FULL-LENGTH PAPER

Inhibition of a malaria host–pathogen interaction by a computationally designed inhibitor

Autumn R. Tobin[1] | Rachel Crow[2] | Darya V. Urusova[3] | Jason C. Klima[4,5] |
Niraj H. Tolia[3,6] | Eva-Maria Strauch[1,7]

[1]Department of Pharmaceutical and Biomedical Sciences, University of Georgia, Athens, Georgia, USA

[2]Department of Microbiology, University of Washington, Seattle, Washington, USA

[3]Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri, USA

[4]Institute for Protein Design, University of Washington, Seattle, Washington, USA

[5]Department of Biochemistry, University of Washington, Seattle, Washington, USA

[6]Host-Pathogen Interactions and Structural Vaccinology Section, Laboratory of Malaria, Immunology and Vaccinology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

[7]Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA

Correspondence
Eva-Maria Strauch, Department of Pharmaceutical and Biomedical Sciences, University of Georgia, 240 W Green St, Athens, GA 30606, USA.
Email: estrauch@uga.edu

Abstract

Malaria is a substantial global health burden with 229 million cases in 2019 and 450,000 deaths annually. *Plasmodium vivax* is the most widespread malaria-causing parasite putting 2.5 billion people at risk of infection. *P. vivax* has a dormant liver stage and therefore can exist for long periods undetected. Its blood-stage can cause severe reactions and hospitalization. Few treatment and detection options are available for this pathogen. A unique characteristic of *P. vivax* is that it depends on the Duffy antigen/receptor for chemokines (DARC) on the surface of host red blood cells for invasion. *P. vivax* employs the Duffy binding protein (DBP) to bind to DARC. We first de novo designed a three helical bundle scaffolding database which was screened via protease digestions for stability. Protease-resistant scaffolds highlighted thresholds for stability, which we utilized for selecting DARC mimetics that we subsequentially designed through grafting and redesign of these scaffolds. The optimized design small helical protein disrupts the DBP:DARC interaction. The inhibitor blocks the receptor binding site on DBP and thus forms a strong foundation for a therapeutic that will inhibit reticulocyte infection and prevent the pathogenesis of *P. vivax* malaria.

KEYWORDS
Plasmodium vivax, protease selections, protein design, protein scaffolds, three-helical bundles

Autumn R. Tobin and Rachel Crow contributed equally to this study.

1 | INTRODUCTION

Malaria is predicted to infect up to 2.5 billion by the year 2050. *Plasmodium vivax* has been the second largest cause of human malaria around the world after *Plasmodium falciparum*. However, substantial increases in malaria incidence occurred between 2014 and 2016, with *P. vivax* representing 64% of all malaria cases.[1] Most strategies deployed to eliminate malaria primarily target falciparum malaria and are less effective in controlling vivax malaria.[2] Basic research on *P. vivax* has been hampered by lack of continuous in vitro culture system. Furthermore, *P. vivax* poses unique challenges to control strategies because of its ability to cause relapses[3]; it can remain dormant in the liver for years before entering its blood stage form in which several clinical symptoms can occur. Primaquine is the drug currently used to eliminate hypnozoites from the liver. However, adverse effects associated with primaquine have been reported, particularly in patients who have a severe deficiency of the enzyme glucose-6-phosphate dehydrogenase, in whom the drug triggers hemolysis. Therefore, primaquine cannot be administered to pregnant women or children because of the risk that the patient might have this enzyme deficiency and new treatment options are needed.

During the life cycles of *P. vivax*, different surface proteins are displayed at various stages of the parasite.

During its blood stage cycle, the entry of host cells by the parasite is mediated by the Duffy binding protein (DBP),[4–12] a member of the erythrocyte binding-like invasion protein family.[13–15] DBP binds to the Duffy Antigen Receptor for Chemokines (DARC) on host reticulocytes using a conserved cysteine-rich Duffy binding-like (DBL) domain known as region II (DBP-II).[4–9,11,12] DBP-II binds DARC via receptor-induced ligand dimerization, sandwiching DARC residues 19–30 between two DBP-II molecules.[5,6] Receptor binding and dimerization are required to engage DARC (Figure 1).[5,6] The stepwise binding mechanism improves the affinity and avidity of the DBP-II:DARC interaction leading from a binding constant of 2 μM to 80 nM after dimerization of DBP (Figure 1).[5]

Antibodies from immune individuals with naturally acquired immunity that block the DARC receptor binding and/or dimerization neutralize *P. vivax*.[5,6,16–18] A single human monoclonal antibody that inhibits the DBP:DARC interaction is a potent neutralizer of *P. vivax* invasion of human red cells demonstrating that disruption of the DBP:DARC interaction is sufficient to prevent *P. vivax* invasion of host cells.[18] DBP-II is a promising vaccine candidate for *P. vivax* malaria given the critical role of the DBP:DARC interaction in *P. vivax* host cell invasion.[5–7,19–25] In addition to its vaccine potential, DBP-II presents an ideal target molecule for a protein-based inhibition, and a therapeutic that can block both
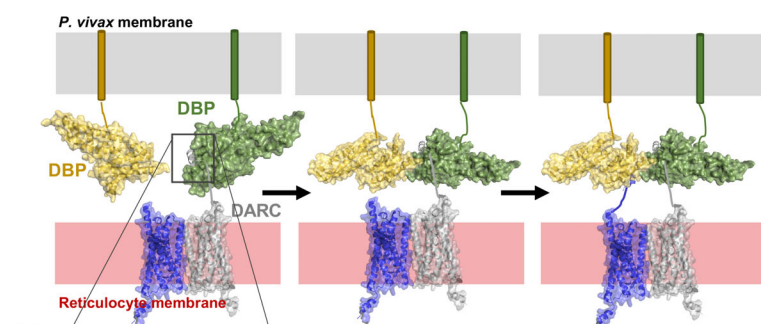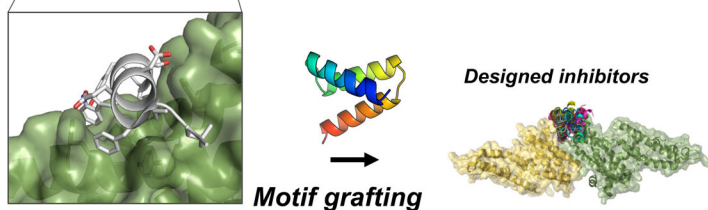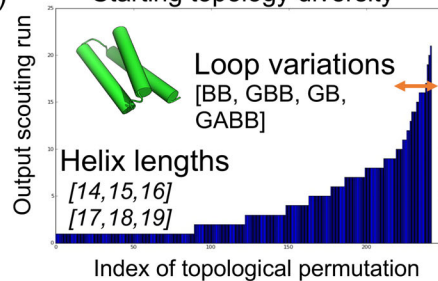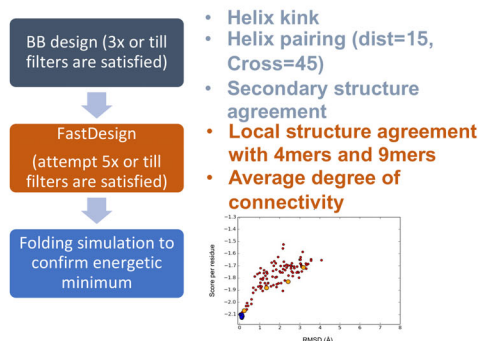
FIGURE 1 (a) Binding mode of Duffy antigen/receptor for chemokines (DARC) with Duffy binding protein (DBP-II) when membrane embedded. DARC is based on the AlphaFold homology model AF-Q16570-F1 (Uniprot). (b) Models of designed receptor mimicries and their interference of DBP binding and dimerization. (c) Histogram illustrating which topology permutation can be successfully made after an initial time limited design trajectory. (d) Overview of the design workflow for the new helical scaffold proteins.

receptor binding as well as interfere with the dimerization is highly desired.

To develop a potent inhibitor, we started by designing a highly stable, helical scaffold library, and used this library to develop a series of receptor mimicries that inhibit the interaction of DBP-II with its host cell receptor. We designed and experimentally evaluated 4,000 helical bundles of the size of 48–68 residues long (Figure S1) and then monitored stability by a direct protease-digestion assay using yeast surface display. Proteins that survived protease-digestion as identified by next-generation sequencing were used as starting material for the design of the receptor mimicries. This stringent selection showed clear cut-offs for scoring terms of these helices (Figure S2) and resulted in a set of stable scaffold proteins. The results highlight stability thresholds for these de novo designed helical bundles, identifying selection criteria for the future design of different topologies as well as the redesign of the original scaffolds produced here. Proteins were redesigned by grafting the core interactions of DARC with DBP-II on the initial designs. Our lead candidate binds with single digit nanomolar affinity to the receptor binding site of DBP. Cell-based assays demonstrate the lead candidate potently inhibits the binding of DBP to DARC on red blood cells with an IC50 of 72.5 nM, and these results form a strong foundation for the development of a potent neutralizing therapeutic reagent.

2 | RESULTS

A protein-based inhibitor that has to serve both as a potential therapeutic as well as diagnostic for treating and diagnosing a *P. vivax* malaria infection would have to be suitably stable to avoid degradation within the body, and possesses characteristics that would enable long-term storage at temperatures suitable for deployment in the field. Furthermore, high expression of protein scaffolds are also desired to ensure affordable production costs. Finally, a designed protein inhibitor must bind to DBP and bury a large enough surface area to mimic the receptor interaction while also sterically blocking the dimer interface of DBP-II.

We therefore first designed a "scaffold" library of all helical three helical bundles with varying sizes and defined loop angles, but aiming to design larger proteins than the previously reported mini-proteins that were 42 residues long and less.[26] Our rational was that the core of the proteins needs to be bigger than three to four amino acids (as with the small mini-proteins) as the helical binding interface with DBP-II is relatively large and contained several hydrophobic residues.[5] We reasoned

that a scaffold presenting this binding interface would need to be larger than previously reported mini-proteins[26] and the hydrophobic residues on the surface should be fewer in number than the number in the core of the protein to avoid misfolding or instabilities. Our newly generated helical scaffold proteins range from 48 to 68 amino acid residues. They provide a starting point for either computational redesign or as a starting scaffold for more traditional approaches for protein engineering using library generation and selections.

2.1 | De novo design of scaffold protein library

For the de novo design of proteins using the macromolecular software suite Rosetta,[27] a blueprint file is needed to describe a single topology with fixed lengths of all helical residues and specific *phi* and *psi* angles for the connecting loops. The backbone (BB) angles for the loop residues are defined using the coarse-grained ABEGO terminology[28] which describes the *phi/psi* angle combinations of the areas populated in natural proteins and summarized by the Ramachandran plot. To exhaust the conformational space possible, we allowed all permutations that had either 14, 15, or 16 residue or in a second set 17, 18, and 19 residues long helices together with one of the four possible loops descriptions (BB, GBB, GB, GABB). We kept the lengths of the helical elements within each set at these ranges to avoid overly unevenly sized helices which could result in unstable overhangs. This resulted in 432 possible blueprints to fold per three-helical bundle. Since not all of these permutations result in reasonable topologies, we allowed in silico folding through fragment insertion[29] following sequence design for a short defined time period to identify reasonable topology descriptions; resulting decoys were filtered based on BB geometry, similarity to existing four-mers in the PDB, helix bending, and packing (Section 4 and SI). Blueprint descriptions that produced large numbers of models were labeled as promising and scaled up in their production time; as expected not all of them produced an output (Figure 1c).

2.2 | Stability selection using yeast surface display

Designs were expressed on the surface of yeast and subjected to three different concentrations of proteases (a mix of both chymotrypsin and trypsin, Figure 2a,b and Section 4). After two rounds of sorting, plasmids were extracted from the sorted pools and starting pool, then
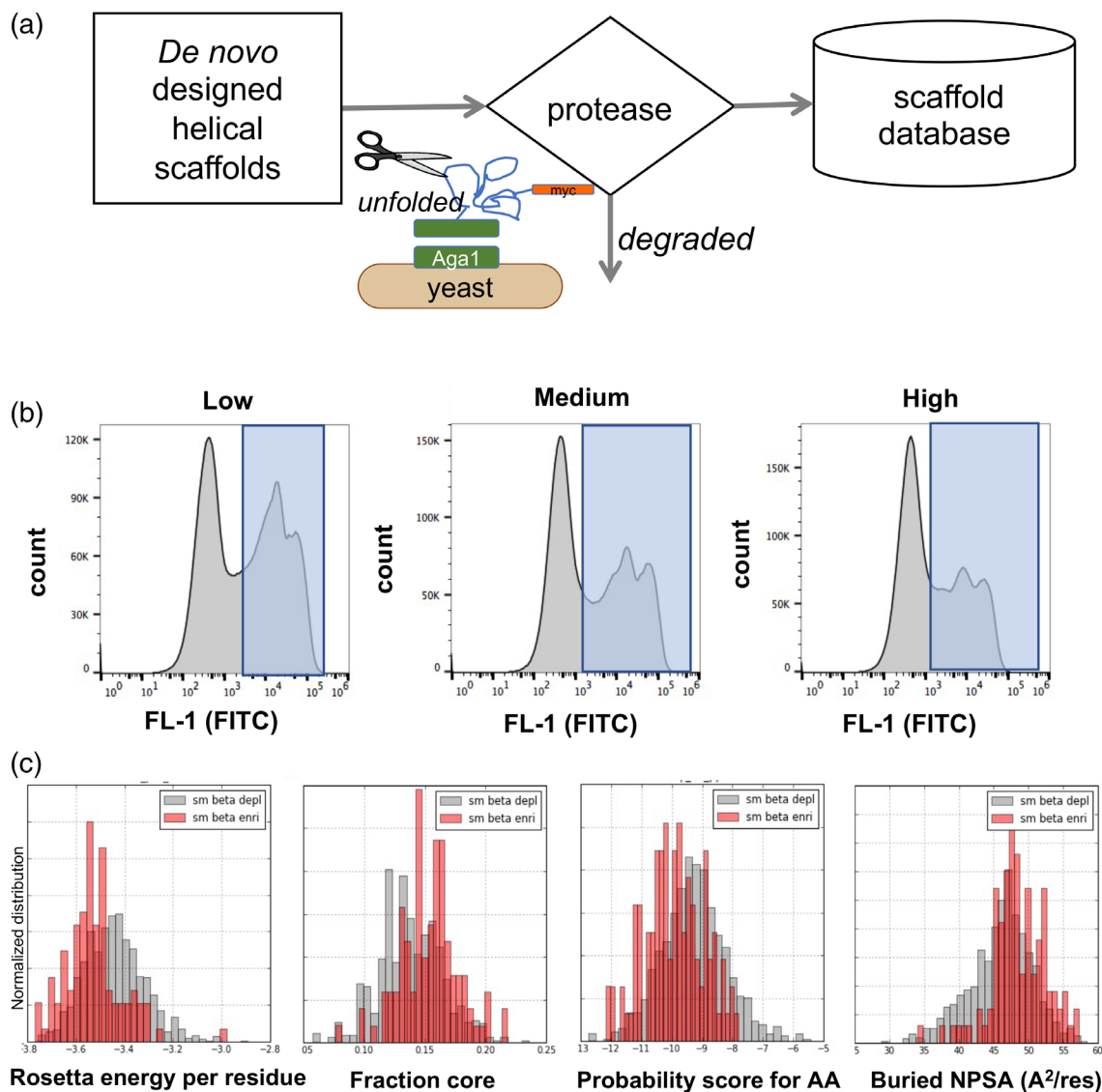
(a)

(b)

(c)

FIGURE 2 Summary of initial yeast surface display screen. (a) Scheme selecting of folded designed proteins using yeast surface display combined with protease digestions. (b) Data of selection of the second round of selection at the three different protease conditions. (c) Comparison between stable and unstable designs for the designed helical bundles.

subjected to next-generation sequencing to obtain enrichment values. Designs with increasing enrichment comparing across the three different concentrations of proteases were considered as stable. While this variation of a screen for folded proteins is less quantitative as a previously reported method in which stability scores can be derived based on computed EC50 values of independent protease titrations[26] it is a much simpler, but efficient way to identify stable proteins. We compared basic Rosetta score terms of successful designs and noticed a minimum threshold for several of the parameters, particularly relevant to the smaller helical bundle set (Figures 2c and S2). Specifically, we noticed that a maximum score per residue (ref2015-based) was noted that required the proteins to be at least at an average of −3.2

Rosetta energy units (REU). Furthermore, as expected, proteins need a hydrophobic core. When comparing stable versus unstable distributions, it appears that the core fraction needs to be above 0.12 with ideally an average buried hydrophobic surface area of above 0.45 Å^2 per residue—at least for the smaller helical bundles. The larger helical bundles reached these thresholds more readily. These cutoff values can be used for either redesigning these scaffold proteins or for developing scaffold sets. In fact, we applied these as filter after grafting and redesigning. Out of a total of 4,000 screened sequences, about 800 design models were identified to be highly protease resistant. These stable proteins scaffolds were then used to graft the DARC binding site onto using Rosetta's motif grafting protocol.[30]

2.3 | Grafting of the DARC binding interface and characterization

Different length variations of the DARC interface helix were incorporated into the identified stable three-helical bundles. Designs with low-binding energy were then evaluated for their folding capability in silico using ab initio structure prediction.[31] Only designs that had clear energy funnels with a minimum close to the designed conformation were ordered (Figure S3).

Sixteen designs were synthesized as gBlock (IDT DNA) and tested via yeast surface display for binding to 10 μM purified, biotinylated DBP-II (Figure S4). All designs showed a distinct binding signal at these high concentrations; designs were then evaluated for their properties as soluble proteins. Twelve designs were successfully cloned, and eight designs expressed soluble protein in *Escherichia coli*. All soluble designs showed a dominant, monodisperse peak at the expected elution volume by size-exclusion chromatography (SEC). Circular dichroism spectrometry established that the designs were indeed folded and helical, and recorded distinct alpha-helical pattern for six of the eight purified de novo designed proteins (Figure S5). All folded designs had a melting point above 95°C and even at that high temperature, a helical profile was recorded except for Dbb10 (Figure S5).

2.4 | Library screening

To optimize binding while maintaining high stability, we first generated a library of all possible point mutations of Dbb4, Dbb7, Dbb10, and Dbb16 and selected for binding using fluorescent-activated cell sorting (FACS). The site-saturation mutagenesis (SSM) libraries were screened for binding to 250 nM and 1 μM of biotinylated DBP-II. To determine enrichment values, plasmids from starting pool and selected populations were extracted and frequencies determined through next-generation sequencing (Figure 3). This allowed to compute their enrichment values. However, Dbb16 appeared to be the tightest binder when comparing enrichment values of all point mutants in parallel which is possible since all their possible point mutations listed were ordered as part of a single oligonucleotide chip (Figure S6), This parallel analysis highlights that a more focus follow-up is possible; and hence we decided to focus on design Dbb16.

We computed the Shannon entropy of its enrichment values to highlight conserved residues as well as positions that could benefit from optimization (Figure 3). When looking at the lower concentration sort, the core contacts F25 and V28 were conserved, as was the salt bridge between E26 and R274 of DBP-II. Interestingly, a phenylalanine was preferred instead of tryptophan at position 29. Also, K24 appeared to be suboptimal and more hydrophobic contact or a tyrosine was preferred, contradicting the Rosetta-based designed residue at this position. Core residues found in Dbb16 were conserved indicating optimal packing of the small protein. Additional contacts to the targeted DBP-II are established by L6 which is conserved and to some extent E2 which has some variability as it is only in the periphery of the interaction. Interestingly, cysteine residues appear to improve binding as well, suggesting that a dimerization of the small protein
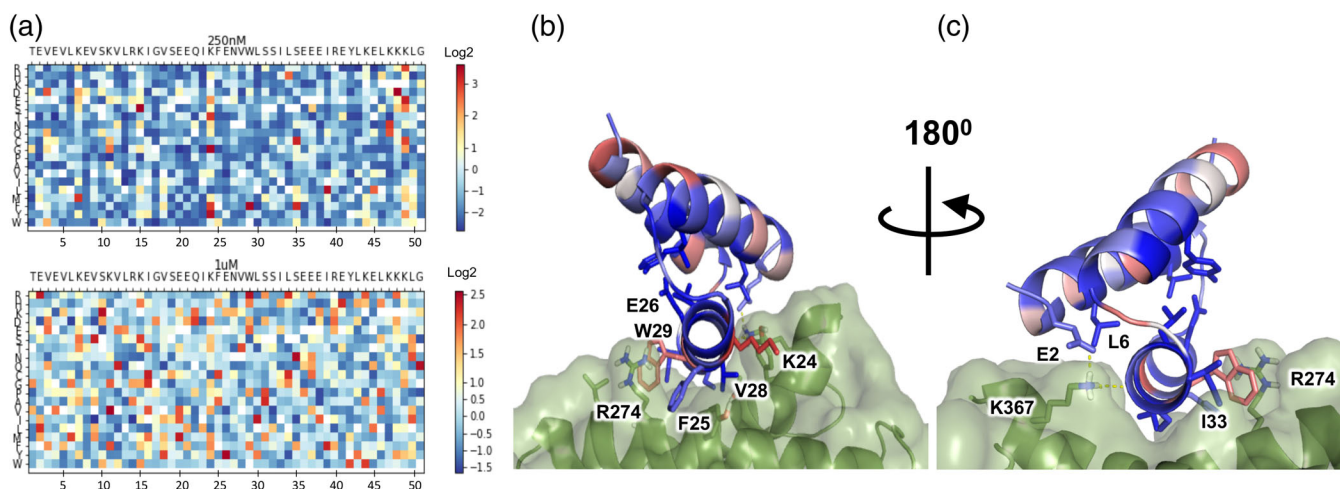

FIGURE 3 Site-saturation mutagenesis (SSM) library results of the lead candidate Dbb16. (a) Heat maps of enrichment values at 250 nM DBP-II and 1 μM Duffy binding protein region II (DBP-II). Red indicates improvement of binding. (b) Model of Dbb16 colored by Shannon entropy computed from the enrichments; blue colors represent more conserved residues. (c) Model of bound Dbb16 turned by 180°.
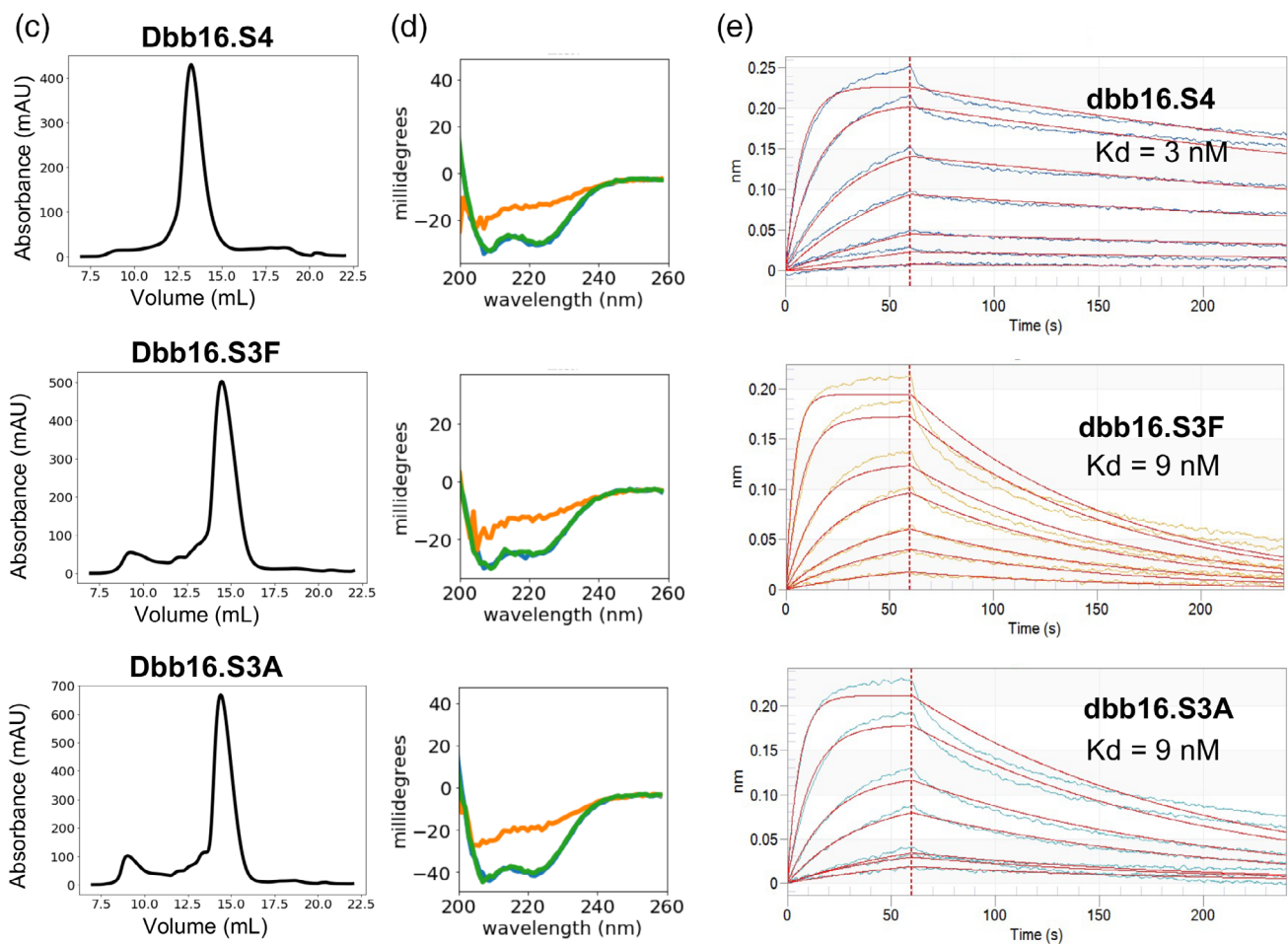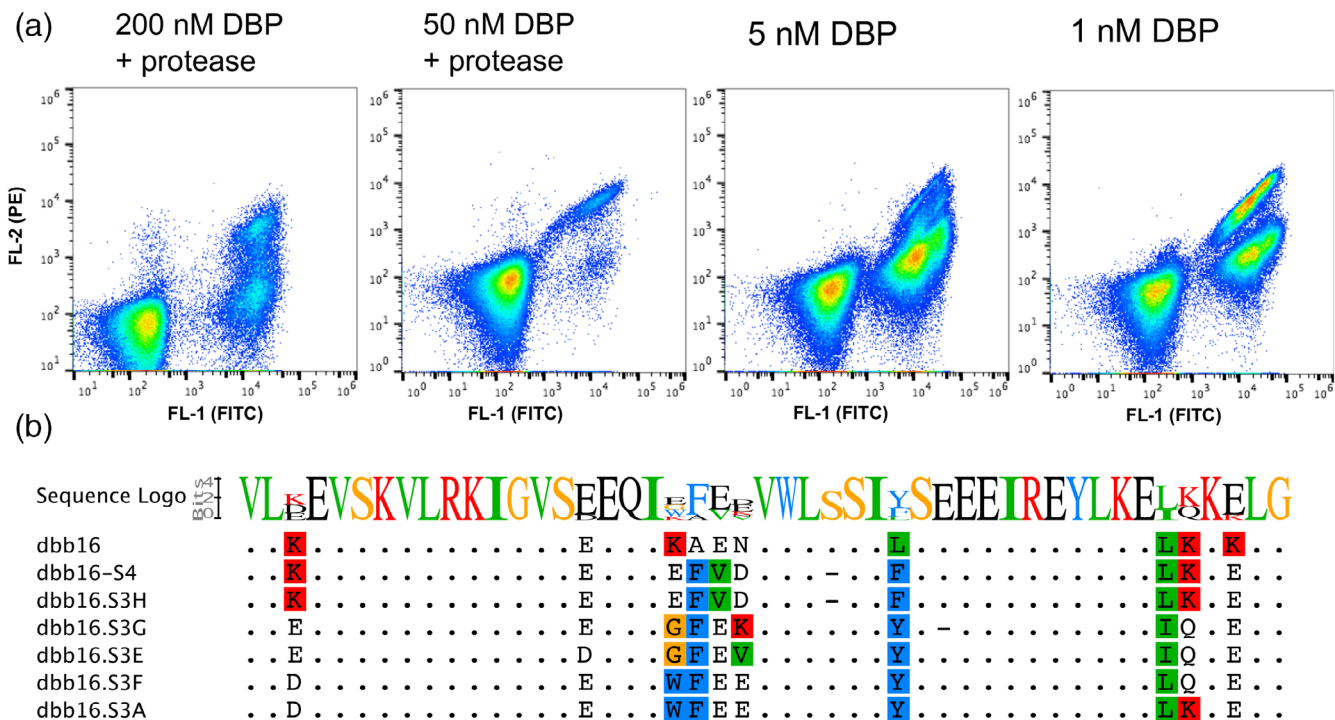
FIGURE 4 Legend on next page.

may be beneficial to further improve the interaction. To optimize binding, we generated a combinatorial library for Dbb16 based on the SSM enrichment data (Table S1); amino acid variations (yet no cysteine residues) were incorporated through degenerative primers with optimized codons using SwiftLib[32] thereby minimizing additional amino acid identities to be encoded.

To ensure high stability, we performed the first two rounds of selection by first incubating displaying yeast cells with the highest concentrations of trypsin and chymotrypsin (as used previously for the screening of the initial de novo designed scaffold library). After the incubation with the protease cocktail, cells were incubated with 200 nM biotinylated DBP-II (bDBP-II) for round 1 and 50 nM bDBP-II for round 2 before sorting. The third and fourth rounds were performed at 5 and 1 nM bDBP-II (Figure 4a). Clones after three and four rounds of sorting were sequenced (Figure 4b). While round 4 converged onto a few sequences, more diversity was seen for round 3.

2.5 | Characterization of affinity maturated Dbb16 variants

Sequenced clones from rounds 3 and 4 were cloned and expressed solubly for further characterization. Proteins were purified as described for earlier variants and SEC showed monodispersed peaks at the monomeric fraction, except for Dbb16.S4. Interestingly, Dbb16.S4 had a deletion in one of its loops' residues. Ab initio structure predictions using both Rosetta[31] and AlphaFold2[33] suggests that the protein variant has the same fold with a shorter loop and a small change at the end of the second helix leaving the interface and core intact (Figure S7). As our affinity maturation process increased the hydrophobic content of the designed interface found in Dbb16.S4, it is likely to be the site for dimerization. We have noted before that highly hydrophobic interfaces of designed protein binders tend to dimerize at their designed interface. Structure predictions using Alhpafold2 provided a possible homodimeric model for Dbb16.S4 (Figure S7). We then measured the secondary structures of the top three variants via CD and found all variants to be helical as before (Figure 4d). Temperature melts showed that the affinity maturated variants do unfold at 95°C, but refold into their original conformation upon cooling down to room temperature. To confirm binding as soluble constructs, we measured binding using biolayer interferometry (BLI) and as expected, the converging sequence after four rounds of sorting bound the tightest (Figure 4e) which bound three times tighter than the round 3 variants. Its improvements are derived from a slower off-rate.

2.6 | Inhibition of DBP binding to RBCs

A tremendous challenge to develop *P. vivax* treatment option is the fact that the parasite cannot be cultivated under laboratory conditions as well as to be grown in the actual Anopheles mosquito. As an approximation to see whether we could inhibit reticulocyte entry mediated through bDBP-II, we utilized previously reported FACS assay that utilizes recombinant protein.[18] Titrating our inhibitor Dbb16.S4 with the protein (0.01 nM–3 μM interval of inhibitor's concentration), we were able to determine an IC_{50} of 72.5 nM (Figure 5a). The inhibitor fully occupies the DARC binding site and extends beyond the original contacts. Further, even though it is a small protein of 5.5 kDa, it should sterically block the dimerization of DBP as our model indicates (Figure 5b).

3 | DISCUSSION

Malaria affects a third of the world's populations and new means to treat and diagnose this disease are necessary, especially for *P. vivax*. Here, we first generated a database of highly stable three helical bundles (3H) as new protein scaffolds that can endure high concentration of protease and have high melting temperatures. These proteins can serve as a starting point for both traditional engineering projects as well as scaffolds for the computational redesign of new protein binders or inhibitors. We further were able to reveal basic parameters that determine minimum thresholds for the 3H bundles in order to be folded. Second, we proceeded to utilize this resource to develop a receptor mimicry to target the DBP-II protein of *P. vivax*. DBP-II is displayed during its merozoite stage and binds to ectodomain of DARC

FIGURE 4 Library selections and biochemical characterization of Dbb16. (a) Fluorescent-activated cell sorting (FACS) rounds for affinity maturation of Dbb16. The concentration of biotinylated Duffy binding protein region II (bDBP-II) was steadily decreased. Round 2 started by first incubating the displaying yeast cells with 0.25 μM trypsin and 0.05 μM chymotrypsin before washing and incubating with 50 nM bDBP-II. (b) Sequences of identified clones after three and four rounds of sorting. (c) Size-exclusion chromatography (SEC) profiles monitored at 280 nm absorption wavelength of soluble identified variants after expression in *Escherichia coli*. (d) Far-ultraviolet CD spectra of variants; 25°C (blue), 95°C (orange), and at 25°C after heating it up to 95°C (green). (e) Biolayer interferometry of variants binding to immobilized bDBP-II.
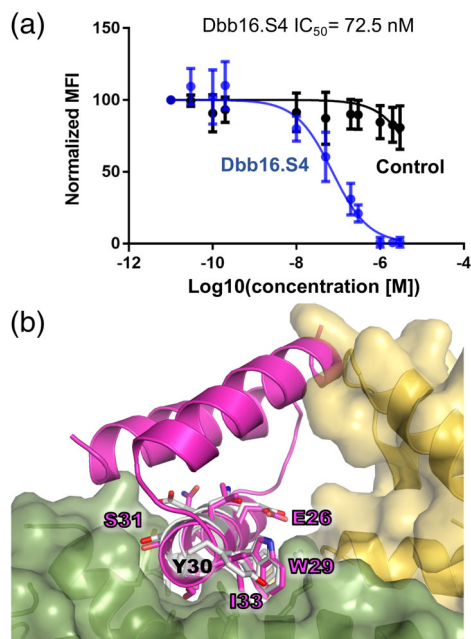
FIGURE 5 In vitro inhibition and proposed mode of action. (a) Inhibition of Duffy binding protein region II (DBP-II) binding to reticulocytes. (b) Model of the inhibitor (pink) bound to one unit of DBP-II (green) which avoids binding to the host receptor Duffy antigen/receptor for chemokines (DARC) (helical interface fragment in grey); the inhibitor would clash with the second DBP-II domain (yellow) and thereby would prevent their dimerization.

as a critical step for reticulocyte invasion of the parasite.[5,6,16–18] We grafted the helical interface of DARC into several hundreds of these starting scaffolds and redesigned for optimized intra- and interface contacts with DBP-II. Designed models to test experimentally were then filtered based on identified thresholds, computed binding energy and in silico refolding evaluations. Six designs were monodispersed and folded proteins out of which we optimized binding affinity for the smallest protein which also introduced a clash with the BB of the second DBP-II domain and thereby could potentially inhibit dimerization.

The lead inhibitor designed and optimized here exhibits single digit nanomolar binding affinity to DBP-II and thereby should effectively disrupt the interaction between DBP-II and DARC and hinder DBP dimerization. Based on our model of the designed protein, a dual mode of action is expected: first blocking of the DBP-II:DARC binding pocket and second preventing homo-dimerization of DBP. The designed protein inhibitor is highly resistant to protease digestions and can endure high temperatures necessary for stability in the body and for storage and distribution. Strikingly, the designed inhibitor readily refolds upon denaturation and assumes its original conformation. We thereby believe our designed and optimized inhibitor could be a valuable therapeutic candidate for treatment of the disease.

4 | MATERIALS AND METHODS

4.1 | De novo helix design protocol

Each protein secondary structure can be described as a sequence describing the bins of the *phi-psi* angles in the Ramachandran plot, which we categorize into five different bins termed ABEGO.[34] To build the tertiary structure of a protein, short, structured fragments with the desired ABEGO sequence are used for its assembly. We generated blueprints for all possible permutation of two groups of helix ranges (Group 1: 14, 15, 16 residues; and Group 2: 17, 18, 19 residues) together with four types of different loop conformations (BB, GBB, GB, GABB). Blueprints were generated using a custom python script (data files). Using the same computing power (2.1 MHz Intel Xeon) and filters, we thereby allowed "scouting" runs of 432 design trajectories (one for each blueprint) producing a range of possible BBs. Since filters were applied, not all of them produced output. We then used the most "productive" blueprints (blueprints that produced more than 10 designs in 2 hr) to generate more models.

The design protocol starts by building a BB in the centroid mode[34] through fragment insertion with desired secondary structure conformations. As simplified scoring function, we used scoring terms cenpack, hbond_sr_bb, hbond_lr_bb, atom_pair_constraint, angle_constraint, and dihedral_constraint. If the BB had the desired secondary structure conformation, and helices not bend past 15°, sequence design was initiated and attempted 10 times in order to fulfill the following criteria: average degree of at least 9 amino acids within a radius of 10 Å, less than 6 alanine residues and a packing score above 0.58. Obtained decoys were then filtered in a second script by fa_atr_per_res, local four-mer geometry, cavities, contact_per_residue, and mismatch of local structure with predicted secondary structure based on Psipred[34]; see filter_HHH.xml protocol for details. Resulting models were then evaluated through ab initio structure prediction to confirm that they were at their global energetic minimum using a modified protocol with reduced numbers of fragments[35] for the prediction process. Rosetta_scripts protocols and python scripts for blueprint generation can be found under https://github.com/strauchlab/DBP.

4.2 | Design of DBP binding proteins

Using the most recent implementation of Epigraft[36] in form of the MotifGraft mover, segments of the core interface helix of DARC (residues 22–30) of varying lengths extracted from the co-crystal structure (PDB ID 4nuv) were aligned with each scaffold using fragment

superposition, not endpoint joining. If the rms distance between the joining BB endpoints was less than 0.25 Å and diverged less than 1.0 Å overall when aligning the replaced scaffold fragment with the interface helix of DARC, grafting was attempted. All predicted interface residues of suitable scaffolds were substituted with alanine and, using BB minimization, we attempted to close broken bonds. Then, both contacts interfacing with the grafted element as well as all residues of the grafted helix except for residues 22, 23, 25, 26, 28, 29, and 30 (which were declared as "hotspots" within the grafting mover) were redesigned. Both intramolecular and intermolecular interfaces were then extensively redesigned using three rounds of the most current RosettaDesign[34] while also allowing rigid-body minimization. The best scoring design for each successful graft solution was kept if the computed binding energy below −24 REU. Designs with a lower than −3 REU per residue score, a ratio of 0.12 of hydrophobic core residues, a high degree of connectivity of 10 were subjected to forward folding and designs that appeared at a clear global minimum were ordered for experimental evaluation. Relaxed DBP structure and helix motif, as well as the Rosetta_scripts grafting protocol can be found under https://github.com/strauchlab/DBP.

4.3 | Library generation and yeast surface display

Amino acid sequences of designed proteins were encoded into DNA using DNAworks2.0 and "ecoli2" codons.[36] Genes and genes encoding SSM libraries were cloned into pETCON using homologous recombination as previously reported.[37,38] For library transformation, EBY100[39] cells were transformed with 2 μg of the gene library together with 1 μg linearized pETCON (SalI, XhoI, NheI digested)[38] using previously reported protocol.[40] Transformation resulted in 2–5 × 10⁷ cells. Yeast cells were induced for about 16–18 hr at 23°C (protase sorts of de novo scaffolds) or 30°C, then washed once with PBSF.

4.4 | Protease digestion of scaffold library

Protease reagent trypsin–EDTA (0.25%) solution was purchased from Life Technologies and stored at stock concentration (2.5 mg/ml) at −20°C. α-Chymotrypsin from bovine pancreas was purchased from Sigma-Aldrich as lyophilized powder and stored at 1 mg/ml in Tris buffered saline (20 mM Tris, 100 mM NaCl, pH 8.0) (TBS) with 100 mM CaCl₂ at −20°C. Each reaction used a freshly thawed aliquot of protease reagent. EBY100 yeast cell cultures were induced in SGCAA for 16–18 hr at 22°C for the first round and 30°C for the second round of sorting.[39]

For each protease condition, 1–3 × 10⁸ cells were aliquoted, pelleted, and washed in 1.00 ml of TBS. Proteolysis was initiated by resuspending the pelleted cells in 1.00 ml of room temperature protease mix at three different protease conditions: 0.05 μM trypsin and 0.01 μM chymotrypsin ("Low"); 0.25 μM trypsin and 0.05 μM chymotrypsin ("Medium"); and 1.25 μM trypsin and 0.25 μM chymotrypsin ("High"). After 5 min of incubation at room temperature, cells were pelleted for 1 min and washed with 1.00 ml of chilled PBSF (20 mM NaPi, 150 mM NaCl, pH 7.4 phosphate buffered saline [PBS] with 1% bovine serum albumin [BSA]); this was repeated once more to wash out the protease. Protease-digested cells were labeled with 10 ng/ml anti-c-Myc antibody conjugated to FITC for ≥15 min on a rotator at room temperature.

For the first round of selections of the SSM libraries, cells were incubated for 2 hr at 23°C with 1 μM bDBP-II, before adding a fourth of the used concentration of streptavidin conjugated to phycoerythrin (SAPE, Invitrogen) and 2 ng/ml anti-c-Myc antibody conjugated to FITC and incubating for another hour. For non-avid conditions in later sorts, cells were incubated for 1 hr at 23°C while rotating, unless noted otherwise, and then washed once with ice-cold PBSF before resuspending in 100 μl with 72 nM SAPE and 2 ng/ml anti-c-Myc antibody conjugated to FITC for 15–20 min on ice. Cells were washed once with 1 ml ice-cold PBSF and stored as pellets on ice before their final resuspension in 1 ml ice-cold PBSF followed by FACS.

4.5 | Library preparation and next-generation sequencing

Plasmids were extracted from 5 × 10⁷ yeast cells for each selection and starting pool, and 6.7 × 10⁷–1.3 × 10⁸ yeast cells from the protease-digestion sorts, as previously described.[41,42] After ExoI (NEB) and Lambda exonuclease (NEB) treatment, the Illumina sequencing primers sequence, flow cell adapters, and selection specific barcodes were added via two nested PCRs using NEBnext. They additionally add 12 entirely degenerate bases at the beginning of the forward and reverse read, ensuring adequate diversity for the Illumina base-calling algorithms. A 250-PE kit was sequenced with 15% phiX at 10 pM concentration using a NextSeq (Illumina). After quality control filtering,[41] we calculated enrichment for each sort by dividing the frequencies of selected by ones seen in the starting pool.

4.6 | Protein expression and purification of designed proteins

Genes encoding the proteins were cloned into pET29b between the restriction sites *NdeI* and *XhoI* and expressed for 4 hr in Terrific Broth (BD Difco) using Lemo21 (NEB) cells at 18°C using IPTG 0.5 mM. Cells were resuspended in 35 ml PBS (150 mM NaCl and 25 mM phosphate buffer at pH 7.4) and lysed using a M110P Microfluidizer or a sonicator. Insoluble cell debris was removed by centrifugation for 20 min at 40,000g. Supernatant was applied to gravity-flow columns containing 1 ml of Ni-NTA for each 500 ml of culture, washed with 50 ml PBS and 50 ml PBS containing 30 mM imidazole. Proteins were eluted with 20 ml of 250 mM imidazole in PBS, followed by Ni-NTA (QIAgen) purification. The elution fraction was subjected to further cleaning via SEC.

4.7 | Protein production and biotinylation of avi-tagged DBP-II

BirA-tagged Sal-1 DBP-II was prepared as described previously.[5–7,19] Briefly, inclusion bodies were solubilized in 6 M guanidinium hydrochloride and refolded via rapid dilution into 400 mM L-arginine, 50 mM Tris (pH 8.0), 10 mM EDTA, 0.1 mM PMSF, 2 mM reduced glutathione, and 0.2 mM oxidized glutathione. Refolded protein was captured on SP Sepharose Fast Flow resin (GE Healthcare) and further purified by SEC (GF200; GE Healthcare) into 10 mM Hepes (pH 7.4) and 100 mM NaCl.

BirA-tagged Sal-1 DBP-II was buffer exchanged into PBS. Then, 50 μl of BiomixA (Avidity), 50 μl of BiomixB (Avidity), and 10 μl of 5 mM *D*-biotin (Avidity) were added to the protein along with BirA ligase, followed by overnight incubation at 4°C. The biotinylation was confirmed by Western Blot using Streptavidin HRP-conjugate (Thermo Scientific). Before use, the reaction mix was buffer-exchanged into PBS.

4.8 | Binding analysis

Titrations were performed at 30°C while rotating at 1,000 rpm on an OctetRED96 BLI system (ForteBio, Menlo Park, CA) using streptavidin-coated biosensors. Sensors were equilibrated for 20 min in PBSTB buffer (PBS, 0.002% Tween 20, 0.01% BSA). Each sensor was loaded with 20 nM biotinylated DBP-II for 100 s. For baseline, association and dissociation time intervals of 60 s, 240 and 300 s were applied, respectively.

4.9 | FACS assay to evaluate inhibition of reticulocyte invasion

Biotinylated purified DBP-II protein in concentration of 10 nM was incubated with 11 different concentrations of Dbb 16T (from 0.03 nM to 1 μM interval of concentrations) for 1 hr at RT. Red blood cells were added and incubated for an additional 1 hr at RT. To detect bound DBP Alexa Fluor 488 Streptavidin conjugate (Fisher) was added to the mix and incubated at RT for 1 hr followed by washing twice with PBS. FITC labeling was measured by flow cytometry. An IC_{50} value for Dbb 16T was calculated in GraphPad Prism from five independent biological replicates. IC_{50} curve for the coil peptide of similar size was plotted as a negative control.

4.10 | AlphaFold2 predictions

Structure prediction of Dbb16 variants was performed using the ColabFold interface[43] using the "no templates" and the multiple sequence alignments option.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Additional data are summarized in the supplementary information. Additional scripts and stable scaffold proteins are available under https://github.com/strauchlab/DBP. Next-generation sequencing data can be sent upon request.

ORCID

Eva-Maria Strauch https://orcid.org/0000-0001-7382-747X

REFERENCES

1. WHO. World malaria report 2017. Geneva: World Health Organization, 2017.
2. Popovici J, Menard D. Challenges in antimalarial drug treatment for vivax malaria control. Trends Mol Med. 2015;21:776–788.
3. Krotoski WA. The hypnozoite and malarial relapse. Prog Clin Parasitol. 1989;1:1–19.
4. Adams JH, Hudson DE, Torii M, et al. The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. Cell. 1990;63:141–153.
5. Batchelor JD, Malpede BM, Omattage NS, DeKoster GT, Henzler-Wildman KA, Tolia NH. Red blood cell invasion by *Plasmodium vivax*: Structural basis for DBP engagement of DARC. PLoS Pathog. 2014;10:e1003869.
6. Batchelor JD, Zahm JA, Tolia NH. Dimerization of *Plasmodium vivax* DBP is induced upon receptor binding and drives recognition of DARC. Nat Struct Mol Biol. 2011;18:908–914.
7. Chen E, Salinas ND, Huang Y, et al. Broadly neutralizing epitopes in the *Plasmodium vivax* vaccine candidate Duffy binding protein. Proc Natl Acad Sci U S A. 2016;113:6277–6282.
8. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med. 1976;295:302–304.
9. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. Science. 1975;189:561–563.
10. Ovchynnikova E, Aglialoro F, Bentlage AEH, et al. DARC extracellular domain remodeling in maturing reticulocytes explains *Plasmodium vivax* tropism. Blood. 2017;130:1441–1444.
11. VanBuskirk KM, Sevova E, Adams JH. Conserved residues in the *Plasmodium vivax* Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. Proc Natl Acad Sci U S A. 2004;101:15754–15759.
12. Wertheimer SP, Barnwell JW. Plasmodium vivax interaction with the human Duffy blood group glycoprotein: Identification of a parasite receptor-like protein. Exp Parasitol. 1989;69:340–350.
13. Adams JH, Sim BK, Dolan SA, Fang X, Kaslow DC, Miller LH. A family of erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci U S A. 1992;89:7085–7089.
14. Malpede BM, Tolia NH. Malaria adhesins: Structure and function. Cell Microbiol. 2014;16:621–631.
15. Salinas ND, Tolia NH. Red cell receptors as access points for malaria infection. Curr Opin Hematol. 2016;23:215–223.
16. Carias LL, Dechavanne S, Nicolete VC, et al. Identification and characterization of functional human monoclonal antibodies to *Plasmodium vivax* Duffy-binding protein. J Immunol. 2019;202:2648–2660.
17. Grimberg BT, Udomsangpetch R, Xainli J, et al. *Plasmodium vivax* invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. PLoS Med. 2007;4:e337.
18. Urusova D, Carias L, Huang Y, et al. Structural basis for neutralization of *Plasmodium vivax* by naturally acquired human antibodies that target DBP. Nat Microbiol. 2019;4:1486–1496.
19. Chen E, Salinas ND, Ntumngia FB, Adams JH, Tolia NH. Structural analysis of the synthetic Duffy binding protein (DBP) antigen DEKnull relevant for *Plasmodium vivax* malaria vaccine design. PLoS Negl Trop Dis. 2015;9:e0003644.
20. Ntumngia FB, Adams JH. Design and immunogenicity of a novel synthetic antigen based on the ligand domain of the *Plasmodium vivax* Duffy binding protein. Clin Vaccine Immunol. 2012;19:30–36.
21. Ntumngia FB, Barnes SJ, McHenry AM, George MT, Schloegel J, Adams JH. Immunogenicity of a synthetic vaccine based on *Plasmodium vivax* Duffy binding protein region II. Clin Vaccine Immunol. 2014;21:1215–1223.
22. Ntumngia FB, Pires CV, Barnes SJ, et al. An engineered vaccine of the *Plasmodium vivax* Duffy binding protein enhances induction of broadly neutralizing antibodies. Sci Rep. 2017;7:13779.
23. Ntumngia FB, Schloegel J, Barnes SJ, et al. Conserved and variant epitopes of *Plasmodium vivax* Duffy binding protein as targets of inhibitory monoclonal antibodies. Infect Immun. 2012;80:1203–1208.
24. Ntumngia FB, Schloegel J, McHenry AM, et al. Immunogenicity of single versus mixed allele vaccines of *Plasmodium vivax* Duffy binding protein region II. Vaccine. 2013;31:4382–4388.
25. Rawlinson TA, Barber NM, Mohring F, et al. Structural basis for inhibition of *Plasmodium vivax* invasion by a broadly neutralizing vaccine-induced human antibody. Nat Microbiol. 2019;4:1497–1507.
26. Rocklin GJ, Chidyausiku TM, Goreshnik I, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. Science. 2017;357:168–175.
27. Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 2011;487:545–574.
28. Lin YR, Koga N, Tatsumi-Koga R, et al. Control over overall shape and size in de novo designed proteins. Proc Natl Acad Sci U S A. 2015;112:E5478–E5485.
29. Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol. 2019;20:681–697.
30. Azoitei ML, Correia BE, Ban YE, et al. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. Science. 2011;334:373–376.
31. Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. Science. 2005;309:1868–1871.

32. Jacobs TM, Yumerefendi H, Kuhlman B, Leaver-Fay A. SwiftLib: Rapid degenerate-codon-library optimization through dynamic programming. Nucleic Acids Res. 2015;43:e34.

33. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–589.

34. Koga N, Tatsumi-Koga R, Liu G, et al. Principles for designing ideal protein structures. Nature. 2012;491:222–227.

35. Klionsky DJ, Abdalla FC, Abeliovich H, et al. Guidelines for the use and interpretation of assays for monitoring autophagy. Autophagy. 2012;8:445–544.

36. Hoover DM, Lubkowski J. DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. 2002;30:e43.

37. Chevalier A, Silva DA, Rocklin GJ, et al. Massively parallel de novo protein design for targeted therapeutics. Nature. 2017; 550:74–79.

38. Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science. 2011;332:816–821.

39. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. Nat Protoc. 2006;1:755–768.

40. Benatuil L, Perez JM, Belk J, Hsieh CM. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng Des Sel. 2010;23:155–159.

41. Strauch EM, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. Proc Natl Acad Sci U S A. 2014;111:675–680.

42. Whitehead TA, Chevalier A, Song Y, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol. 2012;30:543–548.

43. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. Nat Methods. 2022;19:679–682.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.