I-TASSER. The highest C-score was displayed by model 1 (C-score = 0.18) [18]. The 3D structure modelling was designed based on sequence-based secondary structure, coverage alignment, and solvent accessibility

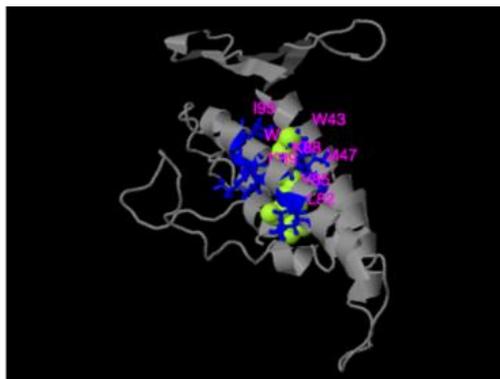
predictions. The chimeric protein possessed conserved residues in the ligand-binding site. It also showed accessible surface area, which is required for the binding. Previous studies on molecular features of CIDR $\alpha$  domains to bind to CD36 and EPCR showed prominent similarities, but with structural inversion of the binding mechanism. In both types, CIDR $\alpha$  domains consist of a core 3  $\alpha$ -helices bundle with an insertion between the 2<sup>nd</sup> and 3<sup>rd</sup> core helices,  $\alpha$ 2 and  $\alpha$ 6, to make smaller helices that lie approximately straight to the core helices [26]. The insert has been identified as a homology block, and the fold of the homology blocks contribute to the formation of ligand-binding sites [26,27].



**Figure 4** The CIDR $\alpha$ 1-PfEMP1 3D structure modelling by I-TASSER server. There were 5 final models ranked based on Confidence score (C-score) in the range of [-5, 2]. Model 1 had the highest C-score (0.18), meaning the highest confidence of the 3D structure model.

#### Ligand-binding site regions

**Figure 5** shows a prediction of the ligand-binding site by COFACTOR and COACH based on structure prediction demonstrated cooperation of conserved residues in ligand-binding site (W43, M47, L82, V85, K88, K89, W92, I95) with the highest C-score of 0.11 (range 0 - 1), with the cluster size 5, PDB hit 3zpqB and the ligand name of 2CV. The C-score is the confidence score of prediction, where the higher score indicates a more reliable prediction. The cluster size 5 demonstrated the total number of templates in a cluster [14,15,19]. The 3zpqB in PDB is a recombinant protein of beta adrenergic receptor that is classified as a membrane protein [28]. The ligand 2CV is HEGA-10 and has been identified as N-2(hydroxyethyl)-N-[(2R,3R,4S, 5S-2,3,4,5,6-pentahydroxyethyl]decamide [29]. Ligand-binding site prediction using COFACTOR deduces protein function using structure comparison and protein-protein networks, while COACH algorithm generates ligand-binding site prediction by combining multiple function annotation results from COFACTOR and structural similarity and matching the target protein with protein in the BioLiP database [15]. However, a comparison with CD36-binding CIDR1 of MC179 showed that the ligand-binding site of the recombinant protein is located at the H1 and H2 [5].



**Figure 5** The ligand-binding site of the CIDR $\alpha$ 1-PfEMP1 recombinant protein. The green-yellow sphere is predicted binding ligand; Blue ball and sticks are binding residues.

### B-cell epitope prediction

The B-cell epitope is a region of antigens that are recognized by B-cell receptors or specific antibodies. These epitopes recognized by B-cells may constitute any exposed solvent region in the antigen and can be of different chemical nature, so epitopes are exposed on the surface protein and accessible to immunoglobulin [30]. Linear B-cell epitopes prediction using the BepiPred identified 5 epitopes, while semi empirically prediction by the Kolaskar Tangaonkar algorithm determined 4 epitopes (**Table 3**). Most of the predicted epitopes were spanning surface-expose regions. Comparison with the secondary structure of CD36-binding CIDR $\alpha$  domain showed that the predicted epitopes are localized at the N-terminal region, H1, H2, and loop a of CIDR $\alpha$  domain and those amino acid sequences included the minimum regions for CD36 binding [5].

**Table 3** The rank of predicted linear B-cell epitopes predicted by BepiPred server and Kolaskar Tangaonkar algorithm.

Rank	Peptide sequence	Peptide length (Amino acids)	Amino acid position	Score	Area of sequence
<b>BepiPred 2.0: Linear sequence</b>					
1	YDDSNKNSKENNNCIQGTWDTFTQY KKVM	29	6 - 34	0.564	Conserved region
2	CINKDNGNTCISGCK	15	61 - 75	0.535	Conserved region
3	KKQKDIGQKGDLSVMT	17	100 - 116	0.530	Conserved region
4	WVEKKKTEW	9	84 - 92	0.516	Conserved region
5	LLQIIQDAYGDAKETEHIRKMLDEEE TAVAGVLGGENKTTI	41	130 - 170	0.512	Conserved region
<b>Kolaskar Tangaonkar</b>					
1	YKKVMSY	7	30 - 36	1.060	Conserved region
2	HDMLIDS	7	45 - 51	1.027	Conserved region
3	TPDFVLQQVLEKNLLLQIIQD	21	116 - 136	1.022	Conserved region
4	CISGCKKCCDCFLKWV	16	70 - 85	1.001	Conserved region

### T-cell epitope prediction to MHC Class I

T-cell epitope is a short peptide within an antigen that is able to stimulate either CD4+ or CD8+ T-cells. The capacity to stimulate T-cells is known as immunogenicity, and it consists of 3 basic steps; antigen processing, peptide binding to MHC molecules, and recognition by a cognate T-cell receptor (TCR). Of these, MHC-peptide binding is the most selective to determine T-cell epitopes [30]. T-cell epitopes prediction of CIDR $\alpha$ 1-PfEMP1 recombinant protein using the NetCTL determined 2 epitopes (**Table 4**). Each epitope consists of 9 amino acid residues because most of human leukocyte antigens (HLA) have strong preference for binding 9mers. The predicted T-cell epitopes are located at the loop b and loop c between H2 and H3 of CIDR $\alpha$  domain that is needed for CD36 binding [5]. Studies on the *P. falciparum*-specific CD8+ T-cell response in malaria are limited due to low *ex vivo* frequency of circulating peripheral *P. falciparum*-specific T cells [31]; however, this study has predicted 2 probable antigens that could activate T-cell response through MHC class I.

**Table 4** Epitope prediction of T-cell to MHC Class I using NetCTL online and VaxiJen v2.0.

Peptide sequence	Amino acid position	Antigenicity score	Antigenicity prediction
EETAVAGVL	154 - 162	0.5037	Probable antigen
KETEHIRKM	142 - 150	0.7122	Probable antigen

### Antigenicity and allergenicity prediction

Prediction of antigenicity and allergenicity was performed using Vaxijen v2.0 and AllerTOP 2.0. The Vaxijen v2.0 analysis on EETAVAGVL and KETEHIRKM epitopes resulted in a score of  $\geq 0.5$ , indicating that a CIDR $\alpha$ 1-PfEMP1 is a probable antigen (**Table 4**). Furthermore, AllerTOP 2.0 software analysis on EETAVAGVL and KETEHIRKM epitopes showed those potentially non-allergen.

Studies found that antibodies are the principal contributors to blood-stage malaria parasite immunity in the field, and for some extent, it also involved T-cell [8,32]. Therefore, malaria vaccine candidates should have the potency to induce humoral and cellular immunity. The prediction of B-cell epitopes based on linear structure and semi empirically physicochemical properties showed that the chimeric CIDR $\alpha$ 1-PfEMP1 had several epitope sequences. Furthermore, the T-cell epitopes prediction on MHC Class I demonstrated that the recombinant protein is antigenic and non-allergenic.

### Conclusions

This study provided information on the chimeric CIDR $\alpha$ 1-PfEMP1 by bioinformatics tools. The physicochemical properties, the secondary structure prediction, ligand-binding site, and epitopes predictions indicated the potency of chimeric CIDR $\alpha$ 1-PfEMP1 protein as a malaria vaccine antigen. The designed vaccine that utilizes immunoinformatic analysis should be experimentally studied to determine the efficiency of further produced vaccines. Therefore, *in vivo* experimental studies are needed for assessing the various aspects of this chimeric protein to develop a malaria vaccine candidate.

### Acknowledgements

The research was financially supported by a Grant Number: 3313/UN25.3.2/LT/2021 from Indonesian Ministry of Education, Culture, Research, and Technology 2021, Indonesia, through Excellence Basic Research of Higher Education scheme.

### References

- [1] World Health Organization (WHO) 2020. World Malaria Report 2020, Available at: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020>, accessed October 2021.
- [2] JD Smith. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. *Mol. Biochem. Parasitol.* 2014; **195**, p. 82-7.
- [3] N Obeng-Adjei, DB Larremore, L Turner, A Ongoiba, S Li, S Doumbo, TB Yazew, K Kayentao, LH Miller, B Traore, SK Pierce, CO Buckee, T Lavstsen, PD Crompton and TM Tran. Longitudinal analysis of naturally acquired PfEMP1 CIDR domain variant antibodies identifies associations with malaria protection. *JCI Insight* 2020; **5**, PMC7406271.
- [4] JD Smith, G Subramanian, B Gamain and DI ML Baruch. Classification of adhesive domains in the

- Plasmodium falciparum erythrocyte membrane protein 1 family. *Mol. Biochem. Parasitol.* 2000; **110**, 293-310.
- [5] MM Klein, AG Gittis, HP Su, MO Makobongo, JM Moore, S Singh, LH Miller and DN Garboczi. The cysteine-rich interdomain region from the highly variable Plasmodium falciparum erythrocyte membrane protein-1 exhibits a conserved structure. *PLoS Pathog.* 2008; **4**, e1000147.
- [6] L Turner, T Lavstsen, BP Mmbando, CW Wang, PA Magistrado, LS Vestergaard, DS Ishengoma, DTR Minja, JP Lusingu and TG Theander. IgG antibodies to endothelial protein C receptor-binding cysteine-rich interdomain region domains of Plasmodium falciparum erythrocyte membrane protein 1 are acquired early in life in individuals exposed to malaria. *Infect. Immun.* 2015; **83**, 3096-103.
- [7] CKY Lau, L Turner, JS Jespersen, ED Lowe, B Petersen, CW Wang, JEV Petersen, J Lusingu, TG Theander, T Lavstsen and MK Higgins. Structural conservation despite huge sequence diversity allows EPCR binding by the pfemp1 family implicated in severe childhood malaria. *Cell Host Microbe* 2015; **17**, 118-29.
- [8] C Harmsen, L Turner, S Thrane, AF Sander, TG Theander and T Lavstsen. Immunization with virus-like particles conjugated to CIDR $\alpha$ 1 domain of Plasmodium falciparum erythrocyte membrane protein 1 induces inhibitory antibodies. *Malar J. Bio. Med. Central* 2020; **19**, 1-11.
- [9] M Bernabeu, SA Danziger, M Avril, M Vaz, PH Babar, AJ Brazier, T Herricks, JN Maki, L Pereira, A Mascarenhas, E Gomes, L Chery, JD Aitchison, PK Rathod and JD Smith. Severe adult malaria is associated with specific PfEMP1 adhesion types and high parasite biomass. *Proc. Natl. Acad. Sci. U.S.A.* 2016; **113**, E3270-E3279.
- [10] A Kessler, S Dankwa, M Bernabeu, V Harawa, SA Danziger, F Duffy, SD Kampondeni, MJ Potchen, N Dambrauskas, V Vigdorovich, BG Oliver, SE Hochman, WB Mowrey, IJC MacCormick, WL Mandala, SJ Rogerson, DN Sather, JD Aitchison, TE Taylor, KB Seydel, JD Smith and K Kim. Linking EPCR-binding PfEMP1 to brain swelling in pediatric cerebral malaria. *Cell Host Microbe* 2017; **22**, 601-14.
- [11] JS Jespersen, CW Wang, SI Mkumbaye, DT Minja, B Petersen, L Turner, JEV Petersen, JPA Lusingu, TG Theander and T Lavstsen. Plasmodium falciparum var genes expressed in children with severe malaria encode CIDR  $\alpha$ 1 domains. *EMBO Mol. Med.* 2016; **8**, 839-50.
- [12] E Gasteiger, C Hoogland, A Gattiker, S Duvaud, MR Wilkins, RD Appel and A Bairoch. *The proteomics protocols handbook*. Humana Press, New Jersey, United States, 2005, p. 571-608.
- [13] Y Zhang. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* 2008; **9**, 1-8.
- [14] A Roy, A Kucukural and Y Zhang. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* 2010; **5**, 725-38.
- [15] J Yang, R Yan, A Roy, D Xu, J Poisson and Y Zhang. The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* 2014; **12**, 7-8.
- [16] A Drozdetskiy, C Cole, J Procter and GJ Barton. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* 2015; **43**, W389-W394.
- [17] B Petersen, TN Petersen, P Andersen, M Nielsen and C Lundegaard. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* 2009; **9**, 1-10.
- [18] J Yang, Y Wang and Y Zhang. ResQ: An approach to unified estimation of B-factor and residue-specific error in protein structure prediction. *J. Mol. Biol.* 2016; **428**, 693-701.
- [19] C Zhang, PL Freddolino and Y Zhang. COFACTOR: Improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res.* 2017; **45**, W291-W299.
- [20] MC Jespersen, B Peters, M Nielsen and P Marcatili. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 2017; **45**, W24-W29.
- [21] AS Kolaskar and PC Tongaonkar. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 1990; **276**, 172-4.
- [22] MV Larsen, C Lundegaard, K Lamberth, S Buus, O Lund and M Nielsen. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* 2007; **8**, 1-12.
- [23] B Peters, S Bulik, R Tampe, PM van Endert and HG Holzhütter. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.* 2003; **171**, 1741-9.
- [24] IA Doytchinova and DR Flower. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* 2007; **8**, 1-7.
- [25] M Mo, CL Hooi, M Kotaka, M Niang, X Gao, JK Iyer, J Lescar and P Preiser. The C-terminal segment of the cysteine-rich interdomain of Plasmodium falciparum erythrocyte membrane protein 1 determines CD36 binding and elicits antibodies that inhibit adhesion of parasite-infected erythrocytes. *Infect. Immun.* 2008; **76**, 1837-47.

- [26] FL Hsieh, L Turner, JR Bolla, CV Robinson, T Lavstsen and MK Higgins. The structural basis for CD36 binding by the malaria parasite. *Nat. Commun.* 2016; **7**, 1-11.
- [27] TS Rask, DA Hansen, TG Theander, AG Pedersen and T Lavstsen. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes - divide and conquer. *PLoS Comput. Biol.* 2010; **6**, e1000933.
- [28] JA Christopher, J Brown, AS Doré, JC Errey, M Koglin, FH Marshall, DG Myszka, RL Rich, CG. Tate, B Tehan, T Warne and M Congreve. Biophysical fragment screening of the  $\beta$ 1-adrenergic receptor: Identification of high affinity arylpiperazine leads using structure-based drug design. *J. Med. Chem.* 2013; **56**, 3446-55.
- [29] National Library of Medicine. Compound Summary, Available at: <https://pubchem.ncbi.nlm.nih.gov/compound/49835947>, accessed February 2022.
- [30] JL Sanchez-Trincado, M Gomez-Perosanz and PA Reche. Fundamentals and methods for T- and B-cell epitope prediction. *J. Immunol. Res.* 2017; **2017**, 2680160.
- [31] J Heide, KC Vaughan, A Sette, T Jacobs and JS Zur Wiesch. Comprehensive review of human plasmodium falciparum-specific CD8+ T cell epitopes. *Front Immunol.* 2019; **10**, 1-23.
- [32] K Miura. Progress and prospects for blood-stage malaria vaccines. *Expert Rev. Vaccines* 2016; **15**, 765-81.